



UNIVERSIDAD DE AZUAY

Facultad de Ciencias de la Administración

Escuela de Administración de Empresas

**“Presupuesto financiero anual de la Junta
Administradora de Agua Potable de Baños empleando
proyecciones con la aplicación informática R y técnicas
de Big Data”**

**Trabajo de graduación previo a la obtención del título de Ingeniero
Comercial.**

Autor:

Manuel Hernán Quito Baculima

Director.

Ing. Marcos Orellana

Cuenca – Ecuador 2017

Dedicatoria

Este trabajo va dedicado a mi esposa Elsitita y a mis queridos hijos, a quienes les quedo en deuda por mi tiempo y dedicación que les he privado, y más aún, que he perdido de disfrutar con ellos de tantos momentos valiosos. Espero poder compensarles con el fruto de este esfuerzo, ya que estoy convencido que, por ustedes, todo sacrificio vale la pena.

Hernán.

Agradecimientos

Quiero agradecer a mi familia por el apoyo que siempre me brindaron, a los miembros del directorio de la Junta Administradora de Agua Potable de la Parroquia Baños, en la persona de su presidente, Sr. Manolo Soto, que me permitieron utilizar los datos para el desarrollo de esta tesis. Un agradecimiento especial a los docentes que durante mis años de estudio me guiaron en el proceso de aprender de la administración, la que considero es de las más complejas de las ciencias, ciencia que tanto gusto me ha brindado al permitirme tener una visión más clara y completa de lo que se requiere para que una sociedad salga adelante, esto es, una buena administración.

Gracias a todos, Dios los bendiga.

Índice de Contenidos

Dedicatoria	i
Agradecimientos.....	iii
Índice de Contenidos	v
Índice de Tablas	ix
Índice de imágenes.....	xi
Resumen	xv
Abstract.....	xvii
Introducción.....	1
Capítulo 1: Estudio de la Junta Administradora de Agua Potable de Baños JAAPB. 5	
Introducción.....	5
1.1. Antecedentes	5
1.1.1. Reseña histórica.....	5
1.2. Visión. Misión y Situación Actual	6
1.2.1. Visión de la JAAPB	6
1.2.2. Misión de la JAAPB.....	6
1.2.3. Situación actual	7
1.3. Proceso de potabilización	7
1.4. Cobertura y sistema de distribución	12
Capítulo 2: Elaboración de presupuestos y técnicas estadísticas para tratamiento y proyección de datos.....	13
Introducción.....	13
2.1. Elaboración de presupuestos.....	14
2.1.1. Proceso continuo de presupuesto	14
2.1.2. Proceso de toma de decisiones para la elaboración de un presupuesto 15	
2.1.3. Pasos para preparar el presupuesto:.....	15

2.2. Modelos determinísticos de pronóstico aplicables en la elaboración de presupuestos.....	18
2.2.1. Enfoques de los pronósticos	19
2.2.2. Horizontes de tiempo del pronóstico	20
2.3. Proceso de Pronóstico	21
2.4. Patrones y técnicas de proyección de datos	23
2.4.1. Patrones de datos y elección de técnicas de pronóstico	23
2.4.2. Precisión y fiabilidad	26
2.5. Modelos de pronóstico basados en series de tiempo	27
2.5.1. Enfoque intuitivo	27
2.5.2. Promedios móviles.....	27
2.5.3. Suavizado exponencial.....	28
2.5.4. Suavizado exponencial con ajuste de tendencia	29
2.5.5. Proyecciones de tendencia.....	30
2.5.6. Variaciones estacionales en los datos.....	31
2.6. Regresión lineal simple y correlación.....	31
2.6.1. Análisis de regresión simple	32
2.6.2. Error estándar	33
2.6.3. Análisis de correlación.....	33
Capítulo 3: Nuevas tendencias de tratamiento de datos para uso en administración empresarial.....	37
Introducción.....	37
3.1. Administración del desempeño empresarial	37
3.1.1. Pronóstico guiado por la demanda.....	37
3.2. Big Data	38
3.2.1. Tipos de datos	39
3.2.2. Términos importantes.....	40
3.2.3. Características de Big Data	40
3.3. Analítica de Datos	41
3.3.1. Análisis descriptivo.....	42

3.3.2.	Analítica de diagnóstico	43
3.3.3.	Analítica predictiva	43
3.3.4.	Analítica prescriptiva	44
3.4.	Inteligencia de Negocio (Business Intelligence)	45
3.5.	Análisis científico de datos	45
3.5.1.	BI versus ciencia de datos	45
3.5.2.	Ciclo de vida del análisis de datos	46
3.6.	Proyección y tratamiento de datos mediante técnicas estocásticas	48
3.6.1.	Pronóstico estocástico	48
3.6.2.	Proceso de crear un modelo probabilístico	49
3.6.3.	Regresión múltiple	49
3.6.4.	Aprendizaje automático	51
3.7.	Herramienta R para el tratamiento automatizado de grandes cantidades de datos	52
3.7.1.	Introducción a R	52
3.7.2.	Características de R	52
3.7.3.	Ventajas de usar R	54
3.7.4.	Cuando se recomienda usar R	54
Capítulo 4:	Aplicación de modelos de proyección de la demanda y elaboración del presupuesto	55
	Introducción	55
4.1.	Marco estratégico de planeación	55
4.1.1.	Problemas e incertidumbres a los que la Junta se enfrenta	55
4.1.2.	Objetivos estratégicos	57
4.1.3.	Presupuesto de inversiones del plan estratégico 2014-2023	59
4.2.	Situación económica actual: precios, pagos, consumos actuales y pasados.	61
4.3.	Análisis de necesidades de información basada en proyecciones	64
4.4.	Proyecciones basadas en modelos tradicionales	65
4.4.1.	Demanda de agua potable	65

4.4.2.	Consumo mensual y tendencias.....	66
4.4.3.	Proyección de usuarios al 2014.....	67
4.5.	Modelo de proyección de requerimiento de agua potable.....	67
4.6.	Aplicación del análisis científico de datos - Análisis exploratorio	70
4.6.1.	Descubrimiento.....	70
4.6.2.	Obtención y preparación de los datos a ser analizados	71
4.6.3.	Preparación de la herramienta y documento de análisis	72
4.6.4.	Carga y limpieza de datos	75
4.6.5.	Análisis exploratorio de datos	77
4.7.	Modelo y Proyecciones basadas en técnicas estocásticas mediante R	88
4.7.1.	Selección y evaluación de técnicas de pronóstico.....	88
4.7.2.	Preparación del modelo	89
4.7.3.	Construcción del modelo.....	97
4.7.4.	Informe	107
4.8.	Elaboración del presupuesto anual.....	108
4.8.1.	Análisis y esquema de elaboración de presupuesto	108
4.8.2.	Incidencia del Plan Operativo.....	109
4.8.3.	Actividades y plan presupuestario.....	114
4.8.4.	Presupuesto año 2014	120
Capítulo 5:	Comparación y análisis de resultados	125
5.1.	Análisis de resultados.....	125
5.2.	Conclusiones y recomendaciones.....	125
Glosario.....		127
Referencia bibliográfica.....		129

Índice de Tablas

Tabla 1 Patrones de estacionalidad	25
Tabla 2 Problemas importantes de cada departamento	57
Tabla 3 Objetivos por cada departamento	58
Tabla 4 Inversiones y Gastos Planificados	59
Tabla 5 Presupuesto detallado de Gastos e Inversiones 2014	60
Tabla 6 Ingresos de la Junta en el año 2013.....	63
Tabla 7 Elementos a los que se puede aplicar proyecciones automatizadas.....	64
Tabla 8 Diversas técnicas de pronóstico, utilidad y requerimientos (Nugus 2009) .	89
Tabla 9 Análisis comparativo vertical de activos.....	111
Tabla 10 Análisis comparativo vertical de pasivo y patrimonio	111
Tabla 11 Análisis vertical de ingresos y costos.....	113
Tabla 12 Costos de implementación de actividades medioambientales.....	116
Tabla 13 Costos de implementación de actividades de producción	117
Tabla 14 Costos de implementación de actividades para distribución de agua	117
Tabla 15 Costos de implementación de actividades administrativas	118
Tabla 16 Costos de implementación de actividades en contabilidad	119
Tabla 17 Costos de implementación de actividades en departamento de sistemas	119
Tabla 18 Resumen de presupuesto de ingresos para el año 2014	121
Tabla 19 Resumen de presupuesto de costos y gastos para el año 2014	122
Tabla 20 Resumen de presupuesto de nuevos costos y gastos para el año 2014	123
Tabla 21 Tabla de depreciaciones presupuestadas para el año 2014.....	124

Índice de imágenes

Figura 1 Usuarios por categoría	7
Figura 2 Lugar de captación en el río Minas (Quito B. y Cárdenas V 2015)	8
Figura 3 Mezcla rápida y adición de químicos	8
Figura 4 Limpieza de la sección de floculación.....	9
Figura 5 Sedimentadores.....	9
Figura 6 Sección de filtrado.....	10
Figura 7 Dosificador de cloro gas	10
Figura 8 Vista general de la planta de tratamiento	11
Figura 9 Tanque de reserva actual JAAPB.....	11
Figura 10 Laboratorio de control de calidad	12
Figura 11 Matriz y sub matrices de la JAAPB	12
Figura 12 Diagrama de presupuesto de ventas y producción. (Shim, Siegel, y Shim 2012).	14
Figura 13 Presupuesto maestro para una empresa de producción. (Horngren et al. 2012)	17
Figura 14 Proceso de pronóstico (Hoshmand 2010).....	21
Figura 15 Patrón de datos horizontal (Hoshmand 2010).....	24
Figura 16 Curvas de tendencia determinística (Hoshmand 2010).....	24
Figura 17 Patrón de datos estacional (Hoshmand 2010).....	25
Figura 18 Gráfica de demanda con tendencia creciente y estacionalidad. (Heizer et al. 2009).....	26
Figura 19 Diferencia ente ventas reales y proyecciones mediante promedios móviles. (Heizer et al. 2009).....	28
Figura 20 pronóstico de suavizado exponencial y ajuste de tendencia. (Heizer et al. 2009)	29
Figura 21 Recta de tendencia calculada mediante mínimos cuadrados. (Heizer et al. 2009)	30
Figura 22 Diagrama de dispersión de ingresos de ventas y gastos de publicidad (Hoshmand 2010).	33
Figura 23 Desvío total, explicada e inexplicada para un valor observado de Y (Hoshmand 2010).	34
Figura 24 Diferentes diagramas de dispersión para diversos valores de dispersión (Hoshmand 2010).	35

Figura 25 Comparación semanal de demanda (ventas) y envíos (suministro) de una empresa de bebidas (Chase 2009).....	38
Figura 26 Cinco V de Big Data (Dietrich et al. 2015).....	41
Figura 27 incremento de valor y complejidad de los distintos niveles de analítica (Erl, Khattak, y Buhler 2016).....	42
Figura 28 Generación de reportes a partir de análisis descriptivo de sistemas operacionales.....	43
Figura 29 Analítica de diagnóstico con resultados que pueden resumirse o detallarse (Erl, Khattak, y Buhler 2016).	43
Figura 30 Herramientas de analítica predictiva e interfaz amigable (Erl, Khattak, y Buhler 2016).....	44
Figura 31 La analítica prescriptiva incluye el uso de reglas de negocio (Erl, Khattak, y Buhler 2016).	44
Figura 32 BI puede utilizarse para analizar datos a través de un tablero de control (Erl, Khattak, y Buhler 2016).	45
Figura 33 Comparación entre BI y Ciencia de Datos (Dietrich et al. 2015).	47
Figura 34 Comparativa de consumo diario	62
Figura 35 Consumo en m3 por categoría de usuario	62
Figura 36 historial de recaudaciones	63
Figura 37 Recaudaciones de acuerdo a categoría de usuario	63
Figura 38 Evolución del consumo promedio mensual de agua potable.....	65
Figura 39 Consumo mensual de agua potable. Suavizado con media móvil n=2	66
Figura 40 Tendencia en el consumo mensual de agua potable.....	66
Figura 41 Proyección de usuarios utilizando una línea de tendencia.....	67
Figura 42 Modelo de proyección de requerimientos de distribución.....	69
Figura 43 Modelo de proyección de requerimientos de capacidad de planta	70
Figura 44 Extracción de datos mediante gestor de bases de datos	72
Figura 45 Preparación de nuevo documento RMarkdown	73
Figura 46 datos de nuevo documento RMarkdown	73
Figura 47 Ingreso de comandos en la consola de RStudio.....	73
Figura 48 Título y script R en documento RMarkdown	74
Figura 49 Informe generado en RMarkdown	75
Figura 50 Script R para carga de datos.....	76
Figura 51 Script R para ajuste de tipo de datos	76
Figura 52 Visualización de parte del contenido de un objeto dataframe.....	77
Figura 53 Script R para obtener información de los campos de un objeto dataframe	77

Figura 54 Script R para obtener datos estadísticos de un listado	78
Figura 55 Script R y resultado gráfico de distribución de usuarios por estado	78
Figura 56 Script R para graficar usando la función ggplot.....	79
Figura 57 Usuarios por sector	79
Figura 58 Script R y gráfica de distribución de consumo.....	80
Figura 59 Gráfica de cajas y violín con distribución de consumo de agua	80
Figura 60 Script R para ver estadísticas básicas y para generar gráficos de cajas e histogramas.....	81
Figura 61 Diagrama de cajas e histograma del consumo de agua potable – categoría residencial	81
Figura 62 Evolución de cantidad de socios y de consumos mensuales con scripts en R.....	82
Figura 63 Scripts R para analizar componentes estacionales de consumo de agua	83
Figura 64 Componente estacional de consumo general y de categoría residencial	83
Figura 65 Datos de sensores climatológicos	84
Figura 66 Comportamiento mensual de la temperatura y Script R para la carga y visualización de datos	85
Figura 67 Consumo pico-medio de agua potable y Script R.....	86
Figura 68 Consumo per cápita por categoría y Script R.....	87
Figura 69 Script R para calcular el consumo per-cápita del año 2013.....	88
Figura 70 Correlación entre cantidad de socios y cantidad de agua consumida.....	90
Figura 71 Resultado de aplicar el modelo lineal de análisis al consumo de agua ...	90
Figura 72 Correlación de consumo de agua para categorías comerciales y cantidad de socios.....	91
Figura 73 Script R para aplicar modelo lineal de consumo de agua clasificado por categoría.....	92
Figura 74 Nube de puntos y línea de estimación de modelo lineal para consumo de agua.....	92
Figura 75 Nube de puntos y estimación de modelo lineal para categorías comercial A y B.....	93
Figura 76 Series de tiempo para consumo y cantidad de usuarios	94
Figura 77 Descomposición de una serie de tiempo de consumos de agua potable	95
Figura 78 Script R para aplicar suavizado mediante medias móviles	96
Figura 79 Proyección de consumo de agua potable y cantidad de usuarios	96
Figura 80 Modelo de estimación de consumo futuro con análisis gráfico.....	98
Figura 81 Resultados de modelo gráfico de estimación de consumo de agua potable para un usuario residencial.....	99

Figura 82 Valores estimados de consumo de agua potable para un usuario residencial	100
Figura 83 Resultados de modelo gráfico de estimación de consumo de agua potable para un usuario comercial.....	100
Figura 84 Valores estimados de consumo de agua potable para un usuario residencial	101
Figura 85 Modelo general de estimación de consumo de agua potable para un listado de usuarios.....	102
Figura 86 Resultados de aplicar el modelo de estimación de consumo de agua potable para un listado de usuarios	103
Figura 87 Script para la carga y tratamiento de datos de cobros.....	104
Figura 88 Evolución y tendencia de cobros agrupados por categoría.....	104
Figura 89 Script y gráfica de pagos por consumo de agua agrupados por tipo de usuario	105
Figura 90 Script y gráfica de pagos mensuales promedio por consumo de agua tipo residencial	106
Figura 91 Informe general en HTML generado con RMarkdown	107
Figura 92 Diagrama de presupuesto de consumo y producción	108
Figura 93 Formato de levantamiento de datos para presupuesto.....	115

Resumen

La gran cantidad de datos con los que las empresas cuentan en la actualidad, sirven muy poco si no se cuenta con personal especializado en usar técnicas modernas de modelización y pronóstico, que ayuden a optimizar el proceso de toma de decisiones.

Este estudio analiza el proceso de elaboración de un presupuesto empresarial y la importancia de pronosticar sus variables clave, mediante modelos de predicción del consumo de agua potable que aplican técnicas orientadas a mejorar la precisión de los presupuestos.

El modelo propuesto, elaborado en R, logró un 98% de precisión en la predicción del consumo de agua potable.

Abstract

The large amount of data that companies have today, are of very little use if businesses do not have personnel specialized in using modern techniques of modeling and forecasting that help to optimize the decision-making process. This study analyzed the process of business budgeting and the importance of forecasting its key variables by means of prediction models of drinking water consumption, which apply techniques to improve the accuracy of budgets. The proposed model developed in R, achieved 98% accuracy in the prediction of drinking water consumption.




Translated by,
Lic. Lourdes Crespo

Introducción

Nos encontramos en un momento del tiempo en el que los datos están fluyendo de manera continua a las empresas gracias a la revolución de la era digital representada sobre todo por la masificación de tecnologías móviles, sociales, analítica y big data, cloud computing, y la inteligencia Artificial(Chandrasekaran 2015). Sin embargo, solamente grandes empresas a nivel mundial están invirtiendo en iniciativas de análisis de datos a gran escala, tal es así que reportes de la consultora TATA, muestran que en Latinoamérica, son las grandes empresas de telecomunicaciones, banca, comercialización y gubernamentales asociadas a prospección, las que apuestan con fuerza en estas iniciativas de tratamiento de datos. Esto lo hacen con el objetivo de aplicar la Inteligencia de Negocios (Business Intelligence) en la toma de decisiones, que les permita realizar predicciones y acciones futuras(Vaikunth , Pai 2016).

Problemática y justificación

La falta de una planificación basada en técnicas estadísticas que aprovechen las nuevas propuestas de análisis científico de datos, tendientes a explotar la gran cantidad de información almacenada en las bases de datos en las empresas, impide un análisis más acertado y técnico en la administración y por ende impide mejorar la efectividad empresarial.

Este trabajo busca analizar las técnicas administrativas tradicionales de elaboración de presupuestos, para luego aplicar los nuevos procesos de análisis basados en Big Data y técnicas estocásticas de proyección, mediante el uso del software R.

Se utilizará como insumo principal los datos de consumo de agua potable y de las recaudaciones de la Junta Administradora de Agua Potable de la Parroquia Baños.

Fundamentación teórica

De acuerdo con Jorge Burbano (Burbano Ruíz 2011), los pronósticos de manejo del efectivo son importantes para evitar situaciones de insolvencia y permitir la valuación científica de la empresa.

Actualmente, con el avance de la informática, se han desarrollado nuevas metodologías de administración empresarial que hacen uso de esta herramienta para mejorar su desempeño, como es el caso de BPM, que son las siglas en inglés de la *Administración del desempeño empresarial* (Business Performance Management) (Turban 2011).

Considerando lo anterior, tenemos que para administrar una empresa y lograr cumplir los objetivos de forma eficiente, es necesario realizar estudios adecuados (Baca Urbina 2013) que involucre los análisis de demanda que permitan determinar la cantidad de bienes o servicios que la población objetivo va a requerir, en este caso, la cantidad de agua potable, la que a su vez incide directamente en los ingresos.

Para proyectar la cantidad demandada, es necesario realizar pronósticos mediante técnicas tradicionales de análisis estadístico y cuantitativo, tales como series de tiempo, promedios móviles, suavizado exponencial y técnicas de regresión (Bonini et al. 2000). En el caso del análisis de series de tiempo, se pueden aplicar análisis de tendencias seculares, variaciones cíclicas y variaciones estacionales en el caso de ser necesarias (Douglas A Lind et al. 2012).

La influencia de la informática que está ya varios años involucrada directamente en los negocios, ha permitido que se genere gran cantidad de datos, que sin embargo debido a la falta de herramientas adecuadas y a la falta de personal capacitado, no se ha logrado explotar el potencial que tienen tales datos para optimizar la toma de decisiones empresariales. Hoy en día esta problemática ha derivado en la necesidad de herramientas que permitan trabajar con grandes volúmenes de datos, aplicando técnicas de análisis estadístico basadas en la aplicación de modelos estocásticos y el análisis científico de datos.

Debemos tener cuidado con el uso apropiado de los datos, como nos lo alerta Jeffrey Leek, profesor de la especialización Data Scientist de la Universidad Johns Hopkin, al expresar que la combinación de datos y el deseo de una respuesta, no asegura que pueda obtenerse una respuesta razonable de un grupo de datos dado, sin importar cuan grandes sean tales cantidades (What about Big Data. Curso Data Scientist).

Frank Ohlhorst en su libro Big Data Analytics, turning big data into big money (Ohlhorst 2013), expresa que en la actualidad los negocios están recolectando cada vez más información, creando así la denominada "Big Data". Información que va desde registros de transacciones comerciales hasta videos e imágenes, pasando por llamadas telefónicas con referencias de posiciones geográficas, datos de sensores climáticos, consumos de recursos, búsquedas en línea, contenido creado en redes sociales, etc. Esta gran cantidad de datos generó el problema inicial de almacenarlos y procesarlos rápidamente, pero una vez solucionado esto, surge el nuevo problema de explorar múltiples fuentes de datos para generar inteligencia (Business Intelligence) que ayude en la toma de decisiones (Decision Support Systems) para

solucionar problemas y producir un incremento de las utilidades, generando valor para la empresa; la solución se está logrando mediante la aplicación de un proceso denominado Big Data Analytics.

La importancia de Big Data se puede verificar en las tecnologías requeridas para su análisis y su aporte a la administración, como son: *Business Intelligence* para generar información que mejore la toma de decisiones de los administradores, *Data Mining* para encontrar patrones de comportamiento que permitan crear modelos de datos con fines predictivos, *Aplicaciones estadísticas* para analizar la información tanto de forma descriptiva como predictiva, para identificar tanto riesgos como oportunidades de negocios, finalizando con el *Modelamiento de datos* que mediante la aplicación de múltiples escenarios permita analizar y visualizar el efecto que tendrían en los resultados empresariales (Ohlhorst 2013).

Objetivo general: Elaborar el presupuesto operacional de la Junta Administradora de Agua Potable utilizando modelos predictivos con el software R y técnicas de análisis de Big Data.

Objetivos específicos:

- Analizar el estado y problemática de la Junta Administradora de agua potable de Baños debido a la falta de información prospectiva.
- Analizar las distintas técnicas administrativas y estadísticas para tratamiento y proyección de datos.
- Analizar las nuevas herramientas Big Data para el tratamiento de grandes cantidades de datos y la aplicación del análisis científico de datos en los negocios.
- Diseñar una propuesta de administración para aplicar los modelos de análisis y proyección de la demanda en la elaboración de presupuestos, utilizando el software R para el tratamiento de Big Data.

Metodología

El estudio se aplicará a la Junta Administradora de Agua Potable de la Parroquia Baños, mediante la recopilación de información sobre los requerimientos de datos proyectados que se necesite para mejorar la elaboración de presupuestos y en consecuencia la toma de decisiones. Esto se realizará mediante entrevistas con los administradores y en base a esa información se analizarán las técnicas clásicas de

proyección aplicadas en el campo empresarial en general y a las empresas de agua potable en particular.

Se recopilarán datos históricos de consumos, pagos, cantidad de usuarios y otros que sean de interés en la aplicación de los modelos más recomendables para el caso. Se diseñará un modelo determinístico para analizar la capacidad y eficiencia del sistema de tratamiento y distribución de agua potable.

El siguiente paso será la aplicación del análisis científico de datos, para lo que se procederán a aplicar las técnicas de recopilación, limpieza, análisis descriptivo y modelización de datos, para aplicar los modelos estocásticos tendientes a la proyección de la demanda de agua potable. Para esto se utilizará el software de tratamiento estadístico R, que permite realizar programas orientados a aplicar las técnicas antes mencionadas de análisis científico de datos de manera automatizada. En el caso de la Junta, se tratarán cerca de 1 millón de registros.

Utilizando los datos contables de los años pasados y los nuevos datos proyectados obtenidos en los pasos anteriores, se realizará el presupuesto operativo anual de la Junta y se diseñarán la propuesta de pasos necesarios para que este proceso administrativo sea replicable y pueda adoptarse en otras áreas.

Capítulo 1: Estudio de la Junta Administradora de Agua Potable de Baños JAAPB.

Introducción

La Junta Administradora de agua potable de la parroquia Baños, es una organización sin fines de lucro, encargada de la provisión del servicio de agua potable a los habitantes de la parroquia Baños y parte de Narancay.

Esta organización está presidida por un directorio elegido por votación popular, la que se realiza cada dos años.

La oficina se encuentra ubicada junto a la plaza central de la parroquia Baños, en donde también se realizan las actividades de recaudación de las planillas correspondientes al consumo de agua potable. La recaudación también se realiza a través de tres cooperativas de la localidad.

La planta de tratamiento de potabilización del agua se encuentra ubicada en las partes altas de la Parroquia, en el barrio Cochapata.

1.1. Antecedentes

1.1.1. Reseña histórica

La Junta Administradora de Agua Potable de Baños, en adelante JAAPB, remonta sus inicios a la década de 1950, cuando el entonces Párroco, Dr. Alfonso Carrión Heredia, promovió el aprovechamiento del agua del Rio Minas mediante un canal de riego. En el año 1957 se inicia la construcción de un canal de cerca de 3 Km, terminándose de construir en 1962 (Quito B. y Cárdenas V 2015).

Según el libro “Baños, ensueños del alma”, la provisión de agua para el consumo humano en la Parroquia Baños en ese entonces utilizaba bombas de succión manuales públicas, que pronto resultaron insuficientes debido a la creciente demanda.

En el año 1967 se forma el Comité Pro mejoras de la Parroquia, los que inician la gestión ante el Municipio de Cuenca para conseguir apoyo para el proyecto de Agua Potable.

En el año de 1969 visita a la Parroquia Baños el entonces Presidente de la República Dr. José María Velasco Ibarra, que firmó una donación gubernamental de 750.000 sucres con el fin de iniciar el proyecto de dotación de agua potable. La Ilustre

Municipalidad de Cuenca a través de su alcalde Alejandro Serrano Aguilar, firma un convenio con la empresa Municipal ETAPA para iniciar la construcción de una planta de tratamiento de agua potable de 25 litros por segundo, misma que se terminó en el año 1971 (Brito Alemán 2005).

De acuerdo a la documentación histórica que reposa en la Junta, el manejo del sistema de agua en la parroquia estaba a cargo de un Comité Vecinal de Agua Potable. En el año 1974, mediante una Junta General de usuarios, se elige al primer directorio del Comité General de Agua Potable de la Parroquia Baños.

En el año de 1985 la Junta se legaliza, denominándose “Comité General de Agua Potable Baños”, cuyo fin social es la administración, conservación y aprovechamiento de las aguas del Río Minas.

Posteriormente se promueve la construcción de una nueva planta con capacidad de procesar 60 litros por segundo, la que reemplaza a la anterior. La inauguración se la realiza el 3 de noviembre de 1992.

En el año 1997 se terminó la construcción de un nuevo tanque de reserva de 1.500m³ de capacidad, el que sigue funcionando actualmente.

1.2. Visión. Misión y Situación Actual

El plan estratégico de la JAAPB sirve como guía de acción para las actividades que se desarrollan dentro de la Junta. Es por lo tanto necesario identificar cual es el objetivo final al que el presente estudio va a aportar, mismos que se encuentran condensados en las declaraciones de visión y misión empresarial.

1.2.1. Visión de la JAAPB

“Ser una institución comunitaria que brinda un servicio de agua potable de calidad, eficiente en la utilización de sus recursos, con responsabilidad social y ambiental, siendo un referente a nivel local y nacional” (Quito B. y Cárdenas V 2014).

1.2.2. Misión de la JAAPB

“Garantizar la provisión de agua potable a los habitantes de la parroquia Baños, buscando siempre mejorar la calidad y sustentabilidad en el servicio, cumpliendo con las normas técnicas de potabilización, con un manejo eficiente de los recursos humanos, hídricos, financieros, tecnológicos y materiales” (Quito B. y Cárdenas V 2014).

1.2.3. Situación actual

La Junta Administradora de agua potable de Baños abastece a 30 sectores ubicados la mayoría en la parroquia Baños, existiendo también un grupo importante de socios situados en parte de la parroquia Yanuncay.

Al año 2013 existen cerca de 25.000 personas beneficiadas por los servicios de agua potable, de las cuales el 98% corresponden a la categoría residencial y el 2% restante pertenece a la categoría comercial, en donde los principales usuarios son las hosterías.



Figura 1 Usuarios por categoría
Elaborado: El investigador.

1.3. Proceso de potabilización

En el “análisis de factibilidad técnico-económica para la construcción de una nueva planta de agua potable para la parroquia Baños” (Quito B. y Cárdenas V 2015), se especifica cada uno de los pasos del proceso de potabilización, tal como se resume en los siguiente puntos.

a. Captación

El proceso de potabilización se inicia en la captación, que es cuando el agua en estado natural se toma en el río Minas, para luego ser transportada y procesada en la planta de tratamiento.



Figura 2 Lugar de captación en el río Minas (Quito B. y Cárdenas V 2015)

b. Conducción

El agua se trae desde el río Minas hasta la planta potabilizadora mediante una tubería de 5Km de longitud.

c. Tratamiento inicial - mezcla rápida

Al llegar el agua a la planta pasa por el pozo de mezcla rápida, donde comienza el proceso de potabilización al chocar y generar turbulencia para permitir la oxigenación del agua.

d. Dosificación de productos químicos

A la salida de la mezcla rápida se aplica sulfato de aluminio al agua, que tiene el efecto de ser coagulante. Adicionalmente se aplica polímero para mejorar la floculación y remover tanto el color como la turbiedad del agua.



Figura 3 Mezcla rápida y adición de químicos
Fuente: Archivos de la JAAPB

e. Floculación

La floculación es un proceso en el que, gracias a los productos químicos, las partículas causantes del color y turbiedad del agua, chocan y se adhieren entre para formar una espuma denominada floculo.

f. Sedimentación

En el proceso de sedimentación el agua pasa a dos compartimientos con placas de fibra de vidrio, las que retienen los floculos y los deposita en la denominada zona de lodos. El agua es recolectada y dirigida por canaletas hacia los filtros.

g. Filtración

Aquí el agua pasa por una capa de arena y grava para filtrar partículas muy pequeñas. De este proceso se obtiene un agua clara y limpia.



Figura 4 Limpieza de la sección de floculación
Fuente: Archivos de la JAAPB



Figura 5 Sedimentadores
Fuente: Archivos de la JAAPB



Figura 6 Sección de filtrado
Fuente: Archivos de la JAAPB

h. Cloración

Previo al almacenamiento del agua en el tanque de reserva se aplica cloro gas con el objetivo de eliminar bacterias microscópicas que pudieran haber pasado los filtros.



Figura 7 Dosificador de cloro gas
Fuente: Archivos de la JAAPB



Figura 8 Vista general de la planta de tratamiento
Fuente: Archivos de la JAAPB

i. Almacenamiento

El agua ya apta para el consumo humano, se almacena en un tanque con capacidad de 1.500 m³, desde donde se distribuirá a los diversos sectores.



Figura 9 Tanque de reserva actual JAAPB
Fuente: Archivos JAAPB.

j. Pruebas de calidad

Constantemente se toman muestras en el tanque de almacenamiento y en diversas zonas atendidas, esto con la finalidad de realizar análisis físicos, químicos y

bacteriológicos, que son requeridos por la norma INEN 1-108 para asegurar la potabilización de agua.



Figura 10 Laboratorio de control de calidad
Fuente: Archivos JAAPB.

1.4. Cobertura y sistema de distribución

El agua se distribuye a los diferentes sectores de Baños mediante 4 matrices principales. La siguiente figura muestra la estructura general de las zonas que actualmente son atendidas



Figura 11 Matriz y sub matrices de la JAAPB
Elaborado: El investigador

Capítulo 2: Elaboración de presupuestos y técnicas estadísticas para tratamiento y proyección de datos

Introducción

La administración consiste esencialmente en fijar metas orientadas al cumplimiento de los objetivos de la organización y luego implementar las actividades necesarias para alcanzar tales metas mediante el empleo eficiente del talento humano, los materiales y el capital.(Welsch y Vázquez Prada Gutiérrez 2005).

Esta primera actividad administrativa, conocida como planificación, requiere de una revisión y comprensión de los hechos pasados y presentes, con el complemento de la previsión futura de sucesos que puedan afectar a la empresa, de tal manera que las actividades planeadas estén acordes a la realidad en la que posiblemente se enfrentará la empresa. El resultado de este análisis será el plan, que para operativizarse eficientemente requiere de los presupuestos.

Para Horngren, un presupuesto es *“a) la expresión cuantitativa de un plan de acción propuesto por la administración para un periodo determinado y b) una ayuda para coordinar aquello que deberá hacerse para implementar dicho plan”* (Horngren et al. 2012). El resultado final esperado en la empresa se refleja en el presupuesto financiero, que cuantifica los posibles valores de ingresos y egresos, resumidos en los estados financieros presupuestados.

Sin la existencia de presupuestos, es difícil que el personal de la empresa sea consciente de si están o no en la ruta correcta para alcanzar sus metas; por lo que la existencia del presupuesto es también importante en el otro elemento de la administración, que es el control, que Welsch lo define como *“el proceso de medir y evaluar el desempeño de cada componente organizacional de una empresa, y efectuar la acción correctiva, cuando sea necesaria, para asegurar el cumplimiento eficiente de los objetivos, metas, políticas y normas de la empresa”* (Welsch y Vázquez Prada Gutiérrez 2005).

Resumiendo, y como lo especifica Murray, un presupuesto es un plan para conseguir y gastar dinero para alcanzar metas específicas en un tiempo determinado; por lo que quienes preparan presupuestos y planes deben establecer objetivos específicos, examinar el pasado, el presente y el futuro, e identificar las acciones específicas y los costos necesarios para alcanzar las metas. Es necesario considerar que la calidad del presupuesto depende de la colaboración, el tiempo y el esfuerzo dedicados a éste por el equipo presupuestario; sin embargo debemos estar

conscientes de que no podemos predecirlo todo, por lo que será necesario monitorizar y actualizar el presupuesto según se avanza en el tiempo (Dropkin, Halpin, y La Touche 2007).

2.1. Elaboración de presupuestos

Para elaborar los presupuestos, necesitamos determinar los ingresos y los costos en los que se espera incurrir en el período de estudio, que normalmente es un año.

En una empresa de producción, los costos se dividen en dos categorías principales: Costos de fabricación y Gastos de operación. Los costos de fabricación se subdividen en materiales directos, mano de obra directa y gastos generales de fabricación.

El presupuesto de ventas, o, en el caso de una empresa de servicio de agua potable, el consumo, es el punto de partida para preparar el presupuesto de producción; es este volumen estimado de ventas o consumo el que influye en casi todos los demás elementos que aparecen en el presupuesto maestro, tal como se puede ver en la figura 12, por lo que se debe mostrarlo tanto en cantidad como en valor. La meta planteada puede estar definida de acuerdo a razones de punto de equilibrio, ingreso, ventas o consumo objetivo (Shim, Siegel, y Shim 2012).

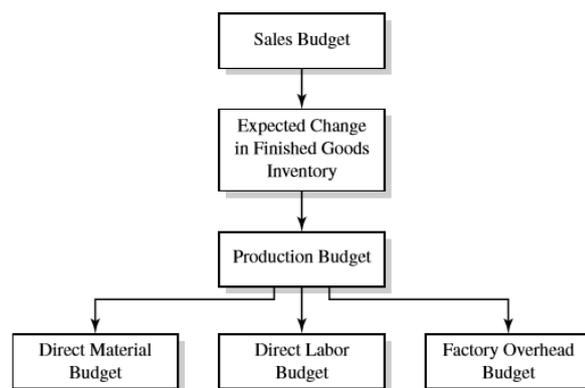


Figura 12 Diagrama de presupuesto de ventas y producción. (Shim, Siegel, y Shim 2012).

2.1.1. Proceso continuo de presupuesto

Horngren define el proceso que el equipo presupuestario, compuesto por gerentes y contadores, debe realizar cada período para obtener el presupuesto general, siendo tales pasos los siguientes:

1. El equipo planea el desempeño de la organización basándose en el desempeño anterior y en los cambios proyectados para el futuro.
2. La alta gerencia proporciona a los mandos subordinados un conjunto de expectativas específicas que se espera que cumplan.

3. Los contadores ayudan a los gerentes a investigar las variaciones con respecto a los planes para tomar acciones correctivas de ser necesario.
4. El equipo presupuestario utiliza la retroalimentación del mercado, los cambios en las condiciones y sus experiencias para hacer planes para el siguiente periodo.

2.1.2. Proceso de toma de decisiones para la elaboración de un presupuesto

La realización de un presupuesto involucra la necesidad de tomar decisiones en base tanto a las expectativas futuras como en el desempeño en los períodos anteriores, por lo cual Horngren propone la aplicación de los siguientes pasos al momento de tomar decisiones en la creación de presupuestos.

1. Identificar problemas e incertidumbres, y posteriormente establecer objetivos cuantificables.
2. Obtener información sobre precios, ventas o consumos actuales y pasados.
3. Realizar predicciones acerca del futuro.
4. Tomar decisiones mediante la elección entre alternativas.
5. Implementar la decisión, evaluar el desempeño y aprender.

La figura 13 muestra un diagrama de las diversas partes del presupuesto maestro para una empresa de producción

2.1.3. Pasos para preparar el presupuesto:

La preparación del presupuesto maestro requiere de la elaboración de una serie de presupuestos parciales, que Horngren los describe como:

1. Presupuesto de ingresos. Este presupuesto se ve influenciado por factores inciden en el pronóstico de ventas, como son: el volumen de ventas en periodos anteriores, las condiciones económicas e industriales generales, los estudios de investigación de mercados, las políticas de fijación de precios, la publicidad y las promociones de ventas, la competencia y las políticas reguladoras.
2. Presupuesto de producción (en unidades). Se debe calcular la cantidad de productos a producir en base al presupuesto de ingresos y a las políticas de inventarios
3. Presupuesto de consumo de materiales directos y presupuesto de compras de materiales directos. En base a la cantidad que se desea producir, se debe calcular el consumo de materiales directos en cantidades y en dólares.

4. Presupuesto del costo de la mano de obra directa. Corresponde a la parte variable del costo.
5. Presupuesto de costos indirectos de manufactura. Aquí se identifican costos de suministros, mano de obra indirecta, energía, mantenimiento, depreciaciones, supervisión, que normalmente son clasificados dentro de los grupos de costos fijos y costos variables.
6. Presupuesto de inventarios finales.
7. Presupuesto del costo de los bienes vendidos.
8. Presupuesto de costos que no son de manufactura. Como son los costos asociados a diseño del producto, marketing y distribución. También es recomendable catalogarlos como costos fijos y variables para luego realizar un análisis de punto de equilibrio y otros.
9. Preparación del estado de resultados presupuestado. Usando los reportes anteriores, se procede a la elaboración del estado de resultados presupuestado. (Horngren et al. 2012)

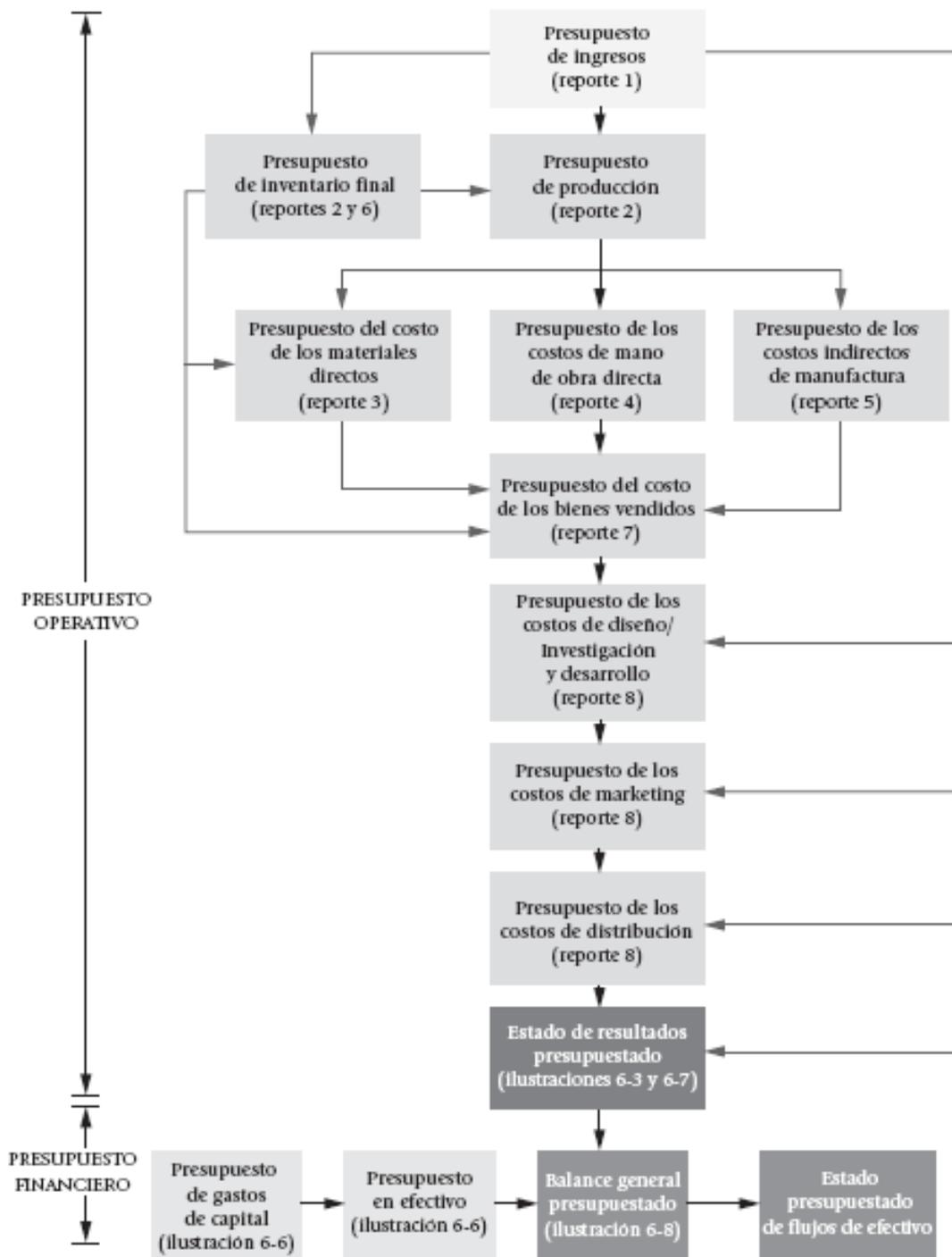


Figura 13 Presupuesto maestro para una empresa de producción. (Horngren et al. 2012)

Welch en su libro *Presupuestos*, propone a su vez un proceso de presupuestación que comprende los siguientes elementos (Welsch y Vázquez Prada Gutiérrez 2005):

1. Objetivos generales
2. Metas específicas
3. Memorando de las premisas de planificación
4. Resumen de propuesta de proyecto

5. Plan estratégico de utilidades de largo alcance
6. Plan táctico de utilidades de corto plazo

Presupuestos

7. Producción
8. Materiales directos
9. Compras
10. Mano de obra directa
11. Servicios del edificio
12. Gastos indirectos de fabricación
13. Inventarios
14. Costo de lo vendido
15. Gastos de distribución
16. Gastos de administración
17. Otros ingresos y gastos
18. Adiciones de capital

Estados financieros planificados

19. Estado de resultados
20. Estado de utilidades retenidas
21. Estado de flujo de efectivo
22. Balance general

2.2. Modelos determinísticos de pronóstico aplicables en la elaboración de presupuestos

El pronóstico es el acto de predecir valores futuros de determinadas variables, utilizando como base los datos históricos o las especulaciones sobre el futuro en el caso de no contar con los datos históricos (Mun 2010).

Shiavi define como variables deterministas a aquellas cuyos valores futuros pueden determinarse con certeza si se conocen su valor actual y algunos parámetros (Shiavi 2007), de modo que existen también para estas variables, modelos matemáticos deterministas, que son aquellos modelos compuestos por ecuaciones que nos permiten determinar el valor de la variable dependiente a partir de los valores de las variables independientes (Keller 2014).

En la tarea de elaborar el presupuesto, es necesario pronosticar, esto es, determinar los comportamientos futuros de ciertas variables que afectan los ingresos y los egresos monetarios en la organización, variables que a su vez conforman un marco de referencia sobre el que se toma las decisiones de planificación; por ejemplo, se podría planificar vender una cantidad arbitrariamente grande de productos, pero esta decisión no sería adecuada a menos que esté delimitada por la evolución de la demanda y oferta del producto.

El objetivo de los pronósticos en los negocios es, basándose en el análisis de datos actuales como históricos, estimar valores futuros de demanda, ventas, rotación de inventarios, requerimientos de activos etc. Es necesario además analizar la posición competitiva de la organización dentro de la industria, así como la posición de la industria dentro del mercado global; esto es cierto aún para aquellas organizaciones sin fines de lucro, en donde se deben analizar los costos presupuestados. El resultado será un modelo, que es la representación de las relaciones observadas entre dos o más variables (Nugus 2009).

Un pronóstico busca combinar los análisis estadísticos y el conocimiento del dominio para desarrollar previsiones aceptables que guiarán las actividades de planificación de negocios en sentido descendente. Se deben definir y medir los indicadores claves del negocio que afectan las ventas y los ingresos, y ofrecer a la alta gerencia opciones de decisión de negocios que estén basadas en un buen análisis estadístico que permita reducir las incertidumbres del futuro (Hoshmand 2010).

En la actualidad con el avance de las tecnologías de la información y el acelerado ritmo de desarrollo de las economías de mercado, se tienen modelos matemáticos muy complicados que están disponibles como funciones de caja negra, en donde lo importante no implica el conocimiento del funcionamiento del modelo, sino más bien en comprender tanto las variables de entrada como los resultados; aun así esto implica un alto grado de preparación académica para quienes diseñan o interpretan los resultados, por lo que todavía se utilizan los métodos tradicionales de pronóstico para la elaboración de presupuestos. Estos métodos tradicionales incluyen la planificación basada en estándares especificados por la administración o por la dinámica cambiante de períodos anteriores (Mutanov 2015).

2.2.1. Enfoques de los pronósticos

Un pronóstico difícilmente será exacto debido a la gran cantidad de elementos incontrolables que pueden suceder en el futuro, por ello se lo considera como el arte y la ciencia de predecir los eventos futuros. La tarea de pronosticar puede ser una

tarea subjetiva en donde la experiencia e intuición del pronosticador o consultor juega un rol importante, o puede ser una tarea objetiva, en donde se diseñen modelos matemáticos que basados en información histórica, permitan proyectar los valores hacia el futuro. Normalmente el proceso en general a pesar de estar basado en modelos matemáticos, no deja de tener componentes subjetivos pues desde la selección de las variables del modelo hasta la valoración final de resultados, deben ser ajustados mediante el buen juicio del administrador. (Heizer et al. 2009)

Los enfoques de los pronósticos pueden clasificarse de modo general como cuantitativos o estadísticos, que son recomendables cuando existen datos históricos comparables, y los enfoques cualitativos o de criterio, para cuando no existen datos históricos.

En el caso de las predicciones cuantitativas, se pueden dividir a su vez en series de tiempo (tienen un elemento de unidades de tiempo), transversales (normalmente ocurren en un mismo momento), o mixtos.

Adicionalmente Nugus considera el enfoque gráfico, ya que éste es particularmente útil cuando es necesario identificar problemas potenciales tan pronto como sea posible, permitiendo visualizar desviaciones e identificar cuándo un método de pronóstico requiere un ajuste (Nugus 2009)

Se consideran a los métodos objetivos más confiables que los subjetivos, debido a que serían menos propensos a que los resultados reflejen los deseos particulares de quien realiza el pronóstico, y que además permite mejorar la precisión mediante la comparación de los resultados obtenidos del modelo con los resultados reales, información que sirve de retroalimentación para ajustar al modelo para el futuro. (Nugus 2009)

2.2.2. Horizontes de tiempo del pronóstico

De acuerdo con Heizer (Heizer et al. 2009), un pronóstico de acuerdo al tiempo futuro que cubre, se clasifica en:

- a. **Pronóstico a corto plazo:** Cubre como máximo un año, aunque lo más común es que sea menor a 3 meses. Sirve para aspectos puntuales de la organización, como es el caso de decidir niveles de producción, programar y asignar trabajos, y planificar compras. En este tipo de pronóstico se utilizan técnicas como promedios móviles, suavizado exponencial y extrapolación de tendencias, dando como resultado pronósticos de mayor precisión que los de largo plazo.

- b. **Pronóstico a mediano plazo:** Normalmente cubre un período comprendido entre 3 meses y 3 años. Sirve para aspectos más generales de la empresa, tales como planear las ventas, la producción, el presupuesto, el flujo de efectivo y como herramienta de análisis de diferentes planes operativos.
- c. **Pronóstico a largo plazo.** Por lo general se proyecta a 3 años o más. Sirven para planear la fabricación de nuevos productos o desarrollo de nuevos servicios, planear gastos de capital, analizar decisiones de ubicación o expansión de las instalaciones, y para planes de investigación y desarrollo.

2.3. Proceso de Pronóstico

Para elaborar pronósticos precisos de manera consistente, se debe utilizar un procedimiento sistemático para pronosticar, procedimiento que pueda ser aplicado rápidamente y modificado según sea necesario. Hoshmand propone un modelo de proceso de pronóstico que cumple tales características y que parte de las preguntas planteadas por la administración en el proceso de planificación, como se puede ver en la figura 14.

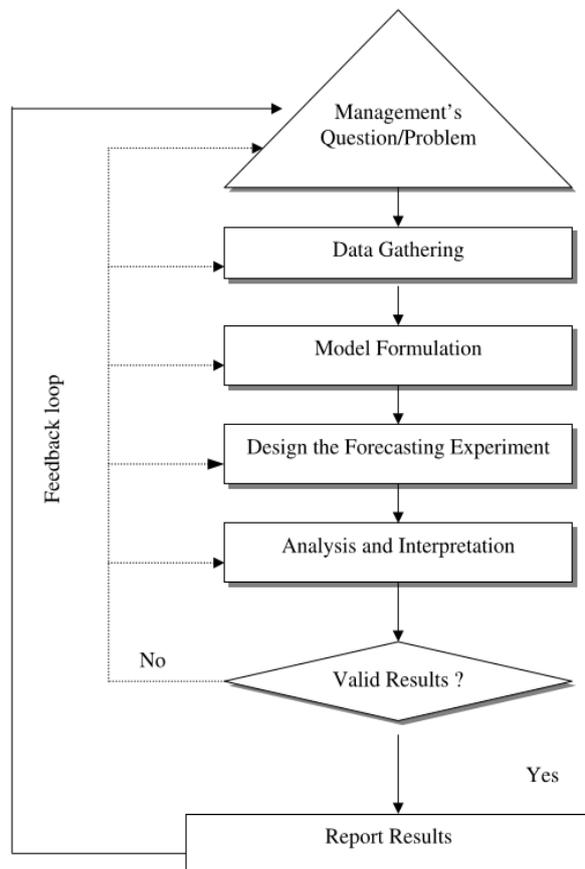


Figura 14 Proceso de pronóstico (Hoshmand 2010).

Las preguntas que se plantea la administración sobre el comportamiento futuro de una variable, deben justificar el esfuerzo requerido de pronosticar, por lo que es necesario preguntarse para qué se necesitan los pronósticos y cómo se utilizarán los resultados. Los pronósticos pueden ser de corto plazo para temas de ventas o producción, y pueden ser de largo plazo para casos de presupuestos de capital. El modelo teórico que se elija para realizar el pronóstico dependerá de las necesidades temporales de la información, de tal manera que, si se requieren pronósticos de corto plazo, se pueden utilizar promedios móviles o suavizado exponencial, mientras que para el largo plazo se puede elegir un análisis de tendencia o un modelo de regresión.

El siguiente paso es recolectar datos para el análisis, datos que pueden ser internos o externos a la empresa. Es importante tener cuidado con el tipo y calidad de los datos que se están recolectando.

Antes de estimar el modelo de pronóstico, es necesario familiarizarse con la naturaleza de la empresa y la industria que representa. Para ello se puede llevar a cabo un análisis del entorno y un análisis de la industria utilizando el modelo de las 5 fuerzas de Porter.

El diseño del modelo implica la realización de pruebas para medir su confiabilidad y validez, que se realiza mediante la comparación de los resultados de las predicciones con los resultados reales, lo que puede llevar a la adición o eliminación de variables, es decir, a un proceso de ajuste del modelo.

El paso final es la presentación de los resultados a la dirección, pero eso no implica que el trabajo haya terminado, será necesario continuar con el proceso de refinación del modelo de acuerdo a los cambios en los supuestos que se presenten en la realidad. Las predicciones deben ser por tanto dinámicas y no estáticas.

Heizer mientras tanto propone una metodología de siete pasos para la realización de pronósticos, los que incluyen:

1. Determinar el uso del pronóstico.
2. Seleccionar los elementos que se deben pronosticar.
3. Fijar el horizonte de tiempo para el pronóstico.
4. Seleccionar las técnicas de pronóstico.
5. Recolectar los datos que se necesitan para realizar el pronóstico.
6. Ejecutar el pronóstico.

7. Verificar la validez de lo pronosticado e implementar los resultados

Las dos propuestas de pasos para realizar pronósticos comparten elementos comunes, que en general implican:

- Definir los objetivos e importancia de realizar el pronóstico.
- Elegir el modelo a aplicar de acuerdo al tipo de variable a pronosticar.
- Recopilar los datos.
- Aplicar y refinar los resultados.
- Validar y comunicar los resultados.

2.4. Patrones y técnicas de proyección de datos

Los modelos de series de tiempo basan su funcionalidad en la suposición de que el futuro es una función del pasado. Es decir, a partir de lo que ha ocurrido en un periodo, se usan datos históricos para hacer un pronóstico (Render et al. 2012).

Una serie de tiempo consiste en una sucesión de datos con periodos de tiempo igualmente espaciados (minutos, días, semanas, etc.).

2.4.1. Patrones de datos y elección de técnicas de pronóstico

El proceso de crear un modelo determinístico implica el conocimiento general de las características de las variables, lo que permitirá elegir un modelo adecuado que más se ajuste al comportamiento de las variables de interés. En el caso de no tener tal conocimiento, será necesario aplicar otras técnicas como la minería de datos, lo que se analizará en la sección de nuevos métodos de pronóstico aplicables a la administración.

En cuanto a las características del comportamiento de las variables, es de especial importancia identificar el patrón de datos, la naturaleza de las relaciones pasadas y el nivel de subjetividad. Las técnicas de pronóstico univariadas se basan en patrones del pasado, mientras que el pronóstico multivariado predice los valores futuros a partir de las relaciones en el pasado. Por otro lado, los pronósticos cualitativos, dependen principalmente de la intuición del pronosticador.

Patrones de datos

Para pronosticar en base a patrones de datos, se parte de la suposición de que el comportamiento observado en el pasado se repetirá en el futuro y la técnica más adecuada será aquella que mejor se adapte a dicho patrón.

Los patrones más comunes según Hoshmand (Hoshmand 2010) son: horizontal, tendencia, estacional, cíclico (como se puede ver en las figuras 15, 68 y 17), y adicionalmente el aleatorio mostrado en la figura 18.

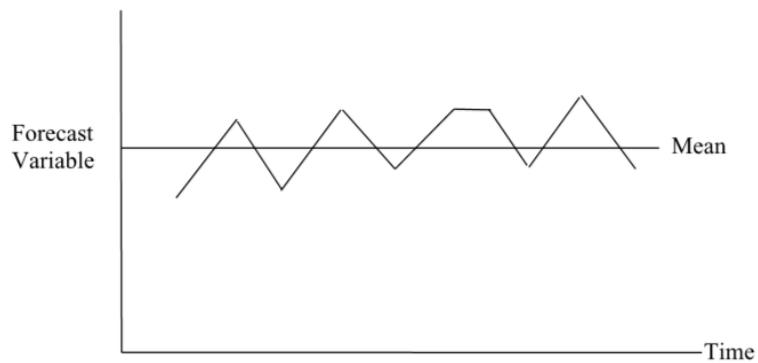


Figura 15 Patrón de datos horizontal (Hoshmand 2010)

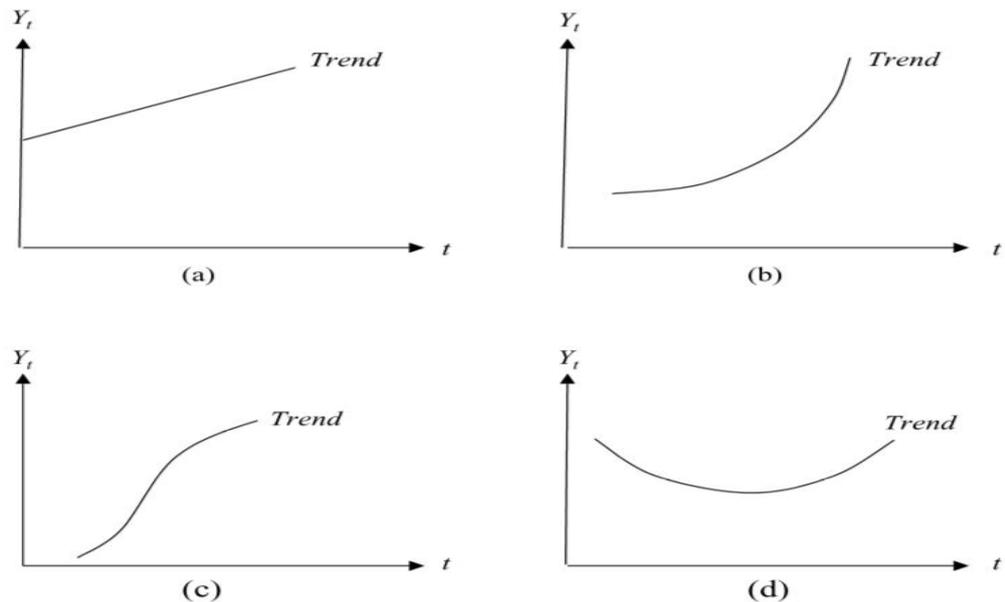


Figura 16 Curvas de tendencia determinística (Hoshmand 2010)

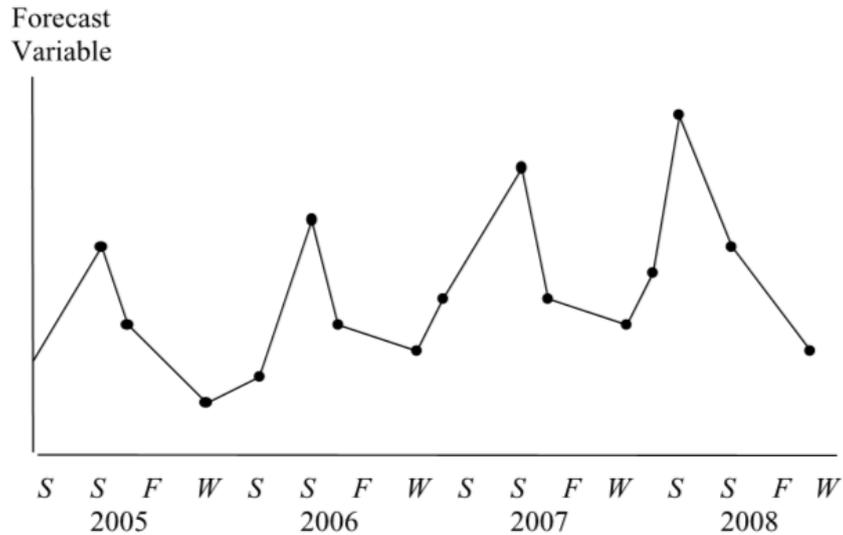


Figura 17 Patrón de datos estacional (Hoshmand 2010).

Heizer describe estos distintos patrones de la siguiente manera:

La tendencia es el movimiento de los datos en el tiempo que gradualmente se dirige hacia arriba o hacia abajo como se puede ver en la figura 16. Las posibles causas se relacionan con cambios en el ingreso, la población o la distribución de edades.

La estacionalidad se refiere a la existencia de un patrón de datos que se repite tras un periodo temporal de días, semanas, meses o trimestres, observándose que en un año son comunes 6 diferentes patrones, los que se resumen en la tabla 1.

En la figura 17 se puede observar un ejemplo de patrón estacional.

Periodo del patrón	Longitud de la "estación"	Número de "estaciones" en el patrón
Semana	Día	7
Mes	Semana	$4-4\frac{1}{2}$
Mes	Día	28-31
Año	Trimestre	4
Año	Mes	12
Año	Semana	52

Tabla 1 Patrones de estacionalidad

Los ciclos son comportamientos que reflejan patrones que se repiten cada cierto período extenso de tiempo, generalmente años, y están asociados a ciclos

económicos que son influenciados a su vez por hechos políticos a nivel internacional, que en consecuencia dificultan la tarea de predicción.

Las variaciones aleatorias son disturbios que se presentan en los datos, cuyo origen puede ser la casualidad o por situaciones inusuales, lo que no hace posible la predicción. La figura 18 muestra una gráfica con tendencia creciente y estacionalidad, alterados por las variaciones aleatorias.

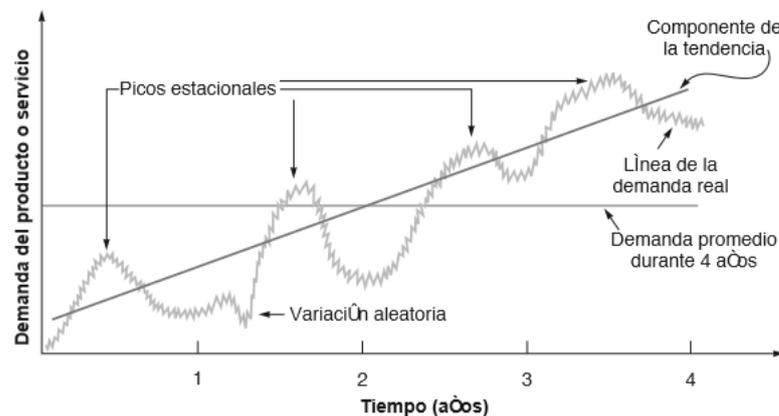


Figura 18 Gráfica de demanda con tendencia creciente y estacionalidad. (Heizer et al. 2009)

2.4.2. Precisión y fiabilidad

Cada pronóstico debe incluir una estimación de su exactitud y una evaluación de su fiabilidad. Esto no sólo permite un margen de error razonable y aceptable, sino que también ayuda a generar confianza en los métodos que se utilizan y en los resultados obtenidos (Nugus 2009).

Error de pronóstico

El error de pronóstico o desviación mide la exactitud de un modelo de pronóstico comparando los valores pronosticados con los reales y obteniendo su diferencia.

$$\text{Error de pronóstico} = \text{Valor real} - \text{Valor pronosticado}$$

Las principales técnicas de medición del error de pronóstico son la desviación absoluta media MAD (mean absolute deviation), el error cuadrático medio MSE (mean squared error), y el error porcentual absoluto medio MAPE (mean absolute percent error).

La desviación absoluta media se calcula sumando los valores absolutos de los errores individuales del pronóstico y obteniendo su promedio

$$MAD = \frac{\sum |Real - Pronóstico|}{n}$$

El error cuadrático medio es el promedio de los cuadrados de las diferencias entre los valores pronosticados y los reales. Por ser elevado al cuadrado, acentúa las desviaciones importantes, por lo que su medida lleva a una mejor calificación a modelos con varias desviaciones pequeñas

$$MSE = \frac{\sum (Errores de pronóstico)^2}{n}$$

El error porcentual absoluto medio es el promedio de las diferencias absolutas encontradas entre los valores pronosticados y los reales. Al ser una medida porcentual, permite su comparación a pesar de utilizar diferentes unidades.

$$MAPE = \frac{\sum_{i=1}^n 100 \frac{|Real_i - Pronóstico_i|}{Real_i}}{n}$$

2.5. Modelos de pronóstico basados en series de tiempo

Heizer clasifica como los principales modelos de series de tiempo a los siguientes:

1. Enfoque intuitivo
2. Promedios móviles
3. Suavizado exponencial
4. Proyección de tendencias

2.5.1. Enfoque intuitivo

Supone que la cantidad demanda en el siguiente periodo será la misma que la cantidad demanda en el periodo anterior.

2.5.2. Promedios móviles

Considera que la demanda del mercado permanecerá estable y usa el promedio de los n valores de datos históricos para generar un pronóstico del siguiente período.

$$Promedio\ móvil = \frac{\sum Demanda\ de\ los\ n\ periodos\ previos}{n}$$

Cuando se considera una tendencia, es posible ponderar cada período mediante un peso diferente usando la siguiente fórmula

Promedio móvil ponderado

$$= \frac{\sum[(Ponderación \text{ de cada período } n)(Demanda \text{ de cada período } n)]}{\text{suma de las ponderaciones}}$$

El valor de ponderación es un entero mayor o igual a 1.

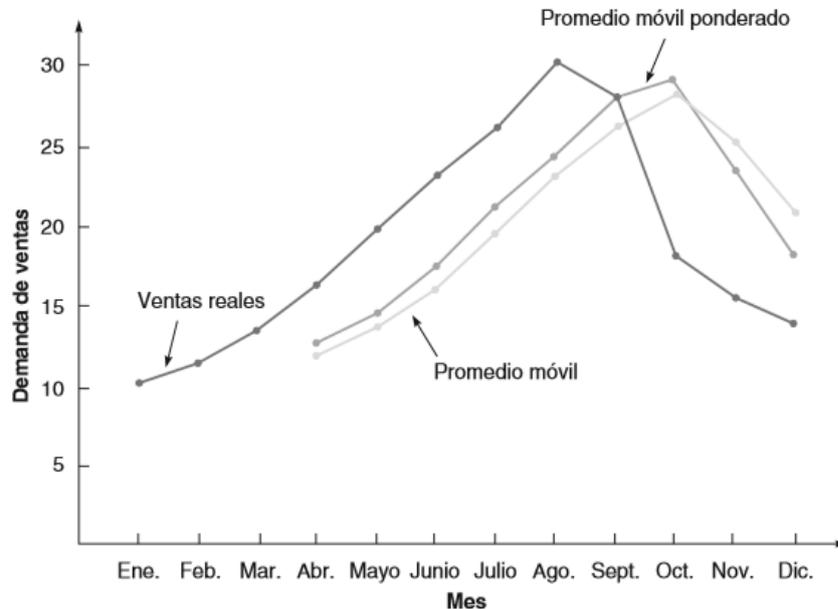


Figura 19 Diferencia ente ventas reales y proyecciones mediante promedios móviles. (Heizer et al. 2009)

2.5.3. Suavizado exponencial

Los datos de la proyección del período anterior se ajustan con una constante de ponderación α que multiplica a la diferencia entre la cantidad de demanda real y la demanda pronosticada anteriormente. Se utiliza cuando se tienen pocos registros de datos históricos.

$$\begin{aligned} \text{Nuevo pronóstico} &= \text{Pronóstico mes anterior} + \\ &\quad \alpha (\text{Demanda real mes anterior} \\ &\quad - \text{Demanda pronosticada mes anterior}) \end{aligned}$$

La constante α se denomina constante de suavizado y puede tener un valor entre 0 y 1, que el pronosticador elegirá de acuerdo a su intuición, aunque por lo general suele asignársele valores de entre 0.05 a 0.50. Los valores bajos de alfa implican dar mucho peso a los datos anteriores, al esperarse un comportamiento estable en la tendencia, o un alto peso a los valores recientes con un alfa elevado para casos en los que se espera que los valores puedan cambiar drásticamente.

2.5.4. Suavizado exponencial con ajuste de tendencia

El suavizado exponencial simple es una técnica débil para tratar las tendencias, pero es utilizada con mucha frecuencia en los negocios y consiste en calcular un promedio suavizado exponencialmente de los datos para después ajustar la desviación encontrada en la tendencia. Las fórmulas son:

Pronóstico incluyendo la tendencia (FIT_t)

= Pronóstico suavizado exponencialmente (F_t)

+ tendencia suavizada exponencialmente (T_t)

$$F_t = \alpha(A_{t-1}) + (1 - \alpha)(F_{t-1} + T_{t-1})$$

$$T_t = \beta(F_t - F_{t-1}) + (1 - \beta)T_{t-1}$$

En donde:

F_t = pronóstico suavizado exponencialmente de la serie de datos en el periodo t .

T_t = tendencia suavizada exponencialmente en el periodo t

A_t = demanda real en el periodo t

α = constante de suavizamiento para el promedio ($0 \leq \alpha \leq 1$)

β = constante de suavizamiento para la tendencia ($0 \leq \beta \leq 1$)

Para comenzar a calcular, necesitamos definir estimaciones iniciales del pronóstico y tendencia del período anterior ($t-1$) que permiten calcular F_t el que a su vez posibilita calcular T_t y con éstos, se determina el pronóstico incluido tendencia FIT_t

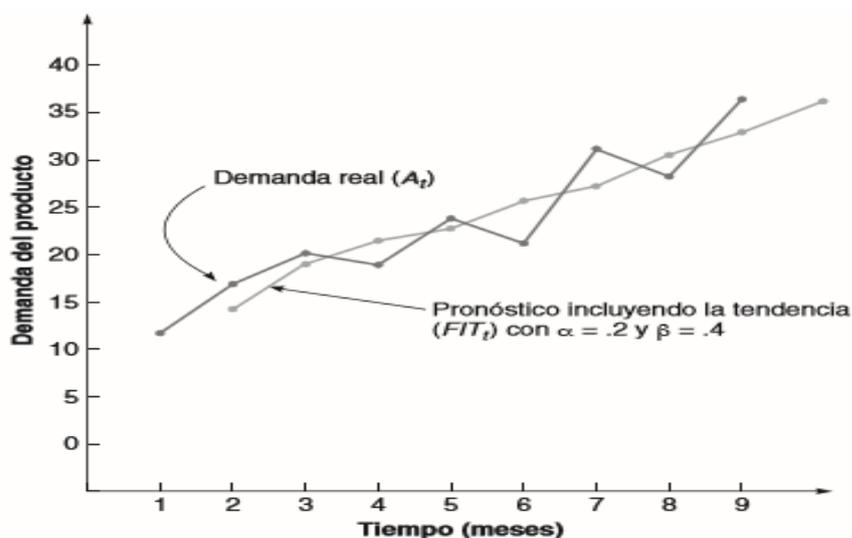


Figura 20 pronóstico de suavizado exponencial y ajuste de tendencia. (Heizer et al. 2009)

La constante de suavizado β alta implica una respuesta más rápida a los cambios recientes de la tendencia, mientras que un valor bajo suaviza la tendencia. De forma similar, un valor alto de α implica una valoración alta de la observación real del período anterior en contra de una valoración baja del pronóstico y tendencias del período pasado. La figura 20 muestra como este tipo de suavizado se ajusta bastante bien a los datos reales.

2.5.5. Proyecciones de tendencia

Consiste en ajustar una curva o recta de tendencia a una serie de datos históricos para luego proyectar esta recta hacia el futuro y con ello pronosticar valores a mediano y largo plazo. Uno de los métodos más usados es el lineal basado en mínimos cuadrados, aunque es posible también calcular curvas exponenciales, cuadráticas o logarítmicas.

Método de mínimos cuadrados. Este método busca la línea recta que minimice la suma de los cuadrados de las desviaciones de la recta y cada una de las observaciones reales. En la figura 21 se visualiza los elementos que sirven de base para el cálculo de la línea de tendencia que tiene la forma siguiente.

$$\hat{y} = a + bx$$

En donde:

\hat{y} = valor pronosticado con el modelo (y hat o y gorro)

a = cruce de la línea de tendencia con el eje y

b = tasa de cambio de la tendencia

x = variable dependiente (período)

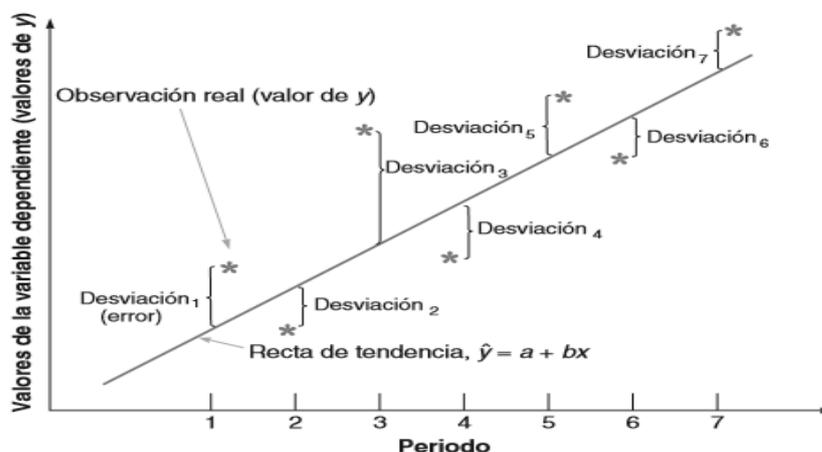


Figura 21 Recta de tendencia calculada mediante mínimos cuadrados. (Heizer et al. 2009)

Es posible calcular de forma manual tanto la pendiente como el cruce, pero como veremos más adelante, se puede utilizar las funciones integradas de R para determinarlas directamente. Lo importante es tener claro el significado de cada componente de la fórmula.

Este método no se recomienda para predecir períodos demasiado lejanos, debido a que la realidad es muy cambiante y debemos tener presente que la relación lineal puede en la realidad no representar de la mejor manera el comportamiento de los datos, por lo cual es necesario verificar la validez del modelo mediante mediciones de error y significancia.

2.5.6. Variaciones estacionales en los datos

Las variaciones estacionales implican movimientos regulares en los datos que se deben a acontecimientos recurrentes como el clima o la temporada.

Como se vio anteriormente, existen 6 períodos comunes que se muestran en la tabla 1, que permiten anticipar y planificar ya sea la producción o las ventas cada vez que haya picos o valles. Las variaciones estacionales se pueden observar en la cantidad en que difieren los valores reales de los valores promedios.

El pronóstico estacional se determina calculando la demanda promedio y multiplicándolo por los factores estacionales.

Si se determinara que existen estaciones mensuales, Heizer recomienda aplicar los siguientes pasos:

1. Encontrar la demanda histórica promedio de cada estación.
2. Calcular la demanda promedio de todos los meses del último año.
3. Calcular un índice estacional para cada estación dividiendo la demanda histórica real de ese mes entre la demanda promedio de todos los meses.
4. Estimar la demanda total anual para el siguiente año.
5. Dividir esta estimación de la demanda total anual entre el número de estaciones y multiplicarla por el índice estacional para ese mes(Heizer et al. 2009).

2.6. Regresión lineal simple y correlación

Dentro de los modelos matemáticos utilizados para la predicción, se encuentran los denominados modelos asociativos, en el que el más utilizado es la regresión lineal.

Estos modelos se denominan así porque incorporan al modelo variables que pueden influir en el valor a pronosticar (Heizer et al. 2009).

2.6.1. Análisis de regresión simple

El análisis de regresión simple es una técnica muy utilizada en la actualidad y busca predecir el valor de una variable basándonos en la relación que tiene con otra variable. El modelo más simple que sirve para relacionar estas variables es el de una recta, por lo que la relación se la denomina relación lineal, y la ecuación que la representa, se denomina *Ecuación de regresión*, y consiste en diseñar un modelo matemático que cuantifique la relación entre la variable dependiente (“y” a predecir), y las variables que se cree están relacionadas con aquella, las que se denominan variables independientes (“x” predictoras), y con el modelo resultante, se puede pronosticar un valor futuro de Y para un determinado valor de X (Hoshmand 2010).

La relación lineal se expresa mediante la siguiente ecuación general de la recta:

$$Y = a + bX$$

Siendo:

Y= variable dependiente, cuyo valor se desea predecir

X= variable independiente que incide directamente en el valor de Y

a= Cruce en Y cuando X es cero

b= coeficiente de regresión

Si bien su principal beneficio es la simplicidad del modelo, es un inconveniente el que en la realidad las relaciones no son lineales, sin embargo, mediante transformaciones de los datos, es posible tener una predicción razonable. Debido además a que existen varios hechos no posibles de capturar en el modelo y otros hechos al azar, un modelo más adecuado es aquel que considere un error en la estimación, error que se representa como ε_t y tiene un valor medio de cero:

$$Y_t = a + bX_t + \varepsilon_t$$

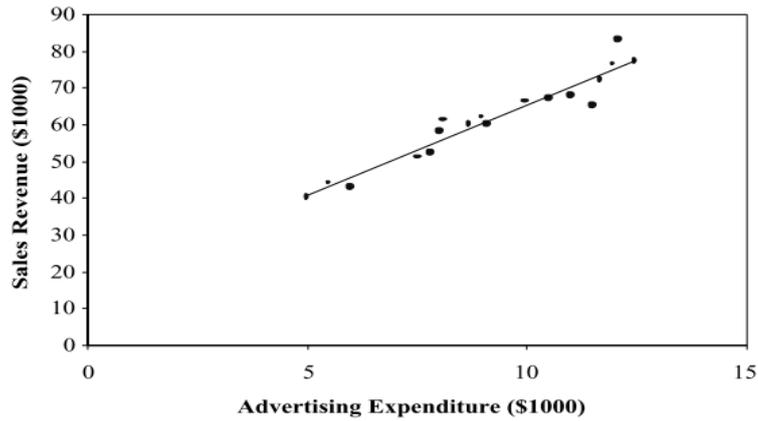


Figura 22 Diagrama de dispersión de ingresos de ventas y gastos de publicidad (Hoshmand 2010). La figura 22 muestra un ejemplo de la relación directa entre el nivel de publicidad y las ventas, en donde se observa que, a mayor gasto en publicidad, mayores serán los ingresos.

2.6.2. Error estándar

Para determinar si la estimación es confiable utilizamos el concepto de error estándar de estimación, que es una medida de la dispersión de los puntos sobre la línea de regresión, de tal manera que es deseable el menor valor posible, lo que implica que el punto real estará más cerca del valor estimado, siendo que este último cae en la línea. La siguiente fórmula permite determinar su valor.

$$s_{y,x} = \sqrt{\frac{\sum(Y - \hat{Y})^2}{n - 2}}$$

En donde:

$s_{y,x}$ = error estándar

Y= variable dependiente

\hat{Y} = valor estimado de la variable dependiente

n= tamaño de la muestra

2.6.3. Análisis de correlación

El análisis de regresión se basa en la relación existente entre dos variables, para poder verificar el grado y fuerza de asociación que existe entre ellas, se utiliza la medida de coeficiente de determinación y el coeficiente de correlación.

El coeficiente de determinación (r^2) indica lo bien que la línea se ajusta a los datos observados, para ello es necesario identificar las desviaciones existentes entre la media de las observaciones y tanto el valor pronosticado como el valor observado. Esto se muestra en la Figura 23.

La distancia vertical entre la línea de regresión y el promedio es la desviación explicada, mientras que la distancia vertical del valor observado y la línea de regresión es la desviación inexplorada. La suma de estas dos desviaciones se denomina desviación total.

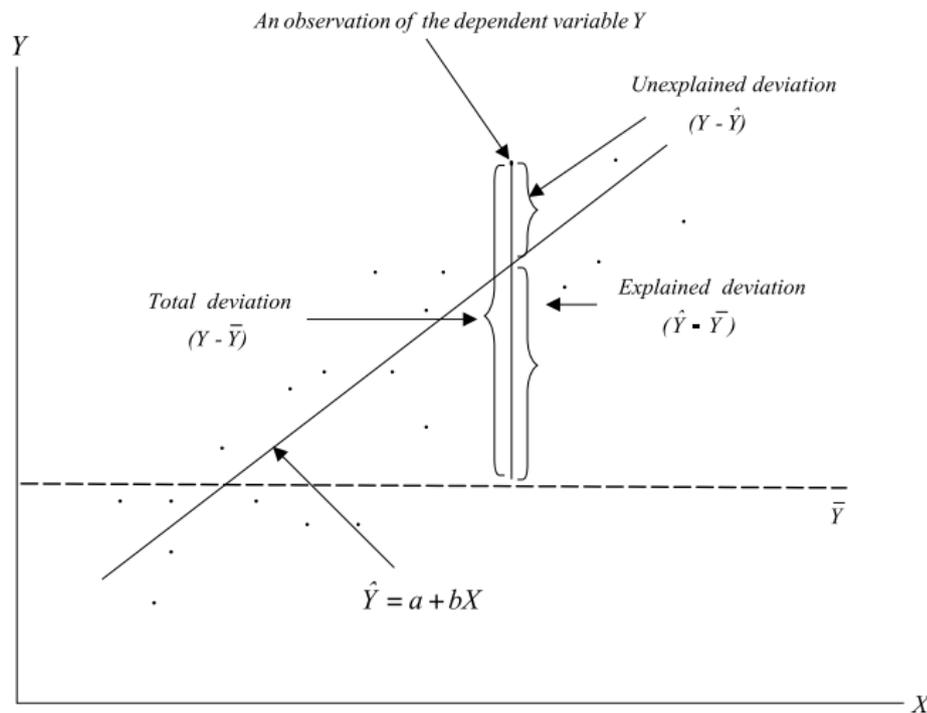


Figura 23 Desvío total, explicada e inexplorada para un valor observado de Y (Hoshmand 2010).

La fórmula del coeficiente de determinación relaciona las desviaciones en la siguiente manera:

$$r^2 = \frac{\sum(\hat{Y} - \bar{Y})^2}{\sum(Y - \bar{Y})^2}$$

El valor resultante es un indicador del porcentaje de variación total que se explica por el modelo, por lo que mientras más alto, mejor.

El otro parámetro que mide la fuerza de la relación lineal entre dos variables es el coeficiente de correlación (r), que es la raíz cuadrada del coeficiente de determinación.

$$r = \sqrt{r^2}$$

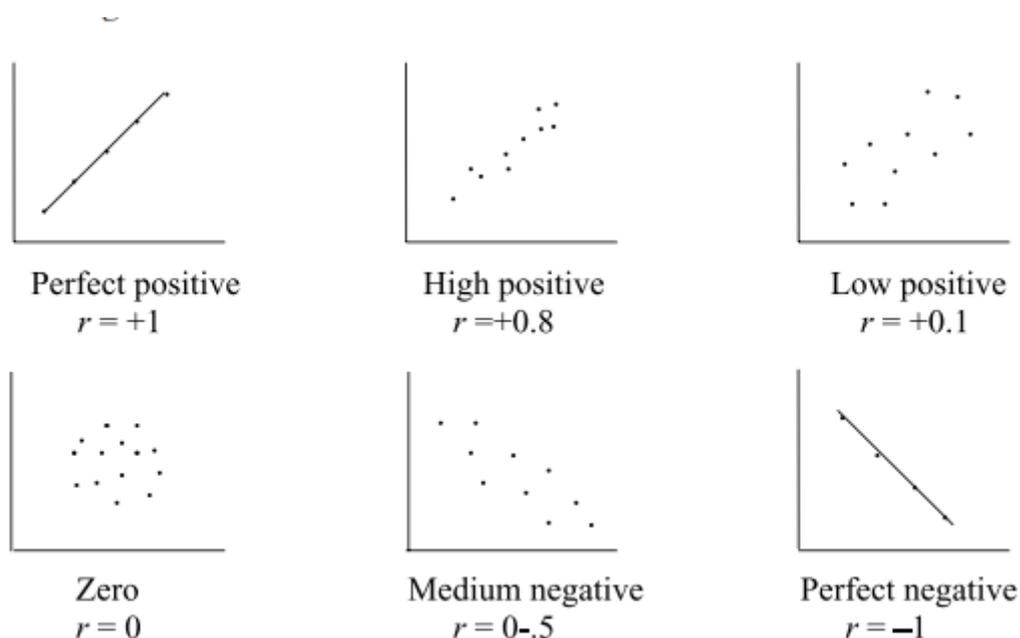


Figura 24 Diferentes diagramas de dispersión para diversos valores de dispersión (Hoshmand 2010).

La figura 24 muestra distintos diagramas de dispersión y sus correspondientes valores de correlaciones, en donde se observa que un valor de uno implica una correlación positiva perfecta, es decir, por cada incremento de X, se tendrá una proporción constante de incremento en Y, y los puntos de la observación caerán sobre la línea de tendencia. Otros posibles valores inferiores muestran una menor dependencia entre las variables, siendo que cuando no hay ninguna relación, el valor de r será cero.

Se recomienda usar el coeficiente de determinación como elemento indicador de la calidad de la relación, puesto que como se indicó con anterioridad, éste representa cuanto del resultado se explica por el modelo (Hoshmand 2010).

Capítulo 3: Nuevas tendencias de tratamiento de datos para uso en administración empresarial

Introducción

Si bien la mayoría de los gerentes desean mejorar la predicción, pocos están familiarizados con las técnicas que se han desarrollado en las últimas décadas, como es el caso del análisis de regresión, las técnicas como Box-Jenkins y las redes neuronales, las que en conjunto han ampliado el campo de la predicción (Hoshmand 2010).

Por lo descrito en el párrafo anterior, en esta sección se describirán algunas técnicas de actualidad recomendables sobre todo para el tratamiento de grandes cantidades de datos (Big Data). Puesto que la mayoría de organizaciones cuentan en la actualidad con un sistema informático que mantiene miles o millones de registros de movimientos, en donde las técnicas tradicionales son poco recomendables, se hace necesario el tratamiento automatizado de tales datos.

3.1. Administración del desempeño empresarial

3.1.1. Pronóstico guiado por la demanda

Chase (Chase 2009), define al pronóstico guiado por la demanda como un nuevo enfoque estructurado de predicción, que se basa en la analítica para detectar, dar forma y predecir la demanda, utilizando modelos matemáticos que aprovechan la gran cantidad de datos existentes en la actualidad gracias a los enormes avances en el almacenamiento de datos y la capacidad de cómputo.

Las empresas ahora pueden explotar los datos mediante el análisis predictivo y la minería de datos, que permiten descubrir patrones de comportamiento del consumidor para anticipar la demanda, medir la efectividad de sus estrategias de marketing y optimizar el desempeño financiero.

Adicionalmente, estas nuevas tecnologías de procesamiento de datos han facilitado el almacenamiento de datos sobre los factores causales que influyen en la demanda, como es el precio, la publicidad, las promociones de ventas, etc.

La importancia de la predicción de la demanda está en que ésta determina la cantidad de materia prima, productos terminados que se requieren enviar, personal a contratar y plantas a construir. Esto es de aún mayor importancia aún, si consideramos que

una organización de provisión de servicios básicos como el agua potable, debe garantizar el suministro que los usuarios requieren.

La predicción guiada por la demanda se basa en la aplicación de la analítica de datos en tres pasos que son: (1) detectar tendencias de la demanda mediante la sincronización de datos de la empresa y del entorno, (2) modelar la demanda usando técnicas analíticas avanzadas, y (3) predecir la demanda, para obtener modelos de pronóstico de mayor precisión.

Se debe tener presente que varios de los métodos cuantitativos en los que se basan las técnicas modernas, ya estaban disponibles desde inicios del siglo pasado, sólo que no había herramientas que faciliten su implementación.

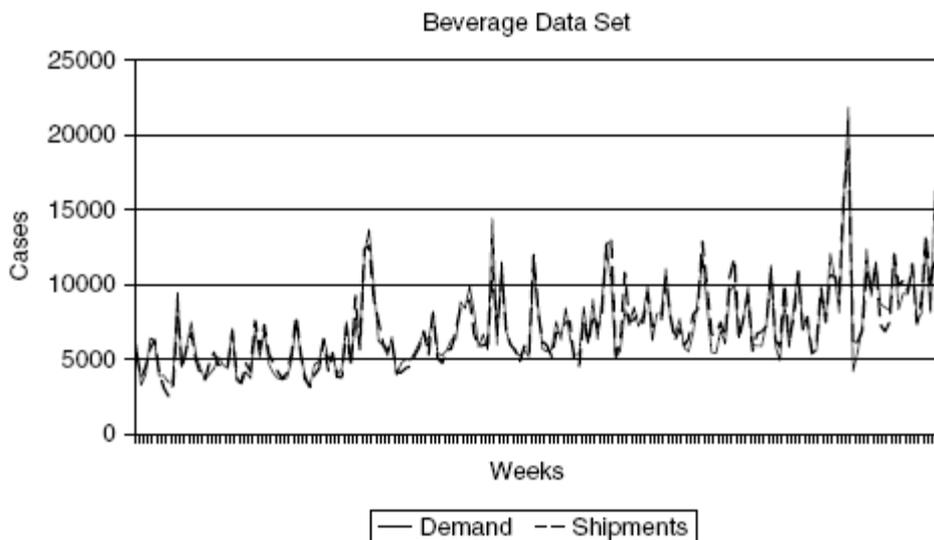


Figura 25 Comparación semanal de demanda (ventas) y envíos (suministro) de una empresa de bebidas (Chase 2009).

El ejemplo de la figura 25 muestra la comparación entre el despacho y la demanda de bebidas, en donde la empresa busca que sus predicciones sean lo más parecidas posible a la demanda real. (Chase 2009)

3.2. Big Data

En los últimos años y gracias a la masificación de los sistemas informáticos, se comienza a hablar de Big Data, que se refiere a la disponibilidad para la empresa de grandes volúmenes de datos.

Big Data es un campo dedicado al análisis, procesamiento y almacenamiento de grandes colecciones de datos que por lo general provienen de fuentes dispares. Se

utiliza cuando las tecnologías y técnicas tradicionales de almacenamiento, procesamiento y análisis de datos son insuficientes como para lograr un resultado adecuado tanto en tiempo como en utilidad. Big Data responde a requisitos distintos, como la combinación de múltiples conjuntos de datos no relacionados, el procesamiento de grandes cantidades de datos no estructurados y la recolección de información oculta de una manera sensible al tiempo. (Erl, Khattak, y Buhler 2016)

Para el tratamiento de datos, se ha utilizado tradicionalmente enfoques estadísticos basados en el muestreo, pero en la actualidad y debido al gran poder de cómputo con el que cualquier organización dispone, sin mencionar la posibilidad de utilizar capacidad de cómputo virtual provisto por la computación en la nube (Cloud Computing), es posible el procesamiento de conjuntos de datos completos.

El informe de McKinsey Global de 2011 indica que una empresa podría incrementar hasta en un 60% sus beneficios operacionales si logra explotar de forma creativa y efectiva las grandes cantidades de datos que tienen a su disposición. El mismo informe estima de muchos otros sectores de la sociedad pueden salir beneficiados con el uso efectivo de Big Data.

Definición de Big Data.

Big Data según el informe de McKinsey es:

“La información cuya escala, distribución, diversidad y / u oportunidad requieren el uso de nuevas arquitecturas técnicas y analíticas para permitir conocimientos que desbloqueen nuevas fuentes de valor comercial.”(Manyika, James et al. 2011)

3.2.1. Tipos de datos

Debido a la variedad de fuentes de datos que una organización puede tener, la información viene en una gran variedad, como son los datos de ventas, archivos de texto, informes en pdf, imágenes, archivos de video, etc. Estos tipos de datos van de los estructurados a los no estructurados.

Datos estructurados: Son aquellos datos que contienen un tipo de datos, un formato y una estructura claramente definidos. Estos datos pueden ser almacenados en bases de datos tradicionales y tratados con técnicas estándar.

Datos semi-estructurados: Son archivos de datos textuales con una estructura identificable, que permite el análisis sintáctico, como es el caso de los archivos codificados con XML o JSON.

Datos cuasi estructurados: Son datos textuales con estructuras erráticas, que, para obtener una estructura adecuada, requiere de mucho esfuerzo, herramientas y tiempo.

Datos no estructurados: Son aquellos datos que no tienen una estructura inherente, siendo algunos casos de estos, los documentos de texto, los archivos PDF, las imágenes y los videos.(Dietrich et al. 2015)

3.2.2. Términos importantes

En las siguientes secciones trataremos temas de actualidad relacionadas con el tratamiento de datos y la obtención de información, donde es necesario tener claro el concepto de un grupo de términos comunes en el mundo de los negocios y Big Data, los que se explicación a continuación:

Conjuntos de datos (Dataset): Las colecciones de datos relacionados donde cada miembro del conjunto tiene las mismas propiedades que los demás, se denominan generalmente conjuntos de datos (Dataset)

Análisis de datos: Proceso de examinar datos para encontrar hechos, relaciones, patrones, ideas y / o tendencias con el objetivo de apoyar para una mejor toma de decisiones.

Metadatos: Son datos sobre las características y la estructura de un conjunto de datos.

3.2.3. Características de Big Data

Erl (Erl, Khattak, y Buhler 2016) establecen que para llamarse Big Data, los datos deben tener alguna de características conocidas como 3V, esto es volumen, variedad y velocidad:

Gran volumen de datos: Cuando es posible tener miles de millones de filas y millones de columnas.

Complejidad de tipos de datos y estructuras: Debido a la variedad de nuevas fuentes de datos, además de bases de datos, se tiene archivos, fotografías, videos, etc., que necesitan ser analizados.

Velocidad de creación y crecimiento de nuevos datos: Los datos se producen a alta velocidad, requiriendo grandes bases de datos y procesos de análisis en tiempo real.

Dietrich propone además un par de características extra que son veracidad y valor como se grafica en la figura 26.

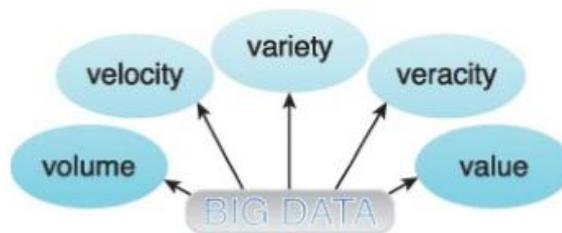


Figura 26 Cinco V de Big Data (Dietrich et al. 2015).

Veracidad: Referido al grado en que los datos reflejen fielmente una realidad, siendo para ello necesario un proceso de limpieza ya que muchos datos pueden estar incompletos o duplicados, generando únicamente lo que se conoce como ruido. Los datos con una alta relación señal-ruido son considerados veraces.

Valor: Es la utilidad de los datos para una empresa, lo que involucra la veracidad y la oportunidad; mientras más altas, mejor será el valor. La oportunidad evidencia que el valor depende inversamente del tiempo y a menor tiempo, mayor será el valor de la información.

3.3. Analítica de Datos

La analítica de datos es una disciplina que va más allá que el análisis de datos, pues abarca además la gestión del ciclo de vida completo de los datos, esto es, la recolección, limpieza, organización, almacenamiento, análisis y gestión de datos. La analítica de datos incluye el desarrollo de métodos de análisis, técnicas científicas y herramientas automatizadas que pueden utilizarse incluso para el tratamiento de grandes volúmenes de datos en entornos de computación altamente distribuidos

El resultado de la analítica de datos es que posibilita la toma de decisiones guiada por los datos (data-driven), de tal manera que las decisiones se basen en datos fácticos y no solamente en la experiencia pasada o la intuición.

La analítica de datos se agrupa dentro de cuatro categorías según lo describe Erl en su libro *Fundamentos de Big Data*:

- Análisis descriptivo
- Analítica de diagnóstico
- Análisis predictivo
- Analítica prescriptiva

El valor de los resultados se incrementa de acuerdo al grado de analítica que se aplique, que a su vez también implica un incremento en la complejidad de realizar los análisis, como se observa en la figura 27 y se explica a continuación.

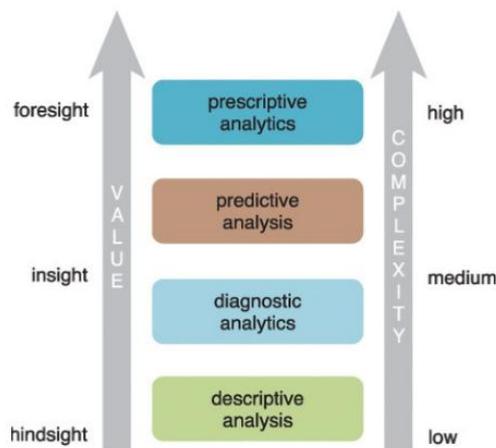


Figura 27 incremento de valor y complejidad de los distintos niveles de analítica (Erl, Khattak, y Buhler 2016).

3.3.1. Análisis descriptivo

Los análisis descriptivos sirven para responder preguntas sobre eventos pasados. El valor que proporciona a la empresa es muy reducido, pero tiene la ventaja de ser un proceso muy sencillo que requiere habilidades básicas de los analistas. La analítica descriptiva por lo general se realiza mediante informes puntuales o paneles de control estáticos que contienen matrices de datos y gráficos, como se muestra en la Figura 28.

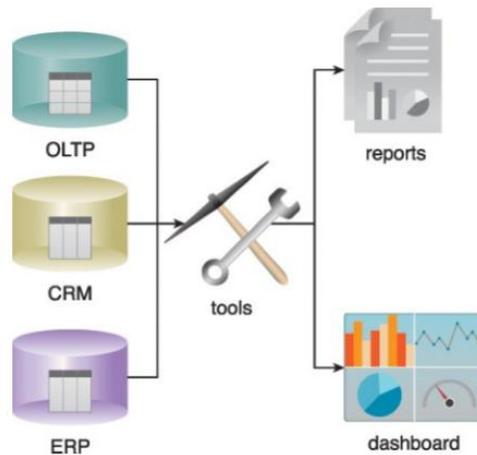


Figura 28 Generación de reportes a partir de análisis descriptivo de sistemas operacionales (Erl, Khattak, y Buhler 2016).

3.3.2. Analítica de diagnóstico

La analítica de diagnóstico busca determinar la causa de un fenómeno que ocurrió en el pasado. Si bien el valor proporcionado a la empresa es mayor, por otro lado, se requiere de personal más calificado para realizar esta tarea, ya que implica realizar análisis de patrones y tendencias para presentar informes flexibles que permiten visualizar los resultados de forma resumida o detallada, como se muestra en la Figura 29.

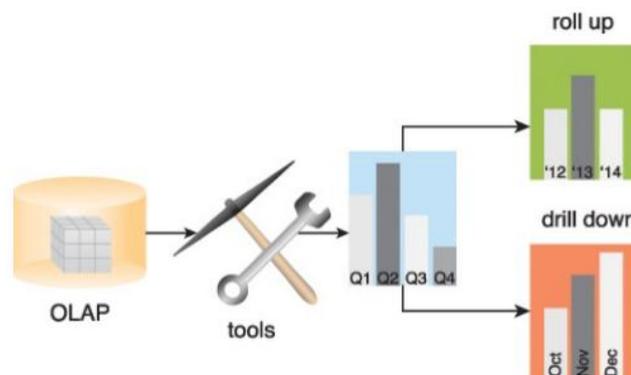


Figura 29 Analítica de diagnóstico con resultados que pueden resumirse o detallarse (Erl, Khattak, y Buhler 2016).

3.3.3. Analítica predictiva.

La analítica predictiva tiene por finalidad determinar el resultado de un evento que podría ocurrir en el futuro, para ello se desarrolla un modelo de predicción. Este tipo de analítica ofrece una información mejor que las anteriores, pero con el costo de mayor complejidad en el desarrollo como en el uso. Las técnicas usadas en el modelo

se basan en patrones y tendencias que se encuentran tanto en los datos históricos como en los actuales, para con ello generar interfaces fáciles de utilizar, como se ilustra en la Figura 30.

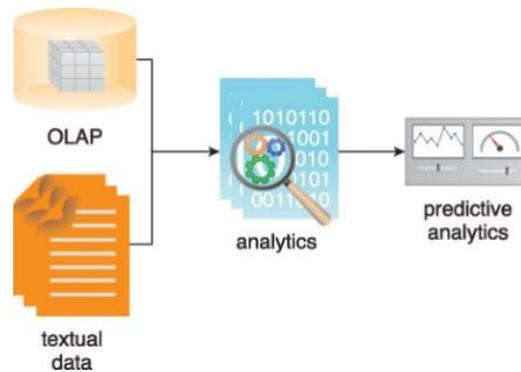


Figura 30 Herramientas de analítica predictiva e interfaz amigable (Erl, Khattak, y Buhler 2016).

3.3.4. Analítica prescriptiva

La analítica prescriptiva dispone y justifica que acciones deben tomarse si se desea obtener un determinado resultado. Para su realización se requiere de herramientas especializadas, así como de conocimiento avanzado para tratar con datos tanto internos como externos y generar un sistema de recomendación como se muestra en la figura 31.

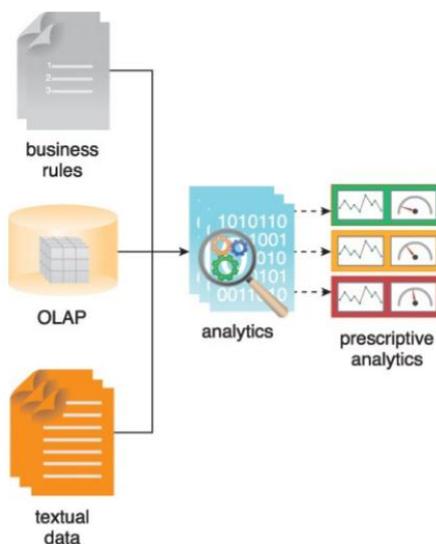


Figura 31 La analítica prescriptiva incluye el uso de reglas de negocio (Erl, Khattak, y Buhler 2016).

3.4. Inteligencia de Negocio (Business Intelligence)

La inteligencia de negocio o BI (por sus siglas en inglés), informa sobre el rendimiento de una empresa mediante el análisis histórico de grandes cantidades de datos generados al interior de la empresa.

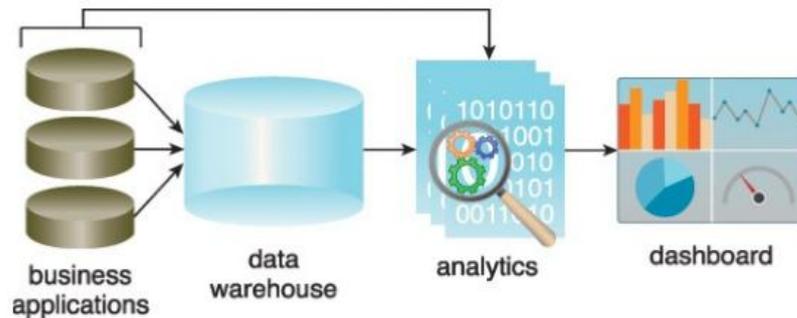


Figura 32 BI puede utilizarse para analizar datos a través de un tablero de control (Erl, Khattak, y Buhler 2016).

Indicadores clave de rendimiento (KPI)

Un KPI es una métrica que sirve para medir el grado de logro de objetivos en áreas importantes dentro de la empresa; se utiliza como puntos de referencia.

3.5. Análisis científico de datos

La información generada en la actualidad es tan variada y veloz, que necesita ser almacenada y procesada en tiempo real. El procesamiento es especialmente difícil cuando los datos no se ajustan a las estructuras de datos tradicionales, por lo que se hace necesario todo un proceso para asegurarse que la información esté en un estado adecuado para procesarse e identificar patrones significativos que lleven a la extracción de información útil.

3.5.1. BI versus ciencia de datos

La diferencia principal entre BI y Data Science, es que BI se enfoca en información pasada que busca explicar el porqué de determinados comportamientos, mientras que la Ciencia de Datos se concentra en el presente y se enfoca en el futuro, utilizando de forma más prospectiva y exploratoria los datos desagregados.

La figura 33 muestra el alcance y tiempo entre BI y la Ciencia de Datos.

Se debe tener en cuenta que la simple aplicación de métodos avanzados a los datos no necesariamente proporciona valor a la empresa, por lo que es importante aprender cómo extraer eficazmente información de los datos y comunicarlos de forma efectiva.

3.5.2. Ciclo de vida del análisis de datos

De acuerdo con Dietrich (Dietrich et al. 2015), Un proyecto de ciencia de datos tiene un ciclo de 6 pasos, en donde con frecuencia se comete el error de asignar poco tiempo a las tareas de entender el problema, considerar los requisitos y planificar. Estos pasos son:

1. Descubrimiento
2. Preparación de datos
3. Planificación de modelos
4. Ejecución del modelo
5. Comunicar los resultados
6. Operacionalización

A continuación, se resume lo que implica cada fase:

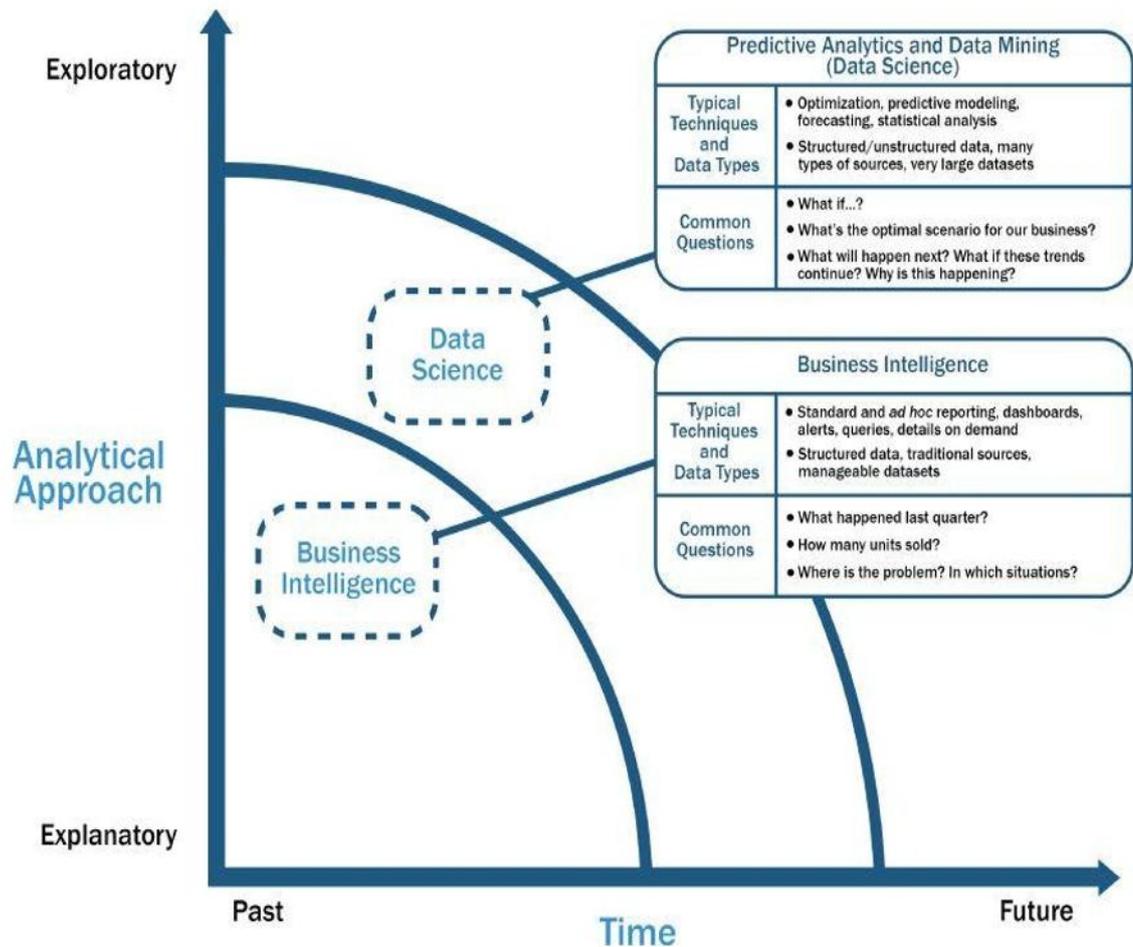


Figura 33 Comparación entre BI y Ciencia de Datos (Dietrich et al. 2015).

Fase 1 - Descubrimiento: El equipo aprende el dominio del negocio, así como de experiencias de los cuales se pueda aprender. Se debe definir el problema del negocio y formular hipótesis iniciales sobre los datos y modelos a obtener.

Fase 2 - Preparación de los datos: Implica trabajar con datos y realizar análisis mientras dura el proyecto. Se requiere ejecutar el proceso extracción, transformación y cargar (ETL) para que los datos sean adecuados para el análisis. Es necesario familiarizarse a fondo con los datos en esta etapa.

Fase 3 - Planificación del modelo: Se determinan los métodos, las técnicas y el flujo de trabajo a seguir en la siguiente fase. Se realiza un análisis exploratorio de los datos para conocer las relaciones entre las variables y seleccionar los modelos más adecuados para procesar las variables clave.

Fase 4 - Construcción de modelos: Se construye el modelo y se preparan datos para las pruebas y su entrenamiento.

Fase 5: Comunicación de los resultados: Se determina si los resultados del proyecto son exitosos o no; adicionalmente se deben identificar los hallazgos clave, cuantificar el valor para el negocio, y resumir y transmitir los resultados a los interesados.

Fase 6-Operacionalización: Se entrega reportes finales, resúmenes, código y documentos técnicos. (Dietrich et al. 2015)

3.6. Proyección y tratamiento de datos mediante técnicas estocásticas

3.6.1. Pronóstico estocástico

Un proceso estocástico es una ecuación matemáticamente definida que puede crear una serie de resultados no deterministas a través del tiempo; Esto significa que los valores pronosticados no dependen únicamente de una determinada constante que lo relacione con la variable independiente, pudiendo obtenerse resultados diferentes cada vez que se pronostica para un mismo valor de x .

Una característica interesante de los procesos estocásticos es que no es imprescindible que existan datos históricos, ya que el modelo no tiene que adaptarse necesariamente a un conjunto de datos históricos, siendo suficiente calcular, estimar o realizar suposiciones sobre los rendimientos esperados y la volatilidad de los datos históricos (Mun 2010).

Los modelos estocásticos son aquellos que pueden ser analizados en términos de probabilidad, y complementan a los modelos determinísticos mediante la adición de datos aleatorios. Los resultados se deben a la realización de experimentos aleatorios, que no es más que una acción o proceso que conduce a uno de varios posibles resultados. (Keller 2014)

Los valores pronosticados caen dentro del espacio de muestra, que es la lista de todos los resultados posibles del experimento.

Cuando utilizamos procesos probabilísticos, a menudo se requiere determinar la probabilidad de que un evento ocurra, esto se logra mediante la suma de las probabilidades de los eventos simples; teniendo por evento al conjunto de uno o más eventos simples en un espacio de muestra.

Como lo ejemplifica Keller (Keller 2014), si tenemos los eventos simples de lograr una calificación y sus probabilidades,

$$P(A) = .20, P(B) = .30, P(C) = .25, P(D) = .15, P(F) = .10$$

La probabilidad del evento *pasar de curso*, será

$$P(\text{Pasar el curso}) = P(A) + P(B) + P(C) + P(D) = .20 + .30 + .25 + .15 = .90$$

3.6.2. Proceso de crear un modelo probabilístico

Se comienza con la creación de un modelo determinista que se aproxime a la relación de interés, para luego agregar un término que represente al error aleatorio del componente determinístico.

Keller ejemplifica esta realidad, mediante la búsqueda de determinar el precio de una casa (y), en donde el modelo determinístico sea una regresión lineal basada en el precio del metro cuadrado (x) será:

$$y = 100,000 + 100x$$

Como el precio de una casa no será exactamente lo que el modelo predice, una mejor aproximación será cuando se lo adicione el componente aleatorio ϵ , que será la diferencia entre el valor real de venta y el precio calculado.

$$y = 100,000 + 100x + \epsilon$$

La forma general del modelo anterior se denomina modelo de regresión lineal simple y tiene la forma:

$$y = B_0 + B_1x + \epsilon$$

En donde:

B_0 es el valor de y cuando x sea cero

B_1 es la pendiente de la recta

y es la variable dependiente

x es la variable independiente

ϵ es la variable de error

3.6.3. Regresión múltiple

Frecuentemente una variable, como la del ejemplo del precio de una casa, depende de varias variables independientes, por lo que se prefiere desarrollar un modelo que

incluya tantas variables independientes como sea necesario para mejorar la capacidad de predicción del modelo.

Considerando que k variables independientes están relacionadas con la variable dependiente, el modelo de regresión tendrá la forma de la siguiente ecuación:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$$

En donde y es la variable dependiente, x_1, x_2, \dots, x_k son las variables independientes, $\beta_0, \beta_1, \dots, \beta_k$ son los coeficientes, y ε es la variable de error.

En su libro *Seeing the Future: How to Build Basic Forecasting Models* (Vu 2015), Vu ejemplifica un modelo simple de regresión múltiple, indicando que es muy común en los negocios y la economía que existan más de un factor que afecta el movimiento de un mercado, lo que lleva a la construcción de un modelo econométrico como el siguiente:

$$\text{Consumo}_i = a_1 + a_2 \text{Ingreso}_i + a_3 \text{precio}_i + e_i$$

En donde Consumo es el gasto semanal de una persona en bienes o servicios, Ingreso refleja lo que la persona percibe mensualmente, precio es precio promedio de un paquete accionario, e es el error de estimación, y las constantes a_2 y a_3 reflejan la tasa de cambio en el consumo debidos al cambio de una unidad en el Ingreso y en el precio respectivamente (a_1 es el valor de y cuando ingreso y precio sean cero). Las variables ingreso y precio (generalizadas como x s) son denominadas variables explicadoras de la variable dependiente y .

Existen seis supuestos clásicos sobre los datos transversales en los modelos de regresión lineal múltiple (Vu 2015):

- i. El modelo generalizado es $y_i = a_1 + a_2 x_{i2} + \dots + a_k x_{ik} + e_i$
- ii. El valor esperado es 0: $E(e_i) = E(y_i) = 0$
- iii. La varianza es igual a la desviación estándar al cuadrado: $\text{Var}(e_i) = \text{Var}(y_i) = \sigma^2$
- iv. La covarianza es igual a 0: $\text{Cov}(e_i, e_j) = \text{Cov}(y_i, y_j) = 0$ para $i \neq j$
- v. Cada x_{ik} no es aleatoria, debe tomar al menos dos valores diferentes, y x s no está perfectamente correlacionada con ningún otro x (caso contrario estamos frente a un problema conocido como problema de multicolinealidad. Un valor menor a 90% es aceptable)
- vi. La distribución del error es normal, con media en 0: $e_i \sim N(0, \sigma^2)$; $y_i \sim ([a_1 + a_2 x_{i2} + \dots + a_k x_{ik}], \sigma^2)$

3.6.4. Aprendizaje automático.

En la actualidad se está buscando aplicar nuevos modelos de análisis de datos mediante técnicas de software y hardware diseñadas para reconocer patrones y desarrollar acciones para responder a estos patrones, éstas técnicas se denominan de aprendizaje automático (Learning Machine).

Pries y Dunnigan describen que el uso de computadoras para detectar los patrones de interés se justifica en que la máquina puede realizar el trabajo mucho más rápidamente de lo que podría hacer un ser humano, inclusive lo puede hacer todo el tiempo. Se recomienda su uso cuando el costo del desarrollo sea menor que el costo de hacerlo por nuestra cuenta o cuando el beneficio sea suficientemente importante.

Los beneficiarios de este proceso son aquellas instituciones que están siendo inundadas por la gran afluencia de datos y tienen el interés de aplicar decisiones guiadas por los datos (*data-driven*).

Si bien la implementación de los modelos actuales de aprendizaje de máquina implican cierto nivel de ensayo y error, en donde los mismos usuarios también están aprendiendo a medida que llevan a cabo experimentos (Pries y Dunnigan 2015); sin embargo, existen otros enfoques automatizados denominados sistemas cognitivos (Kelly y Hamm 2013) que prometen hacerse cargo de la complejidad para ayudar a personas y organizaciones a tomar mejores decisiones. Estos sistemas ayudarían a profesionales del conocimiento, como en el caso de los médicos, a evaluar y tratar a los pacientes, aprovechando sobre todo la capacidad de conocimiento oculta en los grandes volúmenes de datos (Big Data).

Un claro ejemplo de este tipo de sistemas cognitivos es Watson de IBM, que tras ganar el famoso juego Jeopardy, rompió las barreras de comprensión que solamente eran atribuibles a los seres humanos. Esto si bien significa un incremento de productividad para las organizaciones, también implica serios riesgos para el empleo. Cuando se creía que la automatización era una amenaza para los trabajadores de poca educación, dedicados a tareas rutinarias y repetitivas, nos damos en cuenta que en la actualidad una ocupación rutinaria o mejor dicho predecible, es la que la mayoría de los profesionales del conocimiento estamos realizando, de tal manera que pronto un algoritmo podría reemplazarnos en tales actividades, sobre todo cuando existe gran cantidad de datos para entrenar a uno de estos sistemas (Ford 2015).

El software R cuenta con una gran cantidad de librerías orientadas a ser utilizadas por esta técnica de aprendizaje automático. En la siguiente sección se explicarán las características más importantes de esta herramienta.

3.7. Herramienta R para el tratamiento automatizado de grandes cantidades de datos

Las tres primeras fases de las seis fases del Ciclo de Vida de la Analítica de Datos, involucran varios aspectos de la exploración de datos. En general, el éxito de un proyecto de análisis de datos requiere una comprensión profunda de los datos. También requiere una caja de herramientas para la minería y presentación de los datos. Estas actividades incluyen el estudio de los datos en términos de medidas estadísticas básicas y la creación de gráficos para visualizar e identificar relaciones y patrones. Varias herramientas gratuitas o comerciales están disponibles para explorar, acondicionar, modelar y presentar datos, pero debido a su popularidad y versatilidad, el lenguaje de programación de código abierto R se utiliza para ilustrar muchas de las tareas y modelos analíticos.

3.7.1. Introducción a R

R es un lenguaje de programación y un marco de software para el análisis estadístico y generación de gráficos. Disponible para su uso bajo la Licencia Pública General GNU, el software R y las instrucciones de instalación se pueden obtener a través de la Red Completa R Archive and Network.(Dietrich et al. 2015)

R es un lenguaje de scripting para la manipulación y análisis de datos estadísticos. Se inspiró y es en su mayoría compatible con el lenguaje estadístico S desarrollado por AT & T. R se ha vuelto más popular que S o S-Plus, ya sea porque es gratis y porque más personas están contribuyendo a ello.(Matloff 2011)

R es considerado el arquetipo de las herramientas de grandes datos para las estadísticas, y está basado principalmente en líneas de comandos.(Pries y Dunnigan 2015)

3.7.2. Características de R

R tiene una serie de virtudes que se describen a continuación:

- Es una implementación de dominio público del extenso lenguaje estadístico S, y la plataforma R / S es un estándar de facto entre los estadísticos profesionales.

- Es comparable, y a menudo superior, al poder de los productos comerciales en la mayoría de los aspectos significativos - variedad de operaciones disponibles, programación, gráficos, etc.
- Está disponible para los principales sistemas operativos.
- Además de proporcionar operaciones estadísticas, R es un lenguaje de programación de propósito general, por lo que puede utilizarlo para automatizar los análisis y crear nuevas funciones que amplíen las funciones del lenguaje existentes.
- Incorpora características encontradas en lenguajes de programación orientados a objetos y funcionales.
- Es fácil obtener ayuda de la comunidad de usuarios al ser un software de código abierto. Además, muchas de las nuevas funciones son aportadas por usuarios que son prominentes estadísticos.

Programación orientada a objetos. La característica de orientación de objetos permite simplificar la programación al encapsular los resultados de ciertas funciones en objetos, los cuales pueden ofrecer métodos para consultar detalles de un modo sencillo, por ejemplo, al llamar a la función de regresión `lm()` en R, la función devuelve un objeto que contiene todos los resultados: los coeficientes estimados, sus errores estándar, los residuos, etc. Además, R es polimórfico, lo que significa que una sola función se puede aplicar a diferentes tipos de entradas, procesándolas de manera apropiada.

Programación funcional. En la programación en R comúnmente se evita la iteración explícita, cambiando la codificación con bucles por características funcionales de R, que expresan implícitamente un comportamiento iterativo. Esta característica permite que el código se ejecute con mayor rapidez, sobre todo cuando se ejecuta R en grandes conjuntos de datos.

Algunas de las ventajas de las características de programación funcional de R son: Un código más claro y más compacto, velocidad de ejecución mucho más rápida, menor requerimiento de depuración al tener un código sencillo y además permite una transición más fácil a la programación paralela (Matloff 2011)

En definitiva, R es a la vez un lenguaje especializado y una caja de herramientas de módulos dirigidos a cualquier persona que trabaje con estadísticas, que cubre desde la carga de datos hasta la realización de análisis sofisticados y posterior visualización de los resultados.

R puede utilizarse dentro de Hadoop, que es un sistema de aplicaciones distribuidas para trabajar con grandes cantidades de datos (Big Data), por lo que R se convierte en una gran plataforma de desarrollo de prototipos orientados a soluciones que necesitan ejecutarse con grandes cantidades de datos y visualizar de forma diversa los resultados obtenidos del procesamiento.(Warden 2011)

3.7.3. Ventajas de usar R

Algunas de las ventajas del uso de R de acuerdo a Chapman son:

- Se está convirtiendo en la plataforma educativa preferida para el estudio de estadística
- Su uso se extiende a las más diversas áreas de la ciencia
- Ofrece el más amplio conjunto de herramientas analíticas y métodos estadísticos.
- Permite escribir análisis que pueden ser reutilizados
- Se ejecuta en la mayoría de los sistemas operativos
- Puede trabajar con diversos datos cargados en línea y bases de datos SQL
- Ofrece poderosas funciones gráficas que le dan una expresividad mayor que cualquier hoja de cálculo ofrece
- Es posible inspeccionar si se quiere, la validez de las funciones ya que es de código abierto
- R es gratis.

3.7.4. Cuando se recomienda usar R

Chapman (Chapman 2015) recomienda usar R cuando:

- Se requieren métodos únicos, más nuevos o más potentes que los disponibles en otras herramientas.
- Es necesario volver a realizar un análisis repetidas veces.
- Es necesario aplicar un mismo análisis a varios conjuntos de datos.
- Es necesario desarrollar nuevas técnicas de análisis
- Se requiere tener comprensión y control completo de un método existente.

Capítulo 4: Aplicación de modelos de proyección de la demanda y elaboración del presupuesto

Introducción

Se desea elaborar el presupuesto financiero para el presente periodo, de tal manera que sirva de guía en el desarrollo de las actividades de los diferentes departamentos.

El presupuesto debe considerar los crecientes requerimientos de agua potable en las diferentes zonas atendidas debidas al continuo incremento de socios, lo que permitirá tener la materia prima suficiente y el recurso humano necesario, para de esta manera cumplir eficientemente con el trabajo.

Es importante considerar la evolución de pagos, de modo que se pueda planificar adecuadamente los requerimientos de financiamiento para las obras que se plantea realizar durante el presente período.

Se busca además que el proceso desarrolle modelos que usen técnicas modernas de presupuestación, que permitan obtener la mayor información posible para el apoyo en la toma de decisiones. Adicionalmente se busca definir un proceso que sea fácilmente aplicable y altamente confiable, esto es que los valores presupuestados y reales tengan la menor variación posible.

4.1. Marco estratégico de planeación

4.1.1. Problemas e incertidumbres a los que la Junta se enfrenta

Se ha verificado que existe una tendencia creciente de usuarios que demandan cada vez mayores cantidades de agua potable, por lo que se ha planteado el reto de construir una nueva planta de potabilización, ya que la actual planta se encuentra funcionando a los límites de su capacidad. Estas nuevas inversiones requieren que se tenga un plan de financiamiento que permita llevarlas a cabo; por lo que es necesario pronosticar los ingresos que tendrá la Junta dado el actual esquema tarifario y el pronóstico de consumo.

El primer paso para realizar el presupuesto anual es definir con claridad los problemas y objetivos; para ello se diseñó un formato de captura de información, en el que, tras varias reuniones integradas por el presidente de la Junta y los responsables de los distintos departamentos, se llenó y realizó un análisis de los principales problemas que afectan a cada área, esto con la finalidad de identificar sus causas y definir cursos de acción para solucionarlos. Posteriormente se analizaron

los problemas mediante reuniones organizadas de acuerdo a las áreas definidas en la cadena de valor.

Para poder seleccionar los problemas más importantes se aplicó el principio de Pareto, es decir, identificar al 20% de problemas relevantes, mediante la valoración por consenso, del impacto de cada problema.

En tales reuniones se lograron identificar los problemas indicados en la tabla2, a los que se calificó de 1 a 5, 1 para problemas de poco impacto y 5 a problemas urgentes de alto impacto que requieren una atención prioritaria. Además de los problemas (P), también se incluyeron en el análisis las debilidades (D) obtenidas al elaborar la matriz FODA.

ÁREA	TIPO	PROBLEMA	IMPACTO
MEDIO AMBIENTE	P	FALTA DE TRATAMIENTO DE RESIDUOS DE POTABILIZACIÓN	3
	P	AMPLIACIÓN DE LA FRONTERA AGRÍCOLA Y GANADERA (TALA DEL BOSQUE PRIMARIO Y PAJONAL)	3
	P	300 HECTÁREAS DE BOSQUE DE PINO	3
	P	PRACTICA DE DEPORTES EXTREMOS	4
	P	FALTA DE INFORMACIÓN SOBRE EL ÁREA DE RECARGA HÍDRICA Y ÁREAS PROTEGIDAS	3
	P	QUEMA DEL BOSQUE PRIMARIO Y PAJONAL	3

ÁREA	TIPO	PROBLEMA	IMPACTO
PRODUCCION	P	FALTA DE CONOCIMIENTO DEL COSTO DE PRODUCCIÓN	5
	D	FALTA DE REGISTROS DE PRODUCCIÓN Y PRUEBAS DE CALIDAD	5
	D	CAPACIDAD DE LA PLANTA DE TRATAMIENTO AL LIMITE	5
	P	DESPERDICIO EN LOS TANQUE ROMPE PRESIÓN Y DESARENADOR	5
	P	FALTA DE APROVECHAMIENTO DE DESPERDICIOS	2
	P	FALTA DE AGUA POTABLE PARA BAÑOS ALTO	5

ÁREA	TIPO	PROBLEMA	IMPACTO
DISTRIBUCION	P	FALTA DE CAPACIDAD HIDRÁULICA DE LA RED	4
	P	DESABASTECIMIENTO EN HUZHIL Y GUADALUPANO ALTO	5
	P	FALTA DE CAPACIDAD DEL RESERVORIO	3
	P	FALTA DELIMITACIÓN ÁREA DE COBERTURA	4
	P	FALTA DE CONTROL DE PERDIDAS Y FUGAS	5

ÁREA	TIPO	PROBLEMA	IMPACTO
LECTURACION	P	PROBLEMAS DE LECTURACION	5
	P	INCONVENIENTES CON DERECHOS QUE NO SE PUEDEN SUSPENDER POR MOROSIDAD	5

ÁREA	TIPO	PROBLEMA	IMPACTO
FACTURACION	P	VALORES INCORRECTOS EN EL SISTEMA	3
	D	INCUMPLIMIENTO DE LAS NORMAS CONTABLES (MÉTODO DEL DEVENGO)	5
	P	RECLAMOS CONSTANTES DE LOS SOCIOS	3

ÁREA	TIPO	PROBLEMA	IMPACTO
ADMINISTRACION	P	DIFICULTADES DE APLICACIÓN DE CONTROL INTERNO	5
	P	FALTA DE LEYES Y NORMAS INTERNAS ACTUALIZADAS	4
	P	FALTA DE COMPROMISO DE PERSONAL DE TRABAJADORES, EMPLEADOS Y DIRECTIVOS	4

ÁREA	TIPO	PROBLEMA	IMPACTO
CONTABILIDAD / FINANZAS	P	PROGRAMA CONTABLE INADECUADO	4
	P	TARIFAS NO SUSTENTABLES	5

Tabla 2 Problemas importantes de cada departamento
Fuente. Análisis del investigador

4.1.2. Objetivos estratégicos

De acuerdo al análisis de las condiciones en las que se desenvuelve la JAAPB, los problemas, oportunidades, amenazas y el enfoque expuesto en la declaración de misión y visión, se determinaron entre los jefes de área y varios directivos, los principales objetivos a lograr en cada área, mismos que se resumen en la tabla 3.

ÁREA	OBJETIVOS ESTRATÉGICOS
MEDIO AMBIENTE	Descontaminar la quebrada del desfogue
	Restaurar las áreas de bosque y pajonal desgastados
	Restauración del área del bosque de pino
	Minimizar el impacto ambiental
	Reducir los incendios producidos por el hombre y dar una respuesta oportuna

ÁREA	OBJETIVOS ESTRATÉGICOS
PRODUCCIÓN	Disponer de información oportuna para planificar y controlar
	Optimizar la operación y mantenimiento de producción

	Reducción de tasas de desperdicios
	Garantizar la provisión de agua potable a los habitantes de la parroquia Baños, bajo el esquema 24x7
	Mantener la calidad de acuerdo a las normas INEN

ÁREA	OBJETIVOS ESTRATÉGICOS
DISTRIBUCIÓN	Optimizar la calidad y capacidad de la red de distribución
	Reducir el agua no contabilizada y Disminuir la tasa de pérdidas por daños
	Adecuar la tecnología a la infraestructura

ÁREA	OBJETIVOS ESTRATÉGICOS
LECTURACION	Incrementar precisión de lecturas
	Disminuir morosidad
	Mejorar servicio al Cliente
	Reducción de índices de usuarios no tomados la lectura

ÁREA	OBJETIVOS ESTRATÉGICOS
FACTURACIÓN	Contar con información empresarial integrada y oportuna
	Mejorar la atención al cliente
	Mejorar la sustentabilidad

ÁREA	OBJETIVOS ESTRATÉGICOS
ADMINISTRACIÓN	Mejorar los procesos empresariales y la eficiencia laboral
	Ser un referente a nivel local y nacional, fortaleciendo la imagen de la junta
	Manejar eficientemente los recursos humanos, hídricos, financieros, tecnológicos y materiales.

ÁREA	OBJETIVOS ESTRATÉGICOS
CONTABILIDAD / FINANZAS	Uso adecuado y oportuno de los recursos
	Contar con información empresarial integrada y oportuna
	Incrementar la eficiencia y el cumplimiento de las normas contables

ÁREA	OBJETIVOS ESTRATÉGICOS
SISTEMAS	Uso eficiente de sistemas de información
	Proveer de sistemas de información para apoyo en la gestión
	Trasmitir información importante a los usuarios

Tabla 3 Objetivos por cada departamento
Fuente. Análisis del investigador

4.1.3. Presupuesto de inversiones del plan estratégico 2014-2023

En los cuadros a continuación podemos observar el resumen de los gastos estimados para ejecutarse desde el año 2014 hasta el año 2023

Resumen del plan de inversiones

A continuación, se presenta un resumen de los gastos a ser ejecutados desde el 2014 hasta el 2023 y los valores que serán financiados para la JAAPB

JUNTA ADMINISTRADORA DE AGUA POTABLE DE LA PARROQUIA BAÑOS											
PLANIFICACION ESTRATEGICA											
RESUMEN DE GASTOS A SER EJECUTADOS											
DESCRIPCION	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	TOTAL GASTOS
Inversion Equipos- Obras Civiles	204.800	1.654.700	51.500	20.000	527.000	0	0	0	0	0	2.458.000
Gastos Una sola vez	21.093	2.125	1.000	350	0	0	0	0	0	0	24.568
Gastos Anuales	10.750	175.506	13.196	10.986	4.836	4.836	4.836	4.836	4.836	4.836	239.454
TOTAL COSTOS Y GASTOS	236.643	1.832.331	65.696	31.336	531.836	4.836	4.836	4.836	4.836	4.836	2.722.022
FINANCIAMIENTO	115.268	752.068	2.068	32.068	2.068	2.068	2.068	2.068	2.068	2.068	913.880
GASTO A REALIZAR JAAPB	121.375	1.080.263	63.628	(732)	529.768	2.768	2.768	2.768	2.768	2.768	1.808.142

Tabla 4 Inversiones y Gastos Planificados
Fuente. Análisis del investigador

JUNTA ADMINISTRADORA DE AGUA POTABLE DE LA PARROQUIA BAÑOS																	
PRESUPUESTO Y FINANCIAMIENTO AÑO 2014																	
	COD	ACCIONES	PLAZO			PRIORIDAD 1-5	FECHA		Inversion Equipos- Obras Civiles	Gastos			TOTAL DEL PROYECTO	FINANCIAMIENTO			Aprob
			C/P	M/P	L/P		INICIO	FIN		Una sola vez	Anuales	Años		JAAPB	OTRA ENTIDAD	VALOR	
MEDIOAMBIENTE	M02	Campañas de concientización y capacitación sobre usos alternativos del suelo	x	x	x		permanentemente			368	736	10	7.728	3.680	SENAGUA	368	
MEDIOAMBIENTE	M04	Ejecutar proyectos de conservación, forestación reforestación	x	x	x		permanentemente				3.400	10	34.000	17.000	SENAGUA	1.700	
MEDIOAMBIENTE	M06	Ensayos (In situ) de adaptación de especies nativas	x	x	x		permanentemente				700	10	700	700		0	
PRODUCCION	P02	Adquisición e Instalación de Equipos (Macromedidores)	x				01/04/2014	04/01/2015		11.093			11.093	11.093		0	
PRODUCCION	P04	Instalar el sistema de retrolavado	x				15/04/2014	15/4/2015	20.000				20.000	20.000			
PRODUCCION	P06	Reparación de desarenador	x				01/04/2014	30/4/2015	3.000	0			3.000	3.000			
PRODUCCION	P07	Construcción del Tanque rompedor	x				01/06/2014	30-07-2014	6.000	0			6.000	6.000			
PRODUCCION	P08	Implementar y renovar equipos e instrumentos acorde a tecnología actual	x	x	x		continuo		20.900				20.900	20.900			
PRODUCCION	P10	Estudios de nueva planta - convenio Ecuador Estratégico	x				01/11/2013	30/11/2014	112.000				112.000	0	Ecuador Estratégico	112.000	
DISTRIBUCION	D01	Delimitar el área de cobertura	x				01/08/2014	30/12/2014	6.000				6.000	6.000			
DISTRIBUCION	D02	Diagnóstico de las redes de agua potable		x			01/06/2014	30/11/2014	29.000	1.300		1	30.300	30.300			
LECTURACION	L01	Reactivación de los equipos de lectura	x				01/05/2014	08/01/2014		300			300	300			
ADMINISTRACION	A01	Asesoría Legal permanente	x	x	x		01/04/2014	continuo de acuerdo a la necesidad				2.000	10	20.000	20.000		
ADMINISTRACION	A03	Actualización de los reglamentos y políticas internas	x				01/04/2014	30/12/2014		1.000			1.000	1.000			
ADMINISTRACION	A04	Creación de un manual de funciones y procesos	x				01/04/2014	30/11/2014		1.000			1.000	1.000			
ADMINISTRACION	A05	Crear un plan continuo de capacitación del personal	x	x	x		26/01/2014	30/12/2014		3.000	1.000	10	13.000	13.000			
ADMINISTRACION	A06	Incrementar la asociatividad entre los sistemas de agua	x	x	x		01/04/2014	continuo			450	10	4.500	4.500			
ADMINISTRACION	A07	Establecer convenios de cooperación interinstitucional	x	x	x		01/04/2014	continuo	2.000		1.800	10	20.000	20.000			
ADMINISTRACION	A08	Campaña para posicionar la imagen institucional (Marketing y Publicidad)	x				01/06/2014	continuo	2.000		800	10	10.000	10.000			
ADMINISTRACION	A11	Gestionar incremento de caudal de aprovechamiento del río minas	x				01/05/2014	30/12/2014		900			900	900			
ADMINISTRACION	A12	Reactivación de concesión Yunquillayacu	x				01/04/2014	30/12/2014		1.000			1.000	1.000			
CONTABILIDAD	C03	Estudio técnico de producción y costo real	x				01/04/2014	30/04/2014		1.500			1.500	300	BIZMAN	1.200	
SISTEMAS	S01	Implementación de seguridad cuarto de equipos informáticos	x				01/06/2014	30/12/2014	1.200				1.200	1.200			
SISTEMAS	S05	Mantenimiento de sistemas de información actualizados	x	x	x		01/02/2014	continuo			600	10	6.000	6.000			
SISTEMAS	S07	Implementar un sistema de facturación aplicando el método de devengo	x				01/06/2014	30/12/2014	1.500				1.500	1.500			
SISTEMAS	S07	Sistemas automáticos de respaldo	x				01/07/2014	31/07/2014	1.200				1.200	1.200			
TOTALES									204.800	21.093	10.750		327.093	196.893	0	115.268	

Tabla 5 Presupuesto detallado de Gastos e Inversiones 2014
Fuente. Análisis del investigador

4.2. Situación económica actual: precios, pagos, consumos actuales y pasados.

El sistema de precios utilizado en la Junta parte del principio de subsidiaridad, implementando para ello un método de precio incremental, en el que los primeros metros cúbicos de consumo están subsidiados por aquellos que consumen un valor mayor, de tal manera que quienes consumen hasta 15m³ pagarán cerca de 17 centavos por cada m³, llegando de forma escalonada a valores de un dólar por m³ para los grandes consumidores.

4.2.1. Cantidad de agua consumida

Según los datos de las lecturas de los medidores, en el año 2013 se ha facturado 132.015m³ mensuales, por lo que la cantidad de agua que cada usuario requiere diariamente para realizar sus actividades, valor conocido como dotación, se ha calculado en 178 litros por persona.

La figura 134 muestra que la dotación del usuario de Baños es ligeramente mayor al de Guayaquil, pero menor al de Cuenca y Quito. Esto nos indica que nuestro hábito de consumo es similar al de varias ciudades del mundo, pero es necesario disminuir el consumo ya que según información de la OMS, con 100 litros diarios una persona logra cubrir todas sus necesidades.

En la figura 135 se observa que el consumo está concentrado en los usuarios residenciales, que si bien son el 98% de los derechos, sin embargo solo consume un 89% del total de agua facturada; esto se debe a que varios usuarios comerciales (hosterías) consumen valores de entre 15.000 y 30.000 litros diarios.

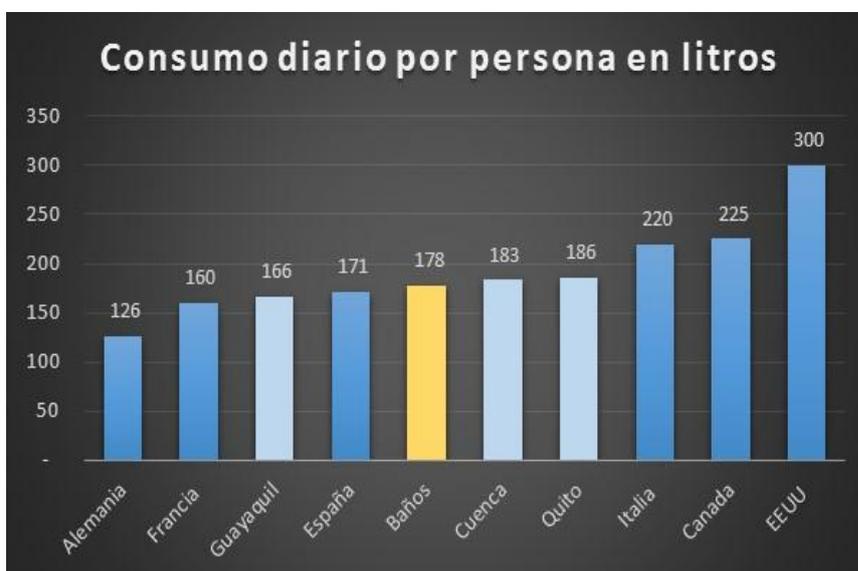


Figura 34 Comparativa de consumo diario
Fuente. Análisis del investigador

4.2.2. Cantidad de usuarios

En diciembre del 2013 se registra la existencia de 5.768 socios, esto implica que cerca de 28.000 personas se benefician directamente del servicio de la JAAPB¹. Las estadísticas muestran que cada año hay un incremento de cerca de 200 socios al sistema.

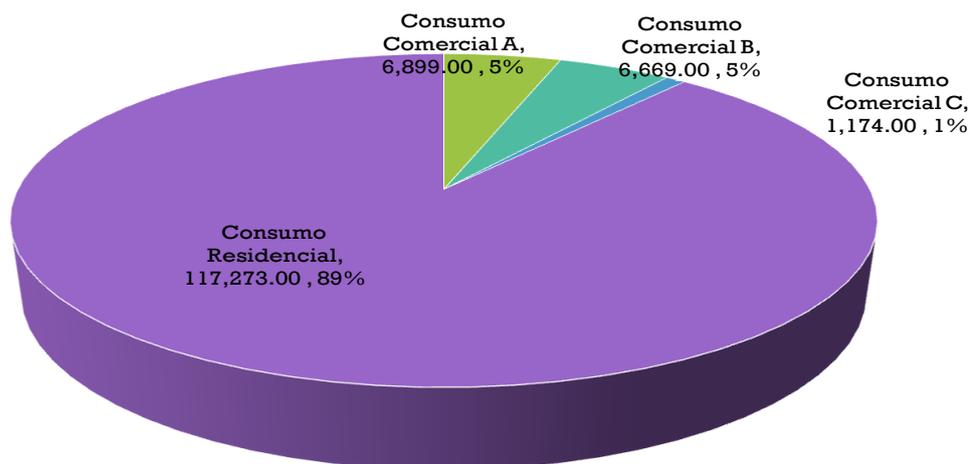


Figura 35 Consumo en m3 por categoría de usuario
Fuente. Análisis del investigador

4.2.3. Ingresos

Los ingresos registrados por la Junta se dividen en ingresos provenientes de: la provisión del servicio de agua potable, la venta de derechos de uso, multas aplicadas, materiales utilizados para instalaciones o reparaciones y otros conceptos de menor cuantía. Estos valores se pueden observar en la tabla siguiente.

INGRESOS JAAPB 2013		
Concepto	Valor Anual	Valor Mensual
Consumo de Agua	362.184	30.182
Derecho de Agua	71.825	5.985
Derecho: Cambio	960	80
Materiales	10.076	840
Mejoras	373	31
Multas	31.237	2.603
Interés	12.249	1.021

¹ Para el cálculo se considera una familia típica conformada por 5 miembros

Total	488.904	40.742
--------------	----------------	---------------

Tabla 6 Ingresos de la Junta en el año 2013
Fuente. Análisis del investigador

Las recaudaciones anuales registradas en la Junta por el consumo de agua desde el año 2007 hasta el año 2013 muestran una tendencia ascendente, siendo en el 2013 un valor de 536.000 dólares.



Figura 36 historial de recaudaciones
Fuente. Análisis del investigador

Analizando el origen de las recaudaciones se verifica que el sector residencial consume un 89% de agua, pero tan sólo provee un 76% de los ingresos.



Figura 37 Recaudaciones de acuerdo a categoría de usuario
Fuente. Análisis del investigador

4.3. Análisis de necesidades de información basada en proyecciones

Para justificar el esfuerzo realizado en el análisis de grandes cantidades de datos, debemos considerar la utilidad que tal análisis generará para la organización, y una de las más importantes, es que permite la mejora del proceso de toma de decisiones (Kudyba 2014).

Es necesario considerar tanto los objetivos del análisis de datos como la disponibilidad de información, así que, al tratarse de un estudio sobre los presupuestos, se parte de las cuentas de los balances presupuestados, identificando aquellas cuentas que son susceptibles a la aplicación de técnicas de proyección.

La JAAPB cuenta con el sistema informático denominado AquaBizman, el que maneja el proceso central de la Junta, esto es, el control de micro medición, la generación de planillas de consumo de agua potable, multas, materiales y derechos, la gestión de cobro, y la gestión de reclamos.

Si bien la información se encuentra desde el año 2000, según una revisión de los registros de toma de lecturas del medidor, es a partir del año 2005 que existen datos completos.

Como parte de este estudio se analizarán 2 componentes del presupuesto del balance de resultados que son factibles de proyectar en base a una gran cantidad de datos, como es el caso de las recaudaciones y los cargos por consumo basados en la micro medición.

En el caso de las recaudaciones, existen 4 puntos de cobro, conformados por 3 cooperativas de ahorro y crédito que cobran de lunes a domingo, más las cajas de la propia institución que tiene un horario de atención solamente de lunes a viernes.

JUNTA ADMINISTRADORA DE AGUA POTABLE DE LA PARROQUIA BAÑOS		
ESTADO DE RESULTADOS		
INGRESOS		
Vtas.Netas Grabadas con Tarifa Cero (Suministro de Agua)	Analizables	Observaciones
Derechos de Instalacion	SI	
Consumo de Agua	SI	
Traspaso de Usuario	NO	Pocos datos
Rendimientos Financieros (Intereses)	NO	Pocos datos
COSTOS Y GASTOS		
Gastos en Personal	NO	Plantilla estable de trabajadores
Contratacion de Servicios	NO	Pocos datos
Produccion de Agua	SI	En base a micromedición
Distribución de Agua	NO	Pocos datos
Mantenimientos	NO	Pocos datos
Gastos en Oficinas y Otros	NO	Pocos datos
Depreciacion Activos Fijos	NO	Pocos datos

Tabla 7 Elementos a los que se puede aplicar proyecciones automatizadas
Fuente. Análisis del investigador

El proceso de micro medición se lo realiza durante los primeros 15 días del mes de acuerdo a la planificación de zonas.

En la tabla 7 se identifican los distintos elementos del balance y las razones que no permiten utilizar las técnicas basadas en Big Data, y es que principalmente no se tiene una gran cantidad de datos para poder ser analizados.

4.4. Proyecciones basadas en modelos tradicionales

En esta sección se revisarán las predicciones generales de consumo de agua potable.

Para realizar el proceso de proyección mediante técnicas tradicionales, se ha procedido a recabar información desde la base de datos y a realizar los análisis utilizando la hoja de cálculo Excel.

4.4.1. Demanda de agua potable

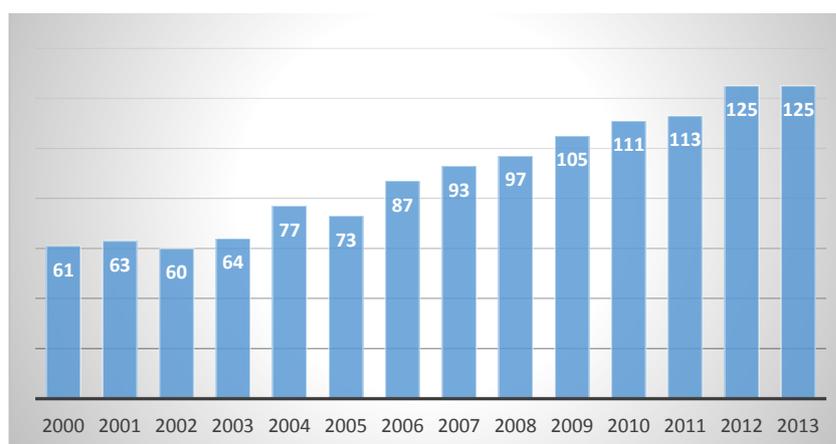


Figura 38 Evolución del consumo promedio mensual de agua potable

Fuente. Análisis del investigador

Se verifica un incremento promedio anual en el consumo de un 6% en los últimos 6 años

4.4.2. Consumo mensual y tendencias

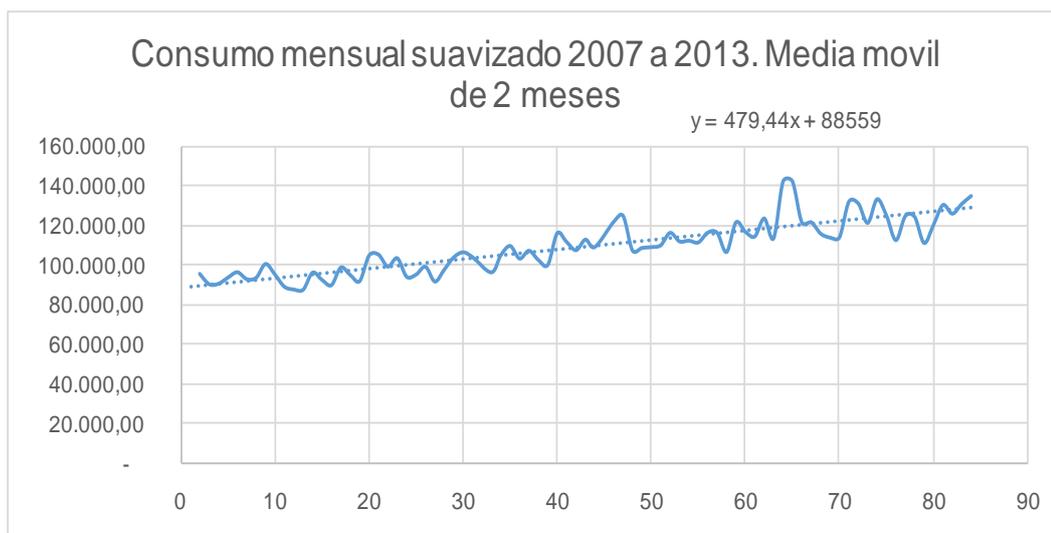


Figura 39 Consumo mensual de agua potable. Suavizado con media móvil n=2
Fuente. Análisis del investigador

La gráfica de consumos mensuales suavizados mediante la técnica de los promedios móviles, muestra que existe una tendencia de incremento de 479m³ cada mes.

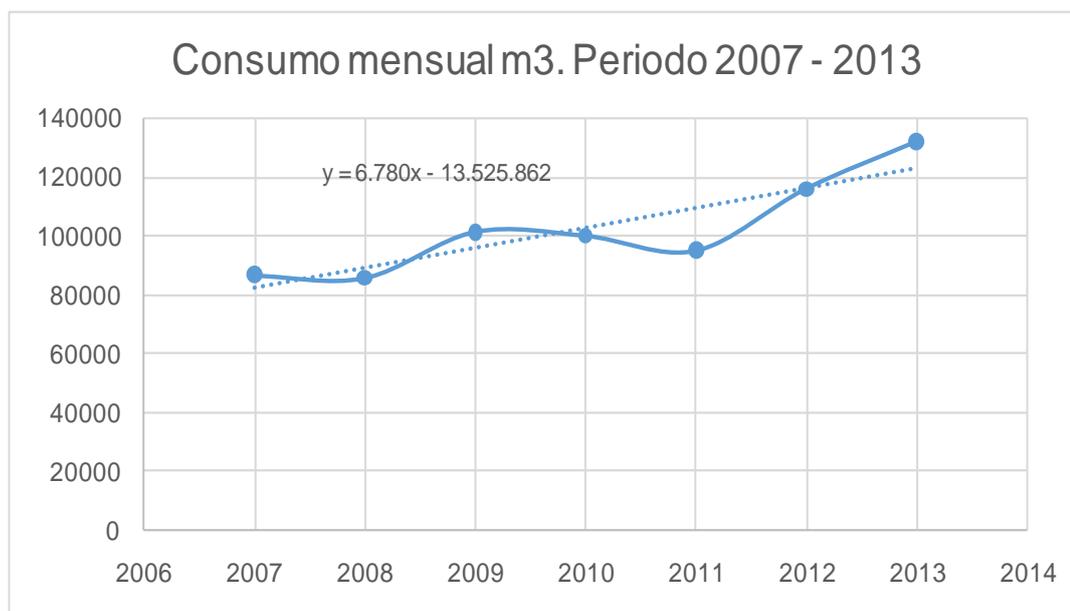


Figura 40 Tendencia en el consumo mensual de agua potable.
Fuente. Análisis del investigador

Las gráficas históricas de los consumos agrupados por año, muestran una tendencia creciente, con una tasa de variación de 6.780m³ cada año.

4.4.3. Proyección de usuarios al 2014

Revisando los datos de los usuarios, se puede observar un incremento constante de cerca de 200 socios por año.

Utilizando los datos históricos y aplicando la función de tendencia de Excel, se estima que en 10 años habrá cerca de 9.460 usuarios, tal como se lo poder ver en la figura 41.

Los 980 sumados al final corresponden a potenciales nuevos usuarios de los sectores denominados Minas y Guadalupano Alto, que se espera puedan acceder a los servicios cuando se construya una nueva planta de tratamiento de agua potable.

$$2024: 8480 + 980 = 9460$$

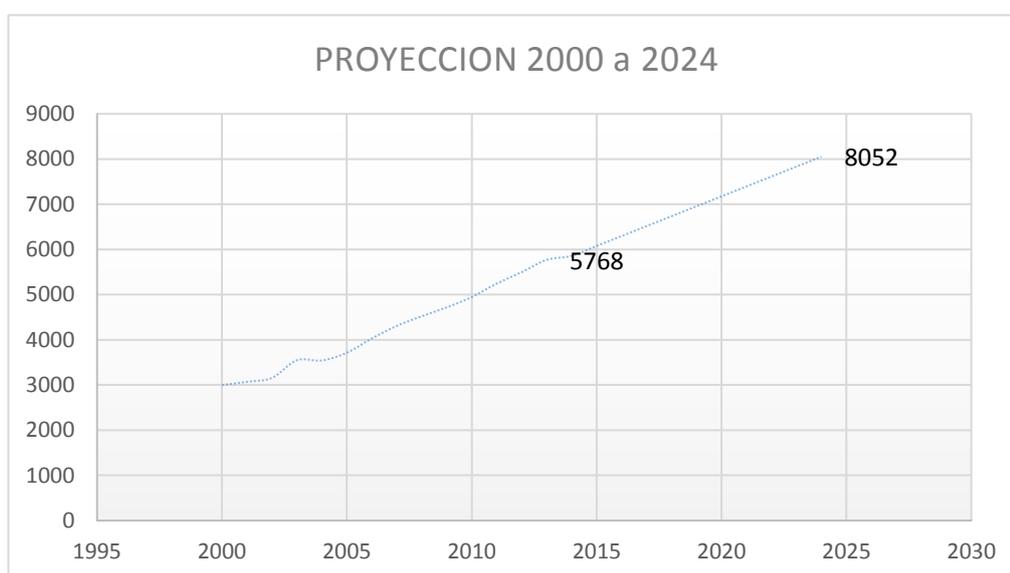


Figura 41 Proyección de usuarios utilizando una línea de tendencia
Fuente. Análisis del investigador

4.5. Modelo de proyección de requerimiento de agua potable

Para la proyección de consumos y requerimientos de agua dentro de 10 años, se ha diseñado un modelo que parte del consumo actual registrado en la micro medición y se ajusta con la capacidad aforada en la planta de tratamiento. Se usa como datos de referencia en el modelo, los resultados informados por otras entidades de agua potable, como es el caso del dato de las pérdidas, que serían de alrededor del 30%, lo que ha permitido llegar a resultados razonablemente acordes con los valores reales de producción y cobros.

El modelo se ha utilizado también para realizar los cálculos de capacidad de producción requeridos hasta el año 2024; para ello se parte del número de socios

proyectados, junto con el nivel de consumo promedio, para como resultado estimar la capacidad de la planta y también el caudal necesario para cumplir con los requerimientos de producción futuros.

La aplicación del modelo se basa en que los 5.268 usuarios del Sistema consumen 132.015m³ de agua al mes, valor al que se deben agregar una estimación del consumo no medido debido a factores como el daño del medidor o conexiones clandestinas.

El valor de consumo no medido, junto con la estimación de las pérdidas en la distribución, se han utilizado como variables para ajustar los cálculos a la capacidad teórica de la planta. En el modelo aparecen estos valores con color rojo, pues serían eventualmente objetivos de control y mejora.

Con las consideraciones anteriores, se determina que se requiere de un ingreso agregado de 53,33 litros por segundo al conjunto de instalaciones domiciliarias.

Las pérdidas en la distribución se deben a fugas y rupturas en las matrices, por lo que la producción debe ser mayor a los requerimientos de los usuarios; Esto significa que la planta debe ser capaz de suministrar a la red un flujo de 71,59 litros por segundo.

De acuerdo al personal técnico de la Junta, el valor estimado de pérdidas para empresas pequeñas es de un 30%, así que en el modelo este valor se lo divide como 25,5 correspondiente a pérdidas en la red y 4,5% como consumo no medido.

La parte derecha del modelo utiliza los valores medios obtenidos en la parte izquierda para determinar las necesidades de producción de agua potable en el futuro. Así que con la estimación de 9.460 usuarios en el 2024, el modelo indica que será necesario que las plantas de tratamiento provean a la red de 113,73 litros por segundo. La figura 42 muestra los efectos de tales estimaciones.

P O T A B L E	Dato	Actualidad	PROYECCION	Proyeccion 2024
	Número de socios	5268		9460
Consumo lecturado mensual m3	132.015,00	225.212,38		
Porcentaje no lecturado	4,5%	4,5%		
m3 no lecurados al mes	6.220,6	10.612		
Requerimientos para socios (m3)	138.236	235.824		
Requerimiento (lts/sg)	53,33	90,98		
D I S T R I B U I O N	REQUERIMIENTO PARA USUARIOS	53,33	90,98	
	Tasa de pérdidas/fugas de red	25,50%	20,00%	
	m3 por pérdias/fugas al mes	47.315,54	58.956,12	
	Requerimientos para distribución (m3)	185.551,14	294.780,60	
	Requerimientos para distribución (lt/s)	71,59	113,73	

Figura 42 Modelo de proyección de requerimientos de distribución
Fuente. Elaborado por el investigador

Para calcular la cantidad de agua cruda² necesaria en el futuro, se parte de que los 71,59 litros por segundo que requiere actualmente la red, implica la necesidad de que ingrese a la planta 77,39 litros por segundo. La diferencia se debe a que en el proceso de tratamiento también existen pérdidas causadas por la necesidad de utilizar agua para lavar 2 veces al día los componentes de la planta y por el desperdicio que significa desviar, mientras se realiza el lavado, el agua cruda que ingresa a la planta.

Según el resultado de un aforo realizado por el personal técnico de la Junta, la capacidad de la planta es de 74.52 litros por segundo; esto implica que la planta al entregar 71.59 lt/sg, trabaja actualmente al 96% de su capacidad. La proyección tiene como variable de decisión el porcentaje de utilización, que al utilizar una tasa esperada del 90%, implica que la capacidad total sumada de las plantas de tratamiento deberá ser 123.97 lt/sg. La tasa de utilización deberá analizarse con detenimiento, puesto que de aquella dependerá el margen de tiempo disponible para seguir cubriendo la demanda de agua potable y el costo asociado de tener plantas con capacidad ociosa. En la figura 43 se muestra la opción de tener una o dos plantas adicionales para cubrir la demanda futura.

El modelo indica que para cubrir la demanda en el año 2024 será necesario una producción de 123.96 litros por segundo y un ingreso a la planta de agua cruda de 134 litros por segundo.

² Agua cruda es como se denomina al agua sin tratamiento de potabilización.

REQUERIMIENTOS DE LA PLANTA DE TRATAMIENTO						
ACTUALIDAD			PLAN DE PRODUCCIÓN	PROYECCION 2014		
P L A N T A	REQUERIMIENTO PARA DISTRIBUCIÓN	71,59		123,96		
	Capacidad planta	74,52				
	% utilización	96%		90%		
		PLANTA 1		PLANTA 1	PLANTA 2	PLANTA 3
	Porcentaje de producción	100%		63%	27%	10%
	Produccion por planta (lt/sg)	71,59		78,10	33,47	12,40
	Tasa de desperdicios por lavado	7,50%		7,50%	7,50%	7,50%
	Desperdicios por lavado (lt/sg)	5,80		6,33	2,71	1,01
	Desperdicios al día (m3)	501,49		547,10	234,47	86,84
	REQUERIMIENTOS DE PLANTA	77,39		84,43	36,18	13,40
				TOTAL	134,00	

Figura 43 Modelo de proyección de requerimientos de capacidad de planta
Fuente. Análisis del investigador

4.6. Aplicación del análisis científico de datos - Análisis exploratorio

La aplicación del análisis científico de datos implica un proceso que parte del descubrimiento del dominio del negocio, la planificación de objetivos y la recolección de datos hasta la generación de informes.

4.6.1. Descubrimiento

En la sección 4.3 se definieron las principales causas que llevan a la necesidad de realizar un análisis científico de datos, que se resume en que se debe considerar la utilidad que tal análisis generará para la organización, permitiendo la mejora del proceso de toma de decisiones. Precisamente el objetivo que se busca, es que mediante este proceso de análisis se puedan tener valores pronosticados de consumos e ingresos cercanos a la realidad, lo que permitirá tomar decisiones apropiadas en cuanto a la planificación de nuevas inversiones y la financiación requerida.

Al tratarse de un estudio sobre los presupuestos, se determinó que las cuentas que son susceptibles a la aplicación de técnicas de proyección son las de consumo de agua potable y cobros de servicios, que son aquellas de las que existe suficiente información acumulada y tienen el mayor impacto tanto en los costos como en los ingresos.

La tabla 7 de la sección 4.3 resume los elementos del balance y las razones de uso de las técnicas basadas en Big Data.

A priori se considera que los métodos factibles de aplicar serían los de series de tiempos para el estudio de los consumos y la técnica de modelo correlacional tanto para el consumo como para los ingresos.

4.6.2. Obtención y preparación de los datos a ser analizados

En esta fase se realiza el proceso denominado ETL (Extracción, Transformación y Carga por sus siglas en inglés).

La información que se encuentra en las bases de datos, normalmente está organizada de una manera tal que garantice el registro de cada uno de los elementos importantes de la transacción; sin embargo no todos los campos son útiles para la tarea de predicción, por lo que al realizar la extracción desde la base de datos, es en donde se ha realizado la primera acción de limpieza al seleccionar las columnas requeridas. Se requerirán procesos ulteriores de limpieza y tratamiento de los datos cargados a la herramienta de análisis, de acuerdo a las necesidades que el proceso de modelización genere.

De cerca de medio millón de datos analizados, se ha encontrado 36 registros con datos erróneos en las lecturas, donde aparecen valores de consumo negativos, los que luego de un estudio minucioso, se determinó que se trata de medidores que, debido a su tiempo de uso, llegaron ya al final de la cuenta máxima y regresaron a cero. Estos datos se han omitido del estudio puesto que generarían valores incorrectos en la predicción.

El proceso de transformación implica el cambio de la estructura de los datos que se encuentran en la base de datos, que luego de pasar por un proceso de normalización, están divididos en múltiples tablas; sin embargo, para el tratamiento analítico es necesario realizar el proceso inverso denominado desnormalización, esto es, unificar en una sola tabla todos los datos relevantes y transformarlos a un formato de intercambio generalmente aceptado, como son json, xml o csv.

La desnormalización se logra mediante la ejecución de consultas en lenguaje SQL realizadas mediante el software gestor de base de datos, que en el presente caso se trata de "Microsoft SQL Management Studio".

Para generar archivos de texto separados con punto y coma (;) en formato “csv”, simplemente se debe exportar los resultados de la ejecución de la consulta, como se puede ver en la siguiente imagen.

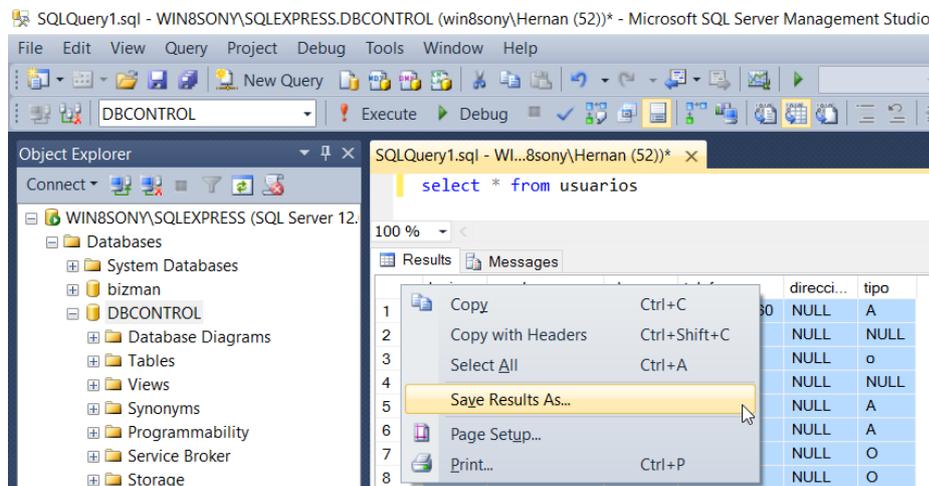


Figura 44 Extracción de datos mediante gestor de bases de datos
Fuente. Análisis del investigador

4.6.3. Preparación de la herramienta y documento de análisis

La herramienta de análisis que se utilizará es R, que se lo descarga gratuitamente de la dirección <https://cran.r-project.org/bin/windows/base/>, la versión a la fecha actual es la 3.4.0.

Adicionalmente descargaremos el entorno de análisis RStudio del sitio <https://www.rstudio.com/products/rstudio/download2/>.

RStudio es una de las herramientas más poderosas para utilizar R ya que integra un editor de código y herramientas de depuración y visualización. Es aquí donde se realizarán los análisis individuales que luego se colocarán como parte de un script general de procesamiento.

Una de las premisas del análisis científico de datos es que los resultados puedan repetirse y ser validados por otros investigadores, además es necesario que los informes de resultados se puedan compartir con la comunidad interesada; es por eso que esta tarea se la llevará a cabo mediante RMarkdown, que es un editor de texto integrado en RStudio, en donde se combina la redacción del informe, el código para cargar y analizar los datos, y los resultados de tales análisis. Para ello realizamos lo siguiente:

En el menú de RStudio elegimos “File, New R Markdown”

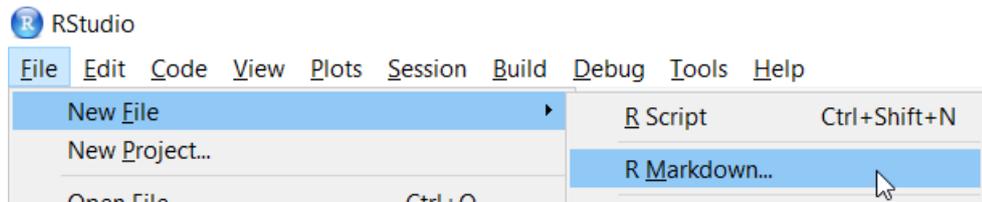


Figura 45 Preparación de nuevo documento RMarkdown
Fuente. Análisis del investigador

Llenamos los datos correspondientes al análisis

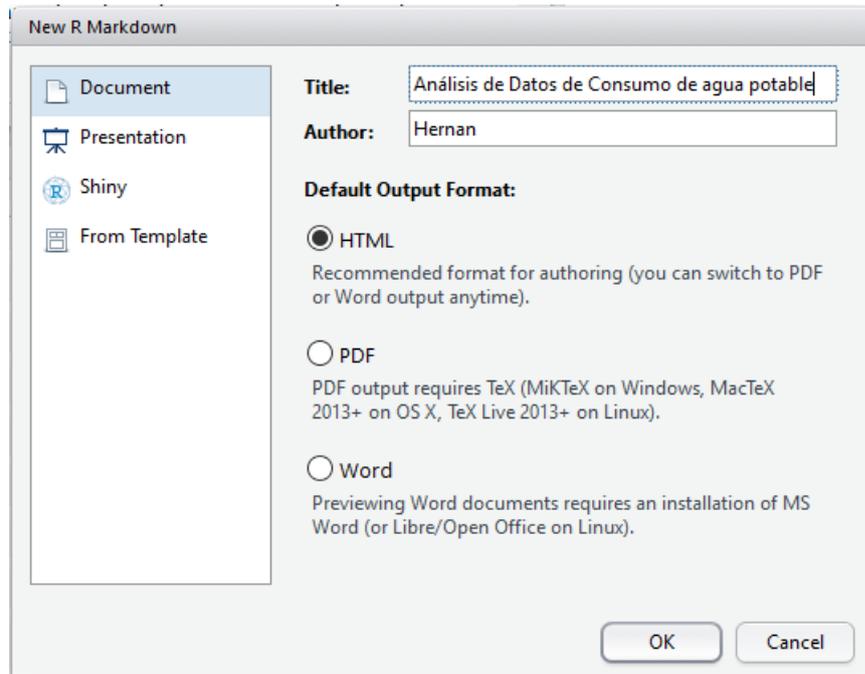


Figura 46 datos de nuevo documento RMarkdown
Fuente. Análisis del investigador

Los comandos se aplican en la consola de RStudio, como ejemplo, en la siguiente imagen se ingresa el comando "version" que imprime los detalles de la instalación actual de R.

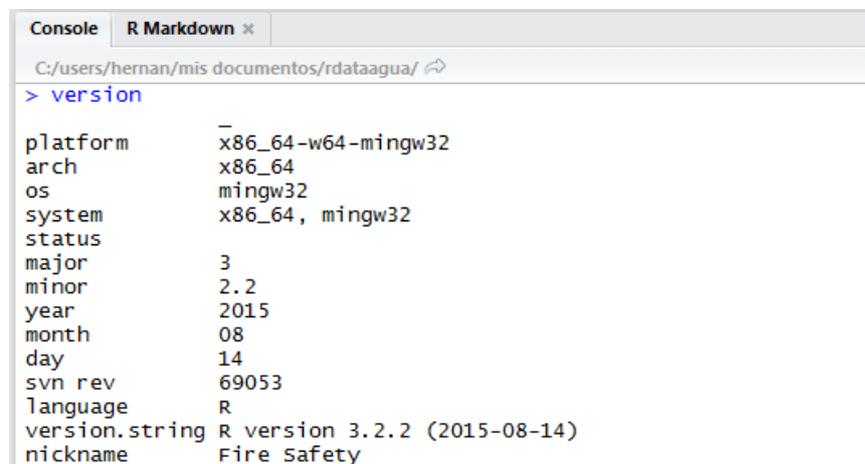
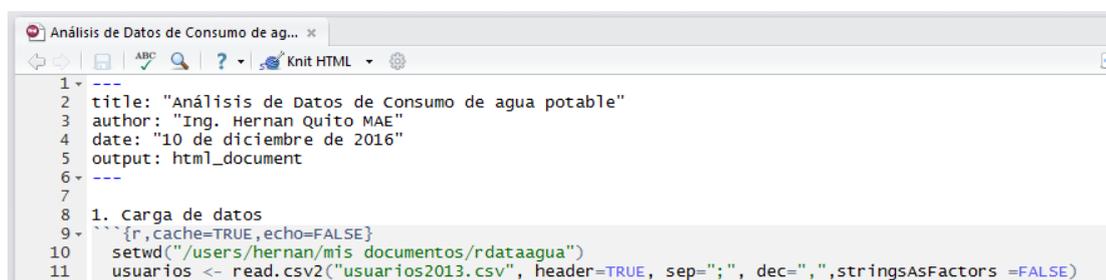


Figura 47 Ingreso de comandos en la consola de RStudio

Luego de comprobar el funcionamiento del comando, se lo puede colocar directamente en el documento RMarkdown antes creado de modo que se vaya generando el script para el informe general.

En la siguiente imagen se puede ver el encabezado del documento y los comandos asociados a la carga de datos.

Los comandos R que deseamos se ejecuten se debe colocar en la sección de script comprendido entre la secuencia de caracteres “`{r}`” y “`````”



```
1 ---  
2 title: "Análisis de Datos de Consumo de agua potable"  
3 author: "Ing. Hernan Quito MAE"  
4 date: "10 de diciembre de 2016"  
5 output: html_document  
6 ---  
7  
8 1. Carga de datos  
9 ```{r,cache=TRUE,echo=FALSE}  
10 setwd("/users/hernan/mis documentos/rdataagua")  
11 usuarios <- read.csv2("usuarios2013.csv", header=TRUE, sep=";", dec="," ,stringsAsFactors =FALSE)
```

Figura 48 Título y script R en documento RMarkdown
Fuente. Análisis del investigador

Las opciones que pueden colocarse en la zona de scripts son “cache=TRUE” para que los datos que se leen o generan en la sección se mantengan en memoria y evitar que se repita el proceso cada vez que generamos el informe; es particularmente interesante utilizar esta opción cuando se cargan los datos, puesto que los dataset suelen tener millones de registros que puede tardar varios minutos solo en cargados en la memoria. Otra opción que se utiliza con frecuencia es “echo=FALSE” que evita que los comandos aparezcan en el informe, sino solamente el resultado de la ejecución de dichos comandos.

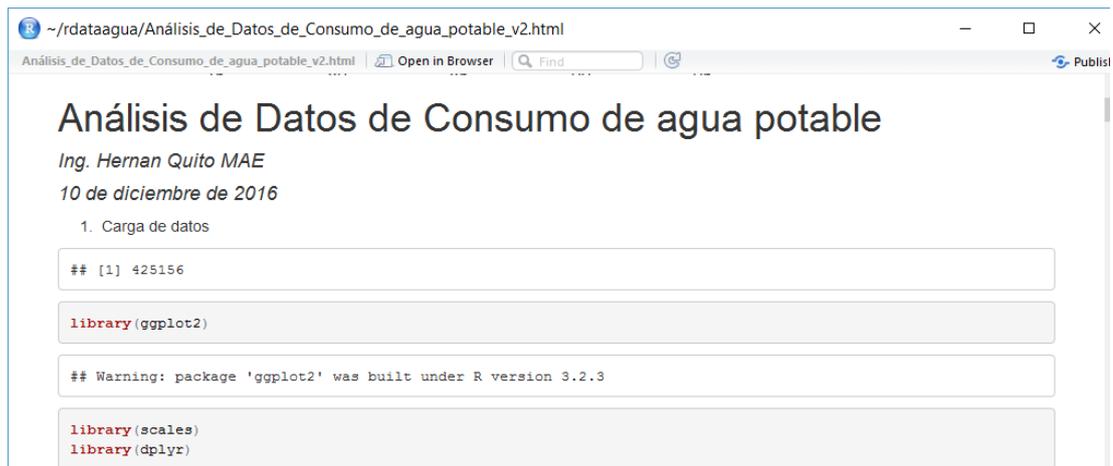


Figura 49 Informe generado en RMarkdown
Fuente. Análisis del investigador

Para generar el informe o reporte se debe pulsar el botón “Knit HTML”, lo que produce un resultado similar al de la figura anterior.

Una de las características clave de R es que existe infinidad de paquetes de herramientas para realizar la carga, el análisis y la visualización de datos. Una importante cantidad de paquetes vienen con la instalación estándar de R, pero la mayoría de paquetes necesarios para actividades especializadas deben ser descargados del repositorio de la CRAN (https://cran.r-project.org/web/packages/available_packages_by_name.html), para luego ser instalados y cargados en el entorno de trabajo.

El proceso de descargar es mediante el comando `install.packages("nombrePaquete")` como se muestra en el siguiente ejemplo que descarga e instala el paquete gráfico “ggplot2” que se utiliza en este trabajo.

```
install.packages("ggplot2")
```

Para poder utilizar las librerías, es necesario realizar la carga del paquete en memoria, lo que se realiza mediante el comando `library(nombrePaquete)`, por ejemplo.

```
library(ggplot2)
```

4.6.4. Carga y limpieza de datos

Los datos que se obtienen en formato de texto con extensión “csv” necesitan cargarse en la memoria para utilizarse como fuente de los diversos análisis. La siguiente figura

muestra el proceso en R, que implica establecer el directorio donde se encuentran los datos y la carga mediante el comando `read.csv2` con los parámetros “`header=TRUE`” que indica que el listado en formato csv tiene cabecera con el nombre de las columnas de datos, “`sep`” indica que el separador de listas es el punto y coma, “`dec`” indica el símbolo de decimal y “`stringsAsFactors`” indica si se debe convertir los campos de texto a factores, que es un tipo especial de dato que sirve para clasificar la información en categorías, y que en este caso se indica no se haga la conversión.

```
setwd("/users/hernan/mis documentos/rdataagua")
usuarios <- read.csv2("usuarios2013.csv", header=TRUE, sep=";", dec=".", stringsAsFactors =FALSE)
totalUsuarios=NROW(usuarios)
consumos <- read.csv2("consumoagua.csv", header=TRUE, sep=";", dec=".", stringsAsFactors =FALSE)
NROW(consumos)

## [1] 425156
```

Figura 50 Script R para carga de datos
Fuente. Análisis del investigador

El proceso de limpieza y preparación de datos involucra lidiar con los valores nulos, reemplazar caracteres no compatibles, convertir valores a fechas entre otros casos. En el siguiente segmento de texto se puede ver la aplicación de funciones, como el caso de crear un nuevo elemento denominado tipo para usarlo en vez del nombre original que tiene caracteres raros que complican la escritura de código. En el caso de la fecha, se pasa el formato de la fecha en el que se encuentra el campo originalmente.

```
#Preparacion de los datos
consumos$codigo=as.factor(consumos$codigo)
consumos$i..periodo=as.Date(consumos$i..periodo)
consumoPeriodo$tipo=consumoPeriodo$i..tipo
consumoPeriodo$periodo=as.Date(consumoPeriodo$periodo)
consumoPeriodoTipo$tipo=consumoPeriodoTipo$i..tipo
pagos$fecha <- as.Date(pagos$i..fecha,"%Y-%m-%d")
```

Figura 51 Script R para ajuste de tipo de datos
Fuente. Análisis del investigador

Se recomienda visualizar en la pantalla las primeras líneas de datos del dataframe, que es el conjunto de datos cargados en memoria por el comando `read.csv2`, mediante el comando `head(dataset)` o `tail(dataset)`. El ejemplo siguiente indica el contenido de las 6 primeras observaciones.

```

> head(consumos)
  i..periodo codigo socio fecha anio nromes mes tipo estadoMedidor
1 2007-01-01 1954 GUERRERO BARRERA - SILVIO IVAN 2007-01-01 00:00:00.000 2007 1 ENERO RESIDENCIAL PUERTA CERRADA
2 2007-01-01 3305 MAYANCELA GRANDA - DAVID ANTONIO 2007-01-01 00:00:00.000 2007 1 ENERO RESIDENCIAL NULL
3 2007-01-01 4437 ALEMAN CHICA - SERGIO ALEJANDRO 2007-01-02 00:00:00.000 2007 1 ENERO RESIDENCIAL NULL
4 2007-01-01 4438 SAQUISARI ZHININ - DANIEL SALVADOR 2007-01-02 00:00:00.000 2007 1 ENERO RESIDENCIAL NULL
5 2007-01-01 759 VELEZ NUNEZ - JORGE LEONARDO 2007-01-03 00:00:00.000 2007 1 ENERO RESIDENCIAL NULL
6 2007-01-01 742 TORRES ORDOA'EZ - CARLOS EUGENIO 2007-01-04 00:00:00.000 2007 1 ENERO RESIDENCIAL NULL

lect_act lect_ant consumo total pago sector estadoLectura valorConsumo valorMantenimiento tipoIngreso
1 2413 2362 51 8.51 PAGADO SAN JACINTO H.B Normal 8.13 0.38 X
2 3099 3033 66 12.93 PAGADO NULL Normal 12.12 0.81 C
3 0 0 0 0.00 PAGADO CENTRO PARROQUIAL Normal 0.00 0.00 X
4 0 0 0 0.00 PAGADO NARANCA Y Normal 0.00 0.00 X
5 1399 1329 70 14.12 PAGADO LOURDES Normal 13.28 0.84 X
6 2949 2920 29 3.64 PAGADO CENTRO PARROQUIAL Normal 3.51 0.13 X

empleado
1 lucho
2 NULL
3 NULL
4 NULL
5 NULL
6 NULL

```

Figura 52 Visualización de parte del contenido de un objeto dataframe
Fuente. Análisis del investigador

4.6.5. Análisis exploratorio de datos

Para el análisis exploratorio se requiere de una serie de procesos estadísticos, análisis de muestras y graficado de variables, de tal manera de que se pueda verificar el comportamiento general y características de los diversos campos que componen el dataset.

Para visualizar un resumen del contenido y características de los campos del dataframe se utiliza el comando `str(nombreDataframe)` como en el siguiente ejemplo se observa que el dataframe “usuario” tiene 6.288 observaciones con 6 campos o variables, que se detallan a continuación con el tipo y unas cuantas muestras del contenido.

```

> str(usuario)
'data.frame': 6228 obs. of 6 variables:
 $ i..codigo : int 1 2 4 5 6 7 8 9 10 11 ...
 $ Usuario : chr "ANDRADE CORDERO, EDGAR VITERO" "AUCAY ARIAS, MARIANA NARCISA" "ALEMAN VERA, LUZ MARIA"
ARIA FILOMENA" ...
 $ Sector : chr "CENTRO PARROQUIAL" "CENTRO PARROQUIAL" "CENTRO PARROQUIAL" "CENTRO PARROQUIAL" ...
 $ estado : chr "ACTIVO" "ACTIVO" "ACTIVO" "PASIVO" ...
 $ tipo : chr "RESIDENCIAL" "COMERCIAL-B" "RESIDENCIAL" "RESIDENCIAL" ...
 $ estadoMedior: chr "NORMAL" "NORMAL" "NORMAL" "CORTADO" ...
> |

```

Figura 53 Script R para obtener información de los campos de un objeto dataframe
Fuente. Análisis del investigador

Cuando se trata de contenido numérico, es posible conocer datos generales mediante los comandos `summary(dataframe$nombreColumna)`, `sd()` para la desviación estándar o `min()` y `max()` para el mínimo y máximo respectivamente, tal como se muestra en el siguiente ejemplo.

```

> summary(consumos$consumo)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-4571.0   0.0   13.0   21.4   30.0  4603.0
> sd(consumos$consumo)
[1] 46.11096
> min(consumos$consumo)
[1] -4571
> max(consumos$consumo)
[1] 4603
<

```

Figura 54 Script R para obtener datos estadísticos de un listado
Fuente. Análisis del investigador

El símbolo “\$” sirve para extraer solamente un campo del dataframe, en el ejemplo anterior, se extrae todo el campo “consumo” del dataframe “consumos” y a estos datos se aplica la función correspondiente.

Para realizar análisis gráficos de la composición de los datos de los usuarios, podemos usar el siguiente script que prepara los datos para generar un gráfico de pastel.

```


```
#preparación de datos para visualización
df=as.data.frame(table(usuarios$estado))
names(df)=c("Estado","Cantidad")
activos=df[df$Estado=="ACTIVO","Cantidad"]
df=df[df$Cantidad>10,]
bp<- ggplot(df, aes(x="", y=Cantidad, fill=Estado))+
 geom_bar(width = 1, stat = "identity")
pie <- bp + coord_polar("y", start=0)
pie +theme(axis.text.x=element_blank()) +
 geom_text(aes(y = Cantidad/3 + c(0, cumsum(Cantidad)[-length(Cantidad)]),
 label = percent(round(Cantidad/totalUsuarios,2), size=4)) + ggtitle("Distribucion de estado
\nde usuarios")

```


```

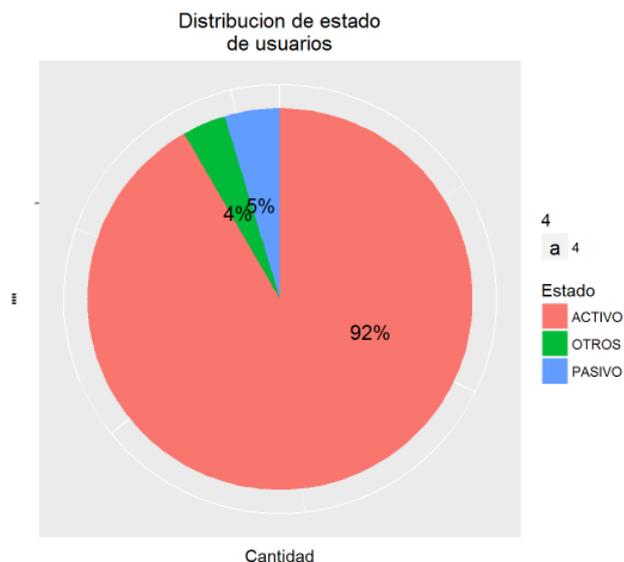


Figura 55 Script R y resultado gráfico de distribución de usuarios por estado
Fuente. Análisis del investigador

De la figura anterior se verifica que un 92% de los socios corresponde a un estado activo, que son los que realmente están utilizando los servicios de la entidad.

El siguiente código genera un gráfico de barras que muestra la cantidad de usuarios por cada sector, en donde se verifica que los principales usuarios son de las zonas más pobladas de la parroquia.

```
#Usuarios por sector. Solo se consideran los activos:
usuarios=usuarios[usuarios$estado=="ACTIVO",]
ggplot(usuarios, aes_string("Sector")) + geom_bar() + coord_flip() + xlab("Sectores") +
  ylab("Cantidad de socios")+ ggtitle("Usuarios por sector")
```

Figura 56 Script R para graficar usando la función ggplot
Fuente. Análisis del investigador

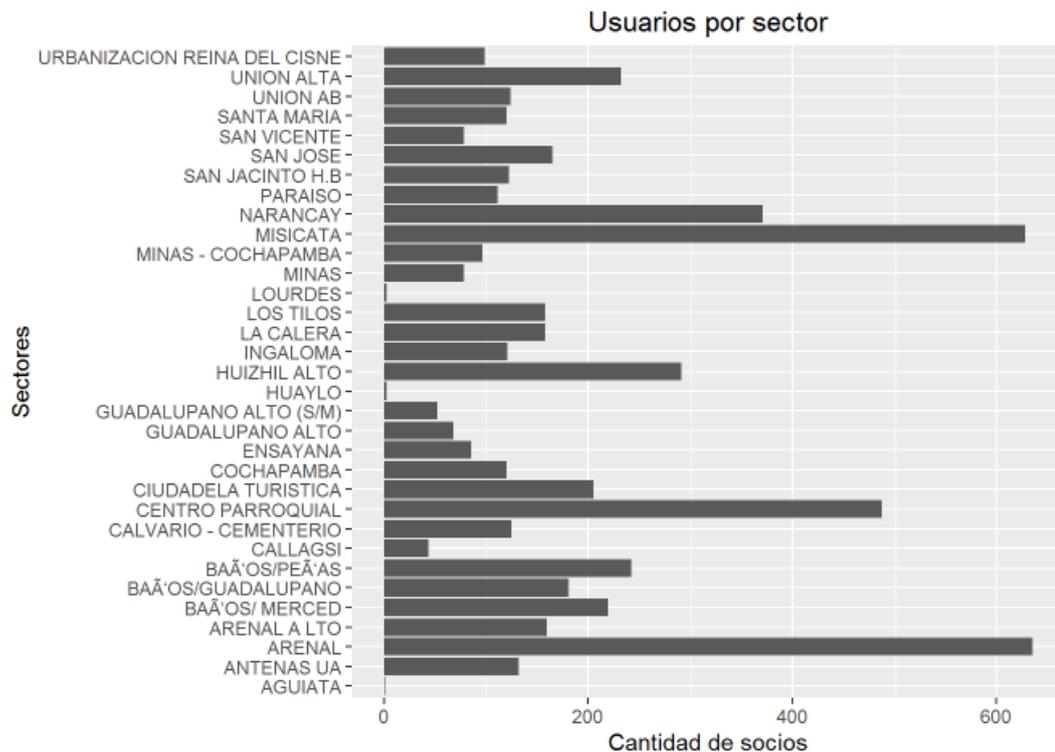


Figura 57 Usuarios por sector
Fuente. Análisis del investigador

Un gráfico que aporta mucha información con solo aplicar un vistazo, es el que genera la función ggplot con el trazo geom_jitter(), ya que permite visualizar la concentración y el nivel de dispersión de las observaciones. En el siguiente script se realiza el proceso denominado subsetting, que consiste en utilizar solo una parte del conjunto de datos, para ello se aplica la forma “dataframe[condiciones, columnas de interés]”; si “columnas de interés” está vacía, indica que se seleccionan todas las columnas. Los datos a graficarse se indican con el parámetro aes(clasificador, valor).

```
#expandiendo para ver el detalle de la distribución de consumos
ggplot(consumos[consumos$consumo<2000&consumos$consumo>0,], aes(tipo,consumo))+geom_jitter()+ ggtitle("Patrón
extendido de consumo por categoría")
```

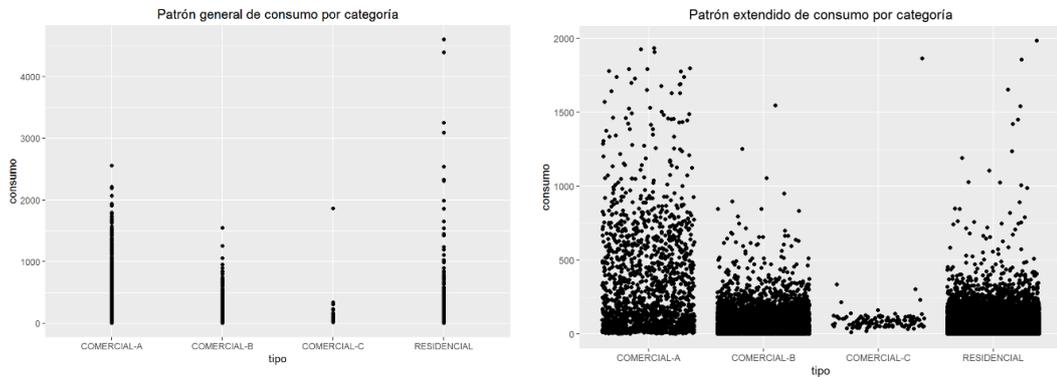


Figura 58 Script R y gráfica de distribución de consumo
Fuente. Análisis del investigador

Se ha aplicado un análisis con un subconjunto de datos en los que se elimina los datos iguales a 0 y aquellos superiores a 2.000m³, esto debido a que existen muy pocas observaciones superiores a este valor, como se ve en la figura (a) y que se deben principalmente a errores en el registro de las lecturas del medidor, mientras que las que tienen cero, no aportan nada en este análisis. El gráfico anterior (b) corresponde a la posición de las observaciones de consumo de agua potable para cada categoría, en donde claramente se muestra que COMERCIAL-C tiene una gran cantidad de observaciones alrededor de los 500m³ mensuales, mientras que los del sector RESIDENCIAL están concentrados principalmente en los consumos menores a 100m³, con pocas observaciones que pasan los 1000m³.

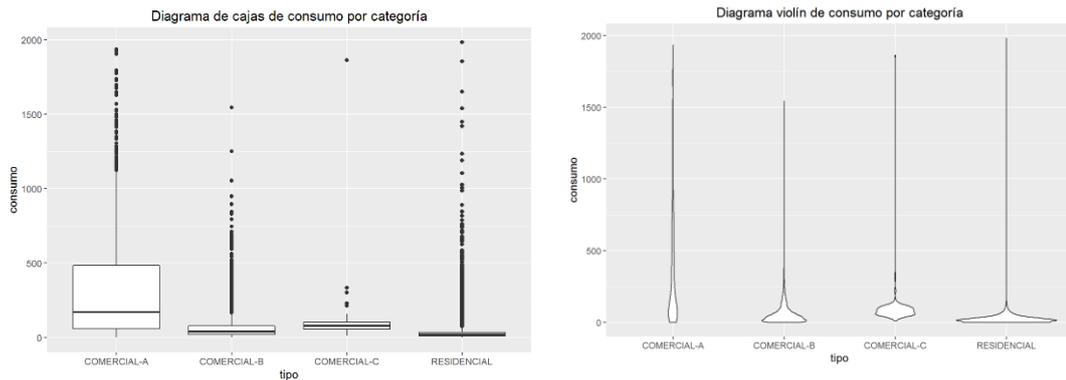


Figura 59 Gráfica de cajas y violín con distribución de consumo de agua
Fuente. Análisis del investigador

En la figura anterior se analiza la composición estadística del consumo de agua potable de acuerdo a las categorías de uso, observándose que un 75% de usuarios comerciales consumen menos de 500m³ mensuales. En general se verifica la existencia de muchas observaciones en los valores bajos de consumo y pocas con

valores elevados; el gráfico del estilo violín muestra esta realidad, en donde el ancho de la parte inferior es proporcional a la cantidad de observaciones encontradas con estos valores bajos de consumo.

Los siguientes comandos permiten extraer solamente los datos correspondientes a la categoría residencial y con consumos menores a 400m³ al mes, lo que permite tener una mejor visualización del comportamiento de la mayoría de consumidores. Los registros con estas condiciones han sido 282.700 según el resultado de la función NROW().

El resultado de la función summary() muestra que un 50% de los usuarios residenciales consumen hasta 22m³ al mes, mientras que en promedio consumen 28.33m³ al mes.

```
par(mfrow = c(1, 2))
x=consumos[consumos$tipo=="RESIDENCIAL"&consumos$consumo<400&consumos$consumo>0,]$consumo
NROW(x)

## [1] 282700

summary(x)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  1.00  12.00  22.00  28.33  37.00 397.00

boxplot(x,col="grey",main="Diagrama de caja\nconsumo Residencial hasta 400m3",xlab = "Consumo",ylab = "Cantidad",cex.main=0.75)
hist(x,col="grey",main="Histograma\nconsumo Residencial hasta 400m3",xlab = "Consumo",ylab = "Cantidad",cex.main=0.75)
```

Figura 60 Script R para ver estadísticas básicas y para generar gráficos de cajas e histogramas Fuente. Análisis del investigador

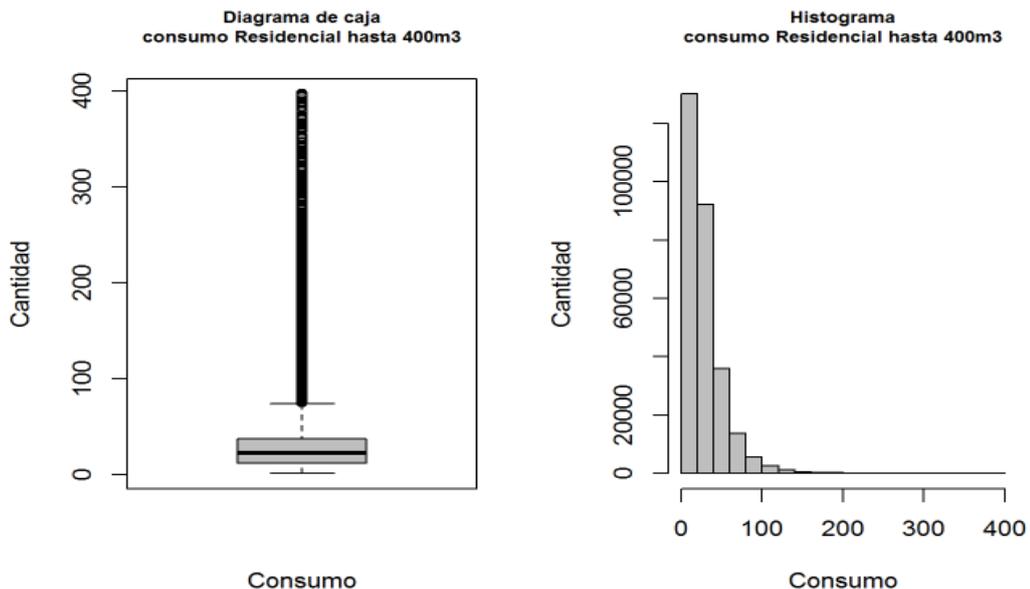


Figura 61 Diagrama de cajas e histograma del consumo de agua potable – categoría residencial Fuente. Análisis del investigador

Tanto el gráfico de cajas como el histograma de los consumos residenciales muestran un comportamiento dominado por la existencia de observaciones menores a 37m³, con pocos valores atípicos superiores a 100m³, un comportamiento previsible para el consumo de agua potable.

Análisis de evolución temporal

Cuando totalizamos los consumos de agua por cada período (mes), podemos ver la gran variabilidad de mes a mes, aunque también se nota una tendencia general de incremento de consumo.

El siguiente script muestra los comandos para resumir tanto la cantidad de consumo como la cantidad de usuarios.

```
#Socios
consumoTP=as.data.frame.table( tapply(consumoPeriodo$socios, consumoPeriodo$periodo, sum))
consumoTP$periodo=as.Date(consumoTP$Var1)
ggplot(consumoTP,aes(periodo,Freq)) + geom_line() + ggtitle("Socios por mes") + xlab("Mes") + ylab("Socios")
```

```
#Consumo
consumoTP=as.data.frame.table( tapply(consumoPeriodo$consumo, consumoPeriodo$periodo, sum))
consumoTP$periodo=as.Date(consumoTP$Var1)
ggplot(consumoTP,aes(periodo,Freq)) + geom_line() + ggtitle("Consumo mensual") + xlab("Mes") + ylab("Consumo")
```

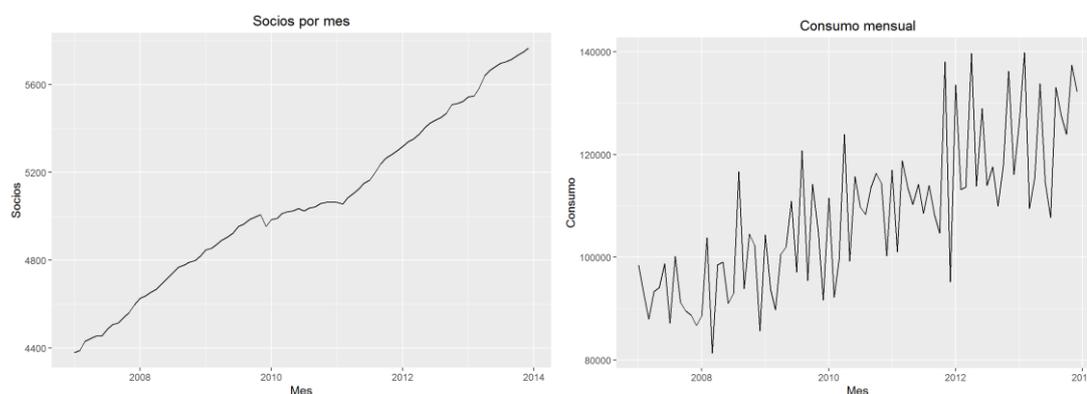


Figura 62 Evolución de cantidad de socios y de consumos mensuales con scripts en R
Fuente. Análisis del investigador

Se puede observar en la primera imagen que existe una tendencia continua al incremento de usuarios, mientras que la segunda imagen evidencia esa alta variabilidad cuando se trata de consumo de agua potable.

Hay diversas razones por las que puede existir tan grande variabilidad, pero la más frecuente en los casos de agua potable, se deben a la existencia de un patrón

estacional derivado de las temporadas de lluvia y sequía por invierno y verano respectivamente.

El siguiente script muestra los pasos necesarios para realizar un análisis considerando la estacionalidad del consumo, para ello primero se requiere totalizar los consumos por período, para luego suavizar su comportamiento mediante la aplicación de la técnica de medias móviles (comando `ma()`), que en el caso del análisis, se ha considerado un período de 3 meses. Se han seleccionado únicamente los 3 últimos años ya que una mayor cantidad produce una imagen con demasiadas líneas que dificultan visualización de los patrones buscados.

```
#componente estacional
consumoTP=as.data.frame(table( tapply(consumoPeriodo$consumo, consumoPeriodo$periodo, sum)
consumoTP$periodo=as.Date(consumoTP$Var1)
consumoTP$anio= format(consumoTP$periodo, "%Y")
consumoTP=consumoTP[consumoTP$anio=="2011"|consumoTP$anio=="2012"|consumoTP$anio=="2013",]
consumoTP$mes= as.numeric(format(consumoTP$periodo, "%m"))
consumoTP$nMes <- months(as.Date(consumoTP$periodo), abbreviate=TRUE)
ggplot(consumoTP,aes(mes,Freq, group=anio, colour=anio)) + geom_line() + ggtitle("Componente estacional") + xlab("meses") + geom_point() + scale_x_discrete(breaks = consumoTP$mes, labels = consumoTP$nMes)+ylab("Consumo")
```

```
#calculo de medias móviles
consumoTP$mm <- ma(consumoTP$Freq,order=3)
ggplot(consumoTP,aes(mes,mm, group=anio, colour=anio)) + geom_line() + ggtitle("Componente estacional - media móvil centrada de 3 periodos") + xlab("meses") + geom_point() + scale_x_discrete(breaks = consumoTP$mes, labels = consumoTP$nMes)+ylab("Consumo")
```

```
#componente estacional residencial
consumoTP=as.data.frame(table( tapply(consumoPeriodo[consumoPeriodo$tipo=="RESIDENCIAL"],$consumo, consumoPeriodo[consumoPeriodo$tipo=="RESIDENCIAL"],$periodo, sum)
consumoTP$periodo=as.Date(consumoTP$Var1)
consumoTP$anio= format(consumoTP$periodo, "%Y")
consumoTP=consumoTP[consumoTP$anio=="2011"|consumoTP$anio=="2012"|consumoTP$anio=="2013",]
consumoTP$mes= as.numeric(format(consumoTP$periodo, "%m"))
consumoTP$nMes <- months(as.Date(consumoTP$periodo), abbreviate=TRUE)
ggplot(consumoTP,aes(mes,Freq, group=anio, colour=anio)) + geom_line() + ggtitle("Componente estacional Residencial") + xlab("meses") + geom_point() + scale_x_discrete(breaks = consumoTP$mes, labels = consumoTP$nMes)+ylab("Consumo")
```

Figura 63 Scripts R para analizar componentes estacionales de consumo de agua
Fuente. Análisis del investigador



Figura 64 Componente estacional de consumo general y de categoría residencial
Fuente. Análisis del investigador

Se puede observar en las gráficas anteriores que el comportamiento total es muy parecido al comportamiento de los consumos de los usuarios residenciales; esto se comprende al considerar que la mayor parte del consumo proviene de dicha categoría, pues a pesar de que en los histogramas se veían que en el sector comercial cada individuo consume valores mucho más altos, su impacto es reducido al ser pocos usuarios comerciales.

El impacto estacional muestra que en los meses de marzo y abril tiende a haber un menor consumo, así también en los meses de junio, julio y agosto; mientras que en septiembre, octubre, noviembre y diciembre se observa una tendencia al mayor consumo, esto se explica principalmente por la incidencia de la temporada de verano con menores lluvias, pero más consumo para riego, y la temporada vacacional (agosto), que implica mayor cantidad de personas en casa consumiendo más agua.

Es posible encontrar sitios en internet que registran valores de factores meteorológicos tales como la temperatura o el nivel de pluviosidad, que inciden directamente en la cantidad de agua que se consume, por lo que son de ayuda para comprender el comportamiento de la demanda de agua potable. Uno de los sitios que publica tales datos es <https://www7.ncdc.noaa.gov/CDO/cdosubqueryrouter.cmd>, donde se puede descargar paquetes de datos con la información de interés.

En la siguiente imagen se puede ver información de sensores ubicados en el aeropuerto Mariscal Lamar con datos desde 1977 hasta 2017.

The screenshot shows the NOAA Satellite and Information Service website interface. At the top, there are logos for NOAA, NESDIS, and the National Climatic Data Center. Below the logos is a navigation bar with the text 'DQC > NOAA > NESDIS > NCDC' and a search field labeled 'Search Field:'. A secondary navigation bar contains links for 'Land-Based Data / NNDC CDO / Product Search / Help'. The main heading is 'Global Summary of the Day (GSOD)'. Below this, it says 'Selected ECUADOR stations - Note: may be slow to load station list on next page'. A section titled 'Select Stations:' contains a table of station data. The 'MARISCAL LAMAR' station is highlighted in blue. Below the table, there are links for 'Order by Station ID - Order by Station Name' and buttons for 'Continue', 'Previous Page', and 'Clear Selections'.

Station Name	Station ID	Start Date	End Date
LOS PERALES	84005099999	02/2013	04/2013
MACARA/J.M.VELASCOI	84205199999	01/1973	06/1973
MACARA/J.M.VELASCOI	84255099999	05/1973	11/1977
MANTA/ELOY ALFARO	84012099999	01/1973	08/1977
MARISCAL LAMAR	84239099999	07/1977	05/2017
MARISCAL SUCRE INTERNATIONAL AIRPORT	84072599999	02/2013	05/2017
MARISCAL SUCRE INTL	84071099999	07/1977	03/2015
MILAGRO	84202099999	07/2011	05/2017
NUEVA LOJA AEROPUER	84063099999	12/1977	08/2013
NUEVO ROCAFUERTE	84132099999	09/2005	05/2017
OTAVALO	84051099999	05/1973	11/1977

Figura 65 Datos de sensores climatológicos
Fuente. Análisis del investigador

El siguiente script realiza el proceso de cargar los datos, cambiarlos al formato y unidad adecuados, y prepararlos mediante la creación de nuevos campos; todo esto de acuerdo a los requerimientos de las técnicas de análisis a ser aplicadas. Así se intenta convertir a fecha (algunos datos están incompletos o nulos y esto hace que falle la conversión) el campo "YEARMODA", para posteriormente extraer el mes y año, que forman el período necesario para graficar y comparar con los gráficos anteriores. Posteriormente se cambia de grados Fahrenheit a Celsius, que es el estándar utilizado en nuestro medio.

```
#Análisis de temperaturas: - carga de datos
temperatura <- read.csv2("temperaturaCuenca.txt", header=TRUE, sep=";", dec=".", stringsAsFactors =FALSE)
temperatura$fecha = try( as.Date( as.character( temperatura$YEARMODA), "%Y%m%d" ) )
temperatura$mes= as.numeric(format(temperatura$fecha, "%m"))
temperatura$anio= as.numeric(format(temperatura$fecha, "%Y"))
temperatura$periodo = paste(temperatura$anio,temperatura$mes,"01",sep="/")
temperatura$periodo = as.Date(temperatura$periodo)
temperatura$TEMP =fahrenheit.to.celsius(temperatura$TEMP, round = 2)
resumenT = temperatura %>% group_by(periodo) %>% summarize(temp=mean(TEMP))
ggplot(resumenT, aes(resumenT$periodo, resumenT$temp)) + geom_line() +geom_point() + ggtitle("Temperatura Mensual") + xlab("Periodo") + ylab("Temperatura")
```

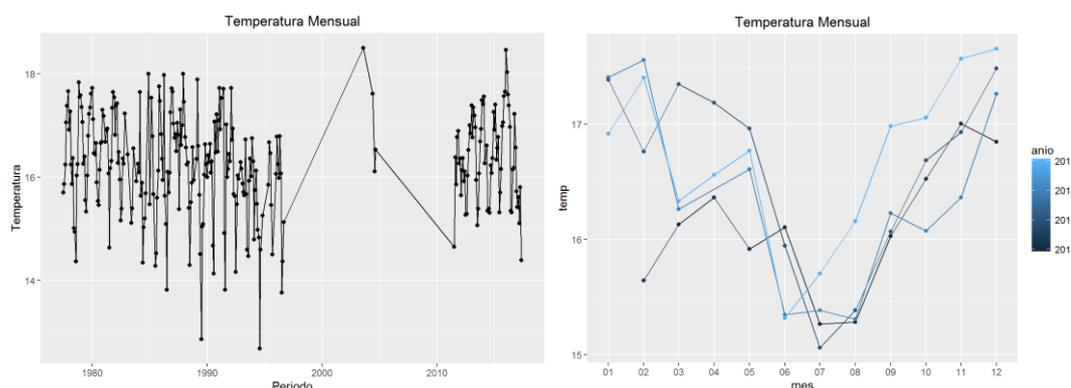


Figura 66 Comportamiento mensual de la temperatura y Script R para la carga y visualización de datos
Fuente. Análisis del investigador

La primera imagen muestra la falta de datos de un gran período comprendido entre 1996 y 2011, por lo que en la segunda imagen se grafica solamente los datos de 2012 a 2015, que son los más cercanos al período de interés de nuestro análisis. Se observa la tendencia a caer la temperatura entre junio y agosto, típica temporada de días soleados y llenos de vientos, pero con fuertes heladas.

Tasa pico-a-promedio mensual

El análisis de la tasa pico-a-promedio mes, es uno de los métodos muy utilizados en el estudio del comportamiento del consumo de agua potable, ya que permite analizar

si la capacidad de la planta de potabilización puede soportar las temporadas de alta demanda o permite diseñar políticas de precios para reducir el consumo.

Para este análisis se deben realizar las transformaciones necesarias al conjunto de datos, agrupándolo por períodos y dividiendo para el promedio de consumo.

```
#Análisis de consumo pico-promedio
resumenConsumo = consumoPeriodo %>% group_by(periodo) %>% summarize(valor=sum(consumo))
resumenConsumo$periodo = as.Date(resumenConsumo$periodo)
resumenConsumo$anio= format(resumenConsumo$periodo, "%Y")
resumenConsumo=resumenConsumo[resumenConsumo$anio=="2011"|resumenConsumo$anio=="2012"|resumenConsumo$anio=="2013",]
mean(resumenConsumo$valor)
```

```
## [1] 119413.2
```

```
resumenConsumo$mes= as.numeric(format(resumenConsumo$periodo, "%m"))
resumenConsumo$picoMedio = resumenConsumo$valor/mean(resumenConsumo$valor)
resumenConsumo$nMes <- months(as.Date(resumenConsumo$periodo), abbreviate=TRUE)
ggplot(resumenConsumo,aes(mes,picoMedio, group=anio, colour=anio)) + geom_line() + ggtitle("Tasa de consumo pico-medio por mes") + xlab("meses") + geom_point() + scale_x_discrete(breaks = resumenConsumo$mes, labels = resumenConsumo$nMes)+ylab("Tasa consumo sobre el promedio")
```

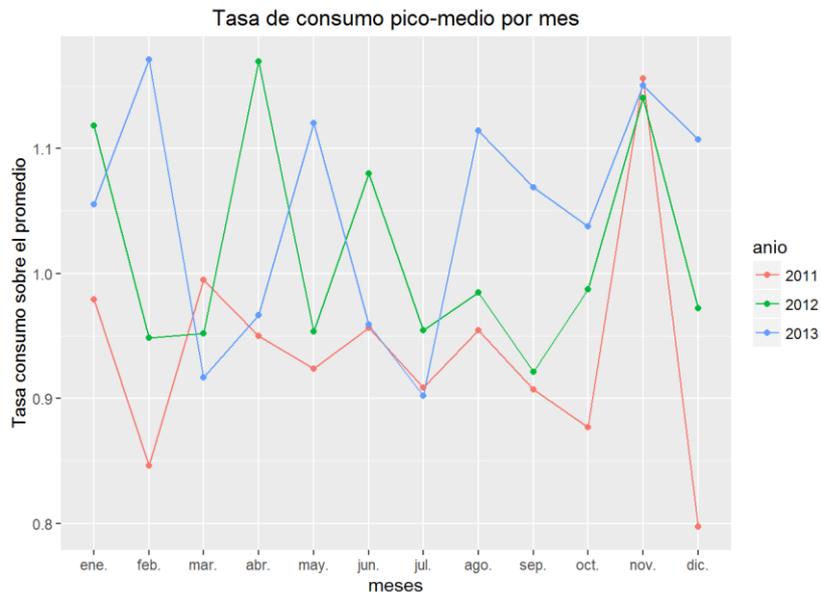


Figura 67 Consumo pico-medio de agua potable y Script R
Fuente. Análisis del investigador

El resultado de graficar la tasa de consumo pico sobre el promedio muestra que, en ciertos meses, sobre todo en noviembre, se puede llegar a consumir un 15% adicional a la media anual.

Consumo per-cápita.

El consumo per –cápita busca calcular cuánto consume cada usuario por día, esto sirve para determinar el comportamiento, en cuanto al consumo de agua, de una comunidad y compararla con otras comunidades.

El cálculo implica totalizar los consumos por cada período, así también se requiere totalizar la cantidad de usuarios en ese período, todo esto clasificado por cada categoría, ya que, de acuerdo con lo analizado en puntos anteriores, cada categoría tiene un comportamiento distinto, que debe ser desglosado para mejorar nuestra comprensión del comportamiento del consumo.

```
# Análisis de consumo per capita por categoría
resumenConsumo = consumoPeriodo %>% group_by(periodo,tipo) %>% summarize(m3=sum(consumo), usuarios=sum(socios))
resumenConsumo$consumoPerCapita=resumenConsumo$m3/resumenConsumo$usuarios
resumenConsumo$periodo=as.Date(resumenConsumo$periodo)
resumenConsumo$anio=format(resumenConsumo$periodo,"%Y")
resumenConsumo=resumenConsumo[resumenConsumo$anio=="2011"|resumenConsumo$anio=="2012"|resumenConsumo$anio=="2013",]
ggplot(resumenConsumo, aes(periodo, consumoPerCapita, group=tipo, colour=tipo)) + geom_line() + geom_point() + ggtitle("Consumo per capita mensuales por categoría")
```

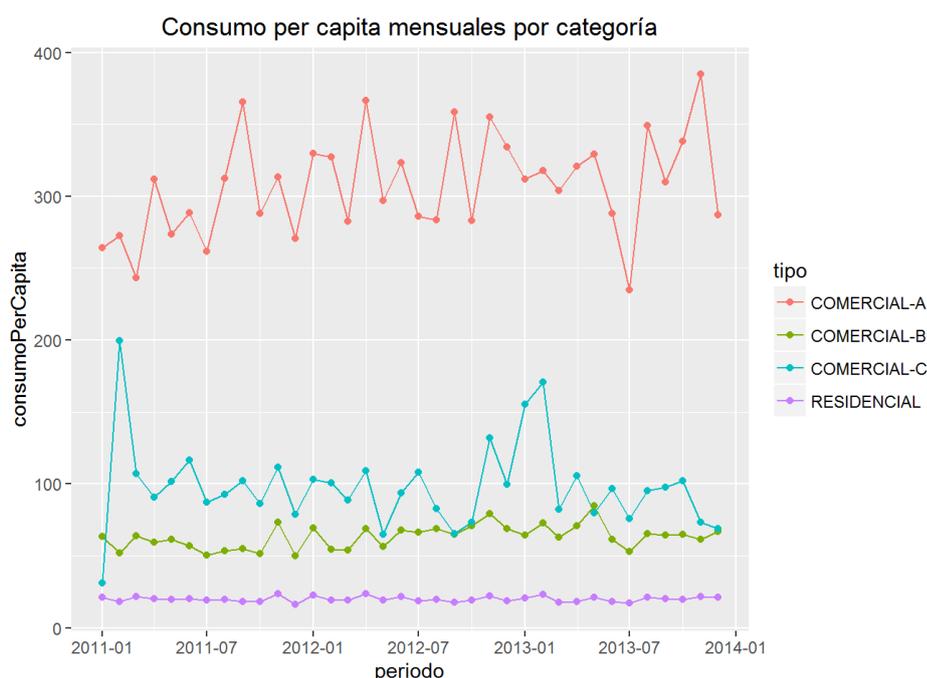


Figura 68 Consumo per cápita por categoría y Script R
Fuente. Análisis del investigador

El consumo per-cápita muestra claramente que los usuarios de la categoría COMERCIAL-A son los que más consumen en comparación con los usuarios residenciales. La diferencia entre estos se puede cuantificar al obtener promedios anuales en donde se observa un consumo de 2.274 litros contra 183 litros.

Los valores antes identificados deben interpretarse considerando los estándares de otras instituciones u organismos internacionales, como es el caso de la ONU que indica que para satisfacer las necesidades mínimas, el ser humano requiere al menos 100 litros diarios, lo que indica que en Baños los usuarios al sobrepasar ese valor, no tienen problemas al respecto. Ciudades como Quito y Guayaquil tienen

comportamientos similares, sin embargo, se debe considerar que el valor subirá si se tiene en cuenta la tasa de pérdidas que afecta a la institución.

```
#Datos anuales per capita
consumoPeriodo$anio=format(consumoPeriodo$periodo,"%Y")
resumenConsumo = consumoPeriodo[consumoPeriodo$sinConsumo==0,] %>% group_by(anio,tipo) %>% summarize(m3=sum(consumo),usuarios=sum(socios))
resumenConsumo=resumenConsumo[resumenConsumo$anio=="2013",]
resumenConsumo$perCapitaDiario =resumenConsumo$m3/resumenConsumo$usuarios*12*1000/365/5
resumenConsumo
```

```
## Source: local data frame [4 x 5]
## Groups: anio [1]
##
##   anio      tipo      m3 usuarios perCapitaDiario
##   (chr)    (chr)   (int)   (int)      (dbl)
## 1  2013 COMERCIAL-A  90613    262      2274.0897
## 2  2013 COMERCIAL-B  84313   1154      480.4045
## 3  2013 COMERCIAL-C   2407     24      659.4521
## 4  2013 RESIDENCIAL 1323695  47542     183.0749
```

Figura 69 Script R para calcular el consumo per-cápita del año 2013
Fuente. Análisis del investigador

El cálculo considera la existencia de 5 miembros en promedio por familia, lo que en el caso de instituciones comerciales lleva a valores elevados, puesto que en la realidad no se cuenta con los datos de la cantidad real de beneficiarios, sobre todo debido a que los más grandes son centros de aguas termales con centenares de clientes que llegan cada día.

La unidad de medida m3 se transforma a litros al multiplicar por 1.000 y al ser datos anuales, se debe dividir por los 365 días, multiplicando por 12 para eliminar el conteo repetido de cada usuario.

4.7. Modelo y Proyecciones basadas en técnicas estocásticas mediante R

En esta fase se procede a organizar la información y a realizar el proceso de análisis descriptivo, en donde además se examinan los datos para identificar las técnicas más adecuadas de tratamiento.

4.7.1. Selección y evaluación de técnicas de pronóstico

Para que un pronóstico sea útil, debe tener un nivel aceptable de precisión, nivel que puede verse afectado considerablemente si los datos no concuerdan con el patrón para el que es fuerte el enfoque de predicción seleccionado.

La selección del método de pronóstico depende de las características de los datos, el objetivo de realizar el pronóstico y el nivel de complejidad que supone tanto para la elaboración del modelo como para su comprensión y análisis.

Nugus resume diversas técnicas de pronóstico y sus características más sobresalientes, las que se pueden ver en la tabla 3.

Otro factor que debemos considerar es la frecuencia con la que se requieren realizar los pronósticos, por ejemplo en los casos de pronósticos repetitivos, éstos tienen la ventaja de poder utilizar la información real como retroalimentación para ajustar y mejorar el modelo (Nugus 2009).

Técnica	Uso	Matemáticas	Datos
Promedios móviles	Pronósticos repetidos sin estacionalidad	Un nivel mínimo de competencia	Esencial datos históricos
Filtrado adaptativo	Pronósticos repetidos sin estacionalidad, donde la naturaleza de cualquier tendencia puede cambiar con el tiempo	Se requiere algunos conocimientos estadísticos	Esencial Datos históricos, pero el nivel y detalle pueden variar
Desvanecimiento exponencial	Pronóstico repetido con o sin estacionalidad	Algunos conocimientos estadísticos para configurar el sistema, pero prácticamente ninguno para operar un sistema bien diseñado	Sólo se requieren datos recientes y previsiones actuales una vez que se ha establecido el factor de suavizado
Regresión lineal simple	Comparar una variable independiente con una variable dependiente en donde existe una relación lineal entre las variables	Algunos conocimientos estadísticos para configurar el sistema, pero prácticamente ninguno para operar un sistema bien diseñado	Una muestra de observaciones relevantes para las variables independientes y dependientes
Regresión lineal múltiple	Cuando más de una variable independiente debe compararse con una variable dependiente	Se requiere algunos conocimientos estadísticos	Una muestra de observaciones relevantes para las variables independientes y dependientes
Análisis de tendencia	Pronosticar a través del tiempo de forma no lineal mediante regresión simple	Se requiere cierta competencia en estadística	Datos históricos con el mayor detalle posible
Análisis de descomposición	Identificar componentes estacionales como parte de otro método de pronóstico	Algunos conocimientos estadísticos para configurar el sistema, pero prácticamente ninguno para operar un sistema bien diseñado	Se requieren los datos históricos para la variable en cuestión
Composición de estimaciones individuales	Pronosticar usando un gran número de estimaciones individuales	No se requiere conocimientos de estadística y sólo un poco de aritmética básica	No se requieren datos pasados

Tabla 8 Diversas técnicas de pronóstico, utilidad y requerimientos (Nugus 2009).

4.7.2. Preparación del modelo

De acuerdo al análisis exploratorio aplicado en la sección anterior a las variables, se debe proceder en esta fase a verificar los posibles modelos que podrían aplicarse, los cuales se tratarán a continuación.

Modelos correlacionales

Los modelos correlacionales permiten pronosticar mediante la cuantificación de la relación existente entre las variables predictoras y las variables dependientes, de tal manera que, si se conoce la una, se pueda determinar la otra. Para esto es muy útil realizar una gráfica de nube de puntos.

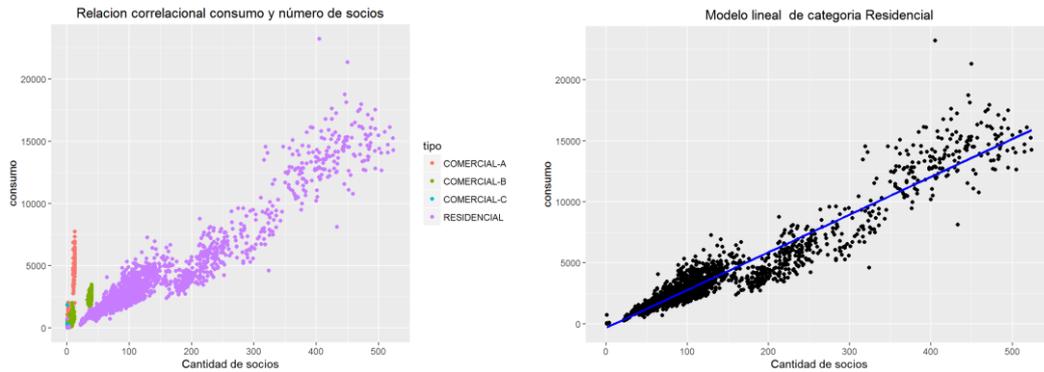


Figura 70 Correlación entre cantidad de socios y cantidad de agua consumida Fuente. Análisis del investigador

La primera gráfica muestra una relación marcada entre la cantidad de socios de un sector y el consumo asociado para los consumidores de la categoría RESIDENCIAL, observándose que, a mayor cantidad de socios, mayor es el consumo. Esto indica que posiblemente exista una relación lineal entre el consumo y la cantidad de socios; sin embargo, esto no se aplica para las categorías COMERCIALES. La comprobación de la existencia de una relación se logra mediante la aplicación de la función estadística de correlación a las variables antes mencionadas.

```
##
## Call:
## lm(formula = subsetCp$consumo ~ subsetCp$socios)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5109.0  -424.8    37.1   402.4 11017.6
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -318.7706    36.4338  -8.749  <2e-16 ***
## subsetCp$socios  30.9437     0.2084 148.450  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1054 on 2088 degrees of freedom
## Multiple R-squared:  0.9135, Adjusted R-squared:  0.9134
## F-statistic: 2.204e+04 on 1 and 2088 DF,  p-value: < 2.2e-16
```

Figura 71 Resultado de aplicar el modelo lineal de análisis al consumo de agua Fuente. Análisis del investigador

Al aplicar la función de modelo lineal se puede ver que efectivamente hay una correlación de 0.95, con un error estándar de 0.2, que da como resultado una relación muy alta de señal ruido (t-value) de 148 que resulta sumamente deseable. En cuanto a la capacidad de predicción de este modelo, se puede verificar que tiene muy buena significancia ($\Pr(>|t|)$), pues representa un valor menor a 2E-16, que se lee como la probabilidad de que los resultados se deban al azar. Asimismo, r^2 de 0.9135 muestra el porcentaje del valor predicho que captura el modelo, esto es que más del 91% del valor predicho lo explica el modelo.

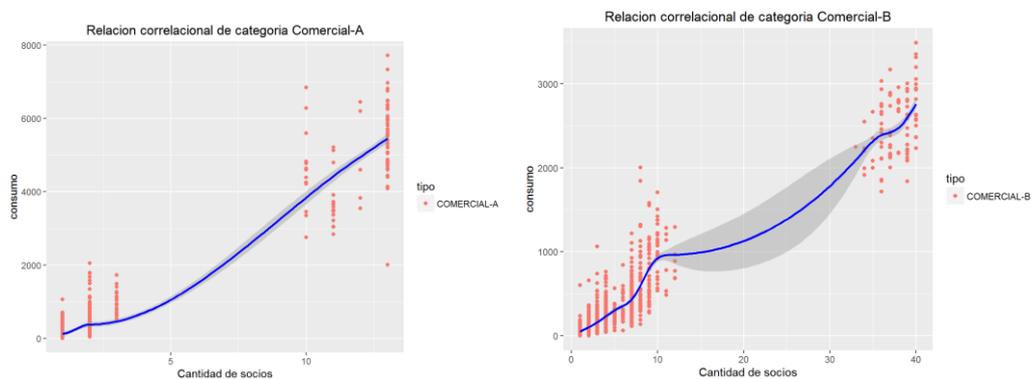


Figura 72 Correlación de consumo de agua para categorías comerciales y cantidad de socios
Fuente. Análisis del investigador

Si bien aparentemente los resultados son alentadores para los usuarios de la categoría RESIDENCIAL, no se puede decir lo mismo de las otras categorías, pues como se observa en las imágenes anteriores, no hay una distribución continua de las observaciones, lo que impide aplicar estos métodos.

Para mejorar la capacidad de predicción del modelo lineal, se debe separar los distintos grupos, en este caso, lo hacemos en función de la categoría.

La siguiente secuencia de instrucciones muestra el proceso para obtener dataframes separados para cada categoría, para luego aplicar el modelo tanto de forma conjunta como de forma separada, obteniendo los gráficos y valores del modelo al final.

Proyecciones con modelo de regresion

```

resumen = consumoPeriodo %>% group_by(tipo, periodo) %>% summarize(m3=sum(consumo), cant=sum(socios))
resumen$periodo=as.Date(resumen$periodo)
resumenRe=resumen[resumen$tipo == "RESIDENCIAL",]
resumenCA=resumen[resumen$tipo == "COMERCIAL-A",]
resumenCB=resumen[resumen$tipo == "COMERCIAL-B",]
resumenCC=resumen[resumen$tipo == "COMERCIAL-C",]
fit <- lm(resumen$m3 ~ resumen$cant)
fitRe <- lm(resumenRe$m3 ~ resumenRe$cant)
fitCA <- lm(resumenCA$m3 ~ resumenCA$cant)
fitCB <- lm(resumenCB$m3 ~ resumenCB$cant)
fitCC <- lm(resumenCC$m3 ~ resumenCC$cant)
summary(fit)$coef

##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3278.11320  314.6514425  10.41824  3.409389e-22
## resumen$cant  18.80556    0.1271072  147.95038  1.554510e-306

ggplot(resumen,aes( cant,m3)) + geom_point() + geom_smooth(method="lm",colour="blue")+ ggtitle("Modelo lineal
general ") + xlab("Cantidad de socios")

```

Figura 73 Script R para aplicar modelo lineal de consumo de agua clasificado por categoría Fuente. Análisis del investigador

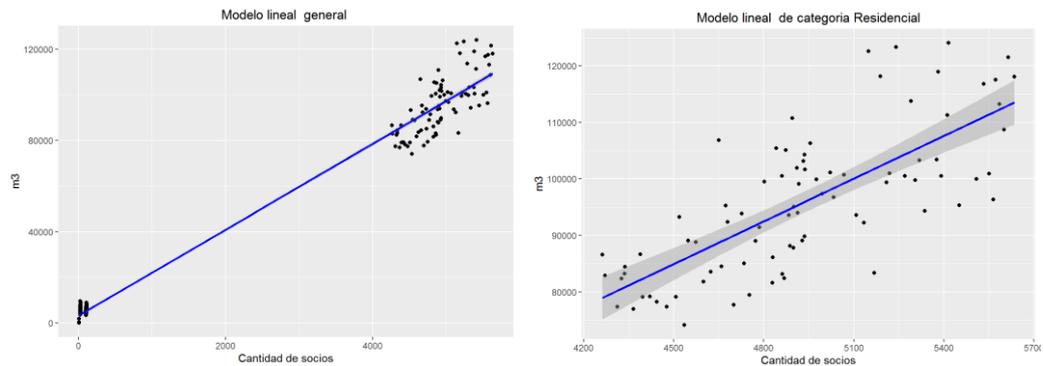


Figura 74 Nube de puntos y línea de estimación de modelo lineal para consumo de agua Fuente. Análisis del investigador

El primer gráfico muestra el modelo aplicado a todos los datos, con dos grupos claramente diferenciados y totalmente separados, lo que significa que este modelo no sería apropiado puesto que en la zona intermedia no se puede pronosticar ningún valor.

Al aislar solamente los datos de los usuarios residenciales, se puede observar en cambio que hay una distribución homogénea en el valor de la variable independiente, lo que permite que si pueda aplicarse este modelo.

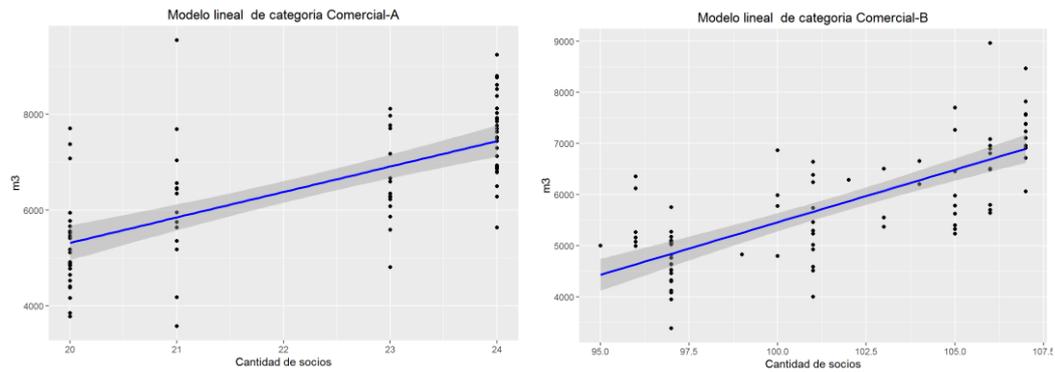


Figura 75 Nube de puntos y estimación de modelo lineal para categorías comercial A y B
Fuente. Análisis del investigador

La relación no es tan clara cuando se trata de pronosticar los consumos de las categorías comerciales como se puede ver en las gráficas anteriores, esto se debe a la limitada cantidad de usuarios en esas categorías, de hecho, el caso del gráfico de sector COMERCIAL-C no tiene ningún sentido, por lo que no se lo ha representado siquiera.

En realidad, los modelos de predicción lineal podrían aplicarse para estimar cuánto debería consumir un sector según la cantidad de usuarios que estén ubicados en dicho lugar; sin embargo, nuestro interés principal es predecir la cantidad de agua potable que será necesaria para abastecer a la demanda futura. Así que, puesto que la cantidad de usuarios se incrementa año tras año, necesitamos un modelo que pueda predecir la tasa de incremento de usuarios y posteriormente el impacto en la demanda de agua potable. En conclusión, se debe buscar otro modelo de predicción.

Medias móviles y series de tiempo

Si consideramos que existe una tendencia continua de crecimiento en el consumo, como en verdad lo hemos observado en las gráficas anteriores, es posible verificarlo mediante la aplicación de un proceso de suavizado con medias móviles.

El cálculo implica construir una serie de tiempo con los datos de interés, en este caso, obtenemos los totales consumidos por cada período mensual en *consumoTP* y *sociosTP*, que se utiliza para crear *consumoTS* y *sociosTS* para un horizonte de tiempo que comienza en enero del año 2007 (fecha desde la que se tienen datos estables) con períodos mensuales ($frequency=12$). La media móvil se aplica con la función `ma()`, cuyo resultado se grafica con la función `plot()`.

```

consumoTP=as.data.frame.table( tapply(consumoPeriodo$consumo, consumoPeriodo$periodo, sum))
sociosTP=as.data.frame.table( tapply(consumoPeriodo$socios, consumoPeriodo$periodo, sum))
consumoTS=ts(consumoTP$Freq, start=c(2007,1), frequency=12)
sociosTS=ts(sociosTP$Freq, start=c(2007,1), frequency=12)
par(mfrow=c(2,2))
plot(consumoTS,main="Serie de tiempo de consumo mensual",xlab="Periodo",ylab="Consumo m3")
plot(ma(consumoTS,3),main="Media movil de consumo mensual\nk=3",xlab="Periodo",ylab="Consumo m3")
plot(sociosTS,main="Serie de tiempo de usuarios mensuales",xlab="Periodo",ylab="Socios")
plot(ma(sociosTS,3),main="Media movil de usuarios mensuales\nk=3",xlab="Periodo",ylab="Socios")

```

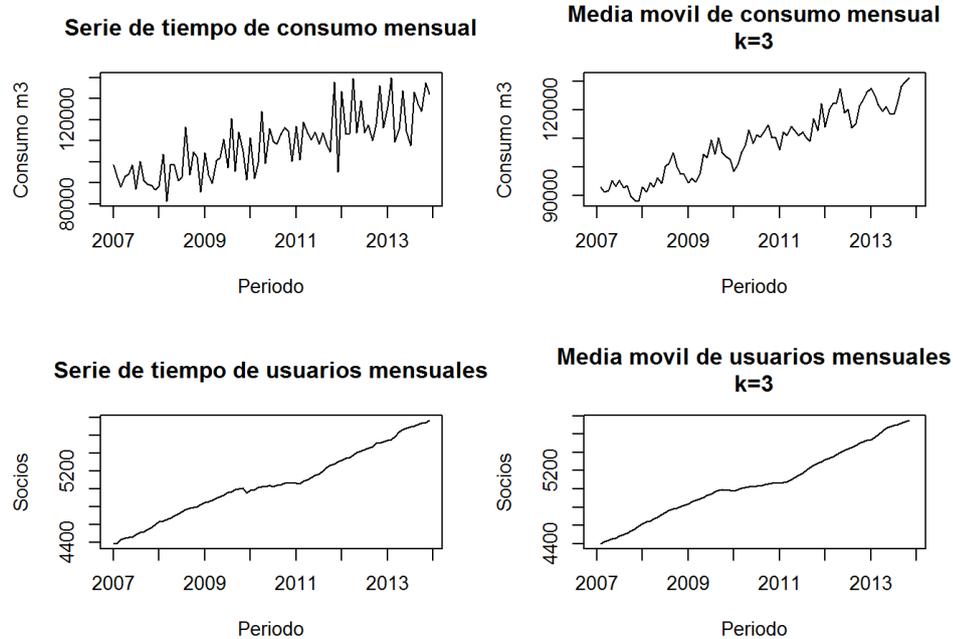


Figura 76 Series de tiempo para consumo y cantidad de usuarios
Fuente. Análisis del investigador

En la figura 76 se observa en la parte izquierda tanto los consumos como la cantidad de socios en cada período desde el año 2007, mientras que en la parte derecha se pueden ver los mismos datos, pero suavizados.

La conclusión a la que se llega es que definitivamente existe una tendencia creciente, sobre todo en el caso de los usuarios. En el caso de los consumos la tendencia existe, pero se ve afectada por grandes variaciones, aparentemente aleatorias, que en realidad puede significar que se trata de un fenómeno estacional.

Análisis estacional

Debido a que se observan grandes variaciones en el consumo mensual de agua, y se presume la existencia de un fenómeno estacional; se debe realizar una desagregación de la serie de tiempo en sus diversos componentes: estacional, tendencia y aleatorios.

Se pueden visualizar estos patrones al graficar el objeto devuelto por la función `stl()`, que ha sido ejecutado con el parámetro de predicción estacional "periodic".

```
#análisis de los datos descompuestos en sus elementos estacionales, tendencia y ruido
plot(stl(consumoTS,s.window = "periodic"),main = "Descomposicion de serie temporal del consumo")
```

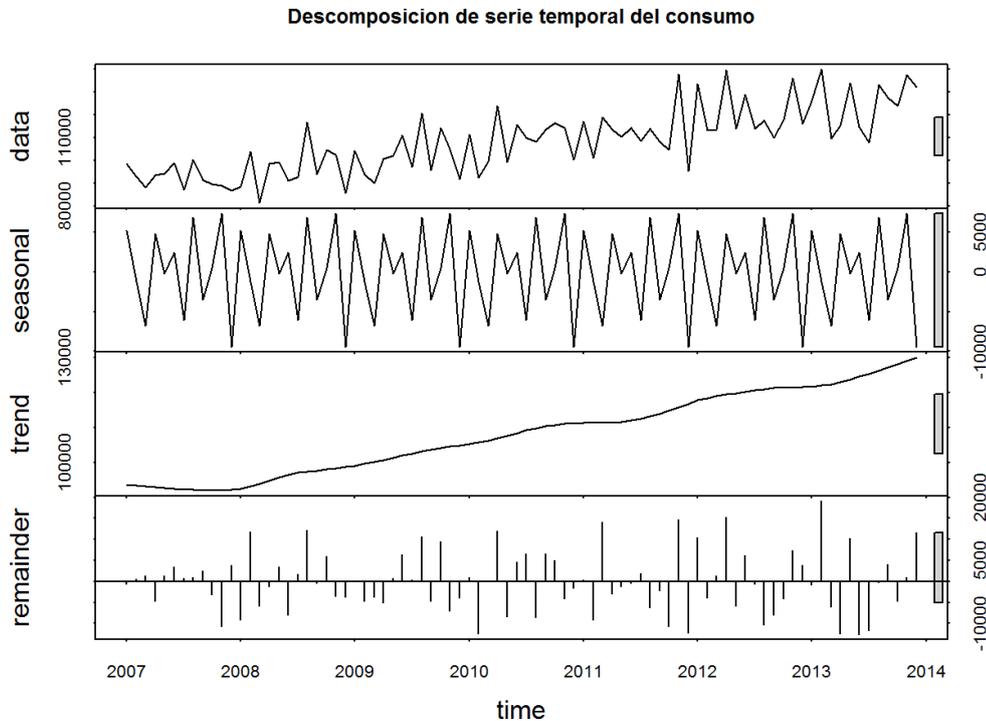


Figura 77 Descomposición de una serie de tiempo de consumos de agua potable
Fuente. Análisis del investigador

La gráfica descompuesta de los patrones existentes en el consumo de agua potable mensual, muestra en la primera parte los datos de originales, en la siguiente parte está la componente estacional periódica, luego está la componente de tendencia de los datos y finalmente los valores aleatorios restantes.

Se observa la existencia de una tendencia creciente claramente marcada, asimismo la componente estacional muestra las típicas variaciones de incremento y disminución que suceden en ciertos meses del año, tal como se pudo ver en los gráficos anteriores, en la sección de análisis exploratorio.

Predicciones con medias móviles y series de tiempo

Al haber analizado el comportamiento de las variables consumo y socios, se concluye que el modelo de series de tiempos puede ser adecuado, de tal manera que se puede utilizar para la predicción la función `SMA()`, que en ejemplo calcula el promedio móvil

de 3 observaciones en el *dataframe* denominado *smaConsumo* y *smaSocios*; estos datos se utilizan para realizar la predicción mediante la función `forecast()`, en la que el número 12 indica la cantidad de períodos que serán calculados. Los resultados se almacenan en el *dataframe* *consumoPry* y *sociosPry*³.

```
par(mfrow=c(1,1))
smaConsumo=SMA(consumoTS,3)
consumoPry=forecast(smaConsumo,12)
```

```
#Proyeccion de usuarios
smaSocios=SMA(sociosTS,3)
sociosPry=forecast(smaSocios,12)
```

Figura 78 Script R para aplicar suavizado mediante medias móviles
Fuente. Análisis del investigador

Para graficar los resultados, el script de la figura 79 utiliza la función `plot()` con el parámetro 12, parámetro que significa incluir esta cantidad de períodos previos en la imagen generada.

Se puede ver que, en el caso del pronóstico del consumo, el resultado es un valor fijo de 130.600m³ mensuales, pero el intervalo de valores posibles en los que el consumo real puede estar varía de entre 122.000 a 140.000 para el primer período pronosticado, incrementándose el valor conforme se pronostica eventos más lejanos. Lo anterior se lo observa en la imagen sombreada externa, que es el intervalo con una confianza del 95%, mientras que la sombreada interna implica una confianza del 80%. Esta gran zona de posibles valores en los que el consumo real pueda caer se debe a la gran variabilidad de los datos, por lo que habrá que aplicar algunas mejoras al modelo final para disminuir esta franja.

```
plot(sociosPry,include = 12,main = "Proyeccion de sociospara el año 2014",xlab = "Período",ylab = "Socios")
```

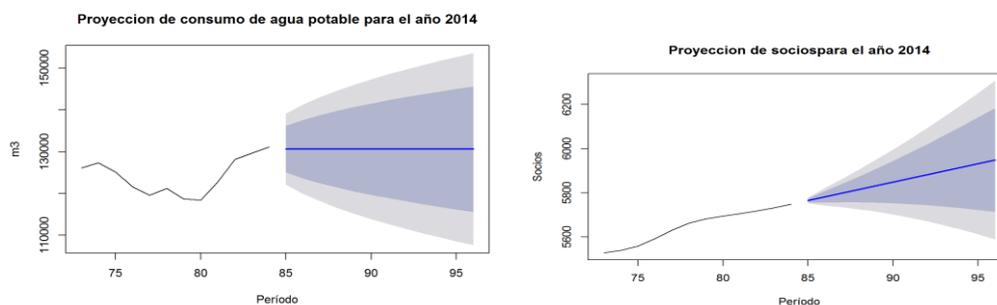


Figura 79 Proyección de consumo de agua potable y cantidad de usuarios

³ Se usa *pry* como indicativo de “proyectados”

El pronóstico de los usuarios es mucho mejor debido a la tendencia estable de su incremento. Es probable que los valores pronosticados estén dentro de un rango menor; así, en el siguiente período se pronostica una cantidad de 5.765 usuarios, con un intervalo de entre 5.753 a 5.778 y una confianza del 95%. El intervalo de confianza se va incrementando con el tiempo, puesto que es razonable que mientras más alejado sea el valor a pronosticar, mayor será la incertidumbre.

4.7.3. Construcción del modelo

Considerando los resultados anteriores, se concluye que el modelo para el pronóstico del consumo de agua potable debe estar compuesto por un análisis de series temporales, mientras que para la cantidad de socios, sería adecuado el modelo de medias móviles.

Para mejorar la precisión del análisis, se han realizado las siguientes consideraciones:

- Existen ciertos valores atípicos que deben ser quitados de los datos de entrenamiento, para esto se aplica la consideración de que el 75% de las observaciones están entre la media y +/- 2 desviaciones estándar, por lo que al eliminar los valores superiores a 2 desviaciones estándar por sobre la media, estaremos utilizando el 87,5% de los datos.
- Como los datos varían mucho entre consumidor y consumidor, lo mejor será realizar un análisis individualizado para cada usuario y el valor resultante totalizar para obtener el pronóstico general
- Los valores en cero se omiten debido a que normalmente esto significa que el usuario tiene únicamente un derecho, pero no una instalación real, además, si un usuario normal tiene consumo cero, frecuentemente se debe a que el medidor se ha dañado, por tanto no refleja lo que realmente debería consumir.

Con estas consideraciones se ha procedido a diseñar 2 modelos, el primero es un modelo que analiza y grafica un conjunto de usuarios, los que se pasan como parámetros en la variable "codigo". Cada código será analizado tomando los datos completos de consumos que se encuentran en la variable "consumos"

```

#Función para análisis y gráficación de datos
seriesIndividuales=function(codigo, consumos)
{
  #extracción de datos de interes del dataset
  for(i in 1:length(codigo) )
  {
    consumosU=consumos[consumos$codigo==codigo[i]&consumos$consumo>0,names(consumos)[c(1,2,6,8,9,12,13,15,16,17,18,19)]]
    print(paste("Datos encontrados para",codigo[i], NROW(consumosU) ) )
    maximo=mean(consumosU$consumo)+2*sd(consumosU$consumo)
    minimo=mean(consumosU$consumo)-2*sd(consumosU$consumo)
    consumosU=consumosU[consumosU$consumo>minimo&consumosU$consumo<maximo,]
    print( paste("Datos sin valores extranios para",codigo[i], NROW(consumosU) ) )
    cate=consumosU$tipo[1]
    anio=as.integer(format(consumosU[1,1],"%Y"))
    mes=as.integer(format(consumosU[1,1],"%m"))
    consumoTS=ts(consumosU$consumo,start=c(anio,mes),frequency=12)
    par(mfrow=c(2,1))
    plot(consumoTS,main=paste("Serie de tiempo de consumo mensual\ncodigo ",codigo[i], " ", cate),xlab="Periodo",
ylab="Consumo m3")
    plot(ma(consumoTS,3),main=paste("Media movil de consumo mensual\nk=3 codigo ",codigo[i], " ", cate),xlab="Per
iodo",ylab="Consumo m3")
    smaConsumo=SMA(consumoTS,12)
    consumoPry=forecast(smaConsumo,12)
    plot(consumoPry,include = 12,main =paste( "Proyeccion de consumo de agua potable para el año 2014\ncodigo "
,codigo[i], " ", cate) ,xlab = "Periodo",ylab = "m3")
    dfp=as.data.frame ( consumoPry)
    consumoTS=ts(round(dfp$`Point Forecast`,0),start=c(2014,1),frequency=12)
    plot(consumoTS,main=paste("Consumo mensual proyectado para el año 2014\ncodigo ",codigo[i], " ", cate),xlab
="Periodo",ylab="Consumo m3")
    print("Valores proyectados para los últimos meses")
    print(tail(as.data.frame ( consumoPry),3))
    print("Valores estadísticos principales")
    print(summary(consumosU$consumo))
    print(paste("Desviación estandar: ",sd(consumosU$consumo)) )
    print(paste("Coeficiente de variación: ",round( sd(consumosU$consumo)/mean(consumosU$consumo)*100,0),"%" ))
  }
}

#Proceso de proyección para un usuario residencial y un comercial

seriesIndividuales(c(1243,279),consumos)

```

Figura 80 Modelo de estimación de consumo futuro con análisis gráfico
Fuente. Análisis del investigador

El resultado de aplicar el modelo anterior a los códigos 1243 y 279, que corresponden a un usuario residencial y uno comercial (hostería), se muestran en la siguiente figura.

```
## [1] "Datos encontrados para 1243 73"  
## [1] "Datos sin valores extraños para 1243 70"
```

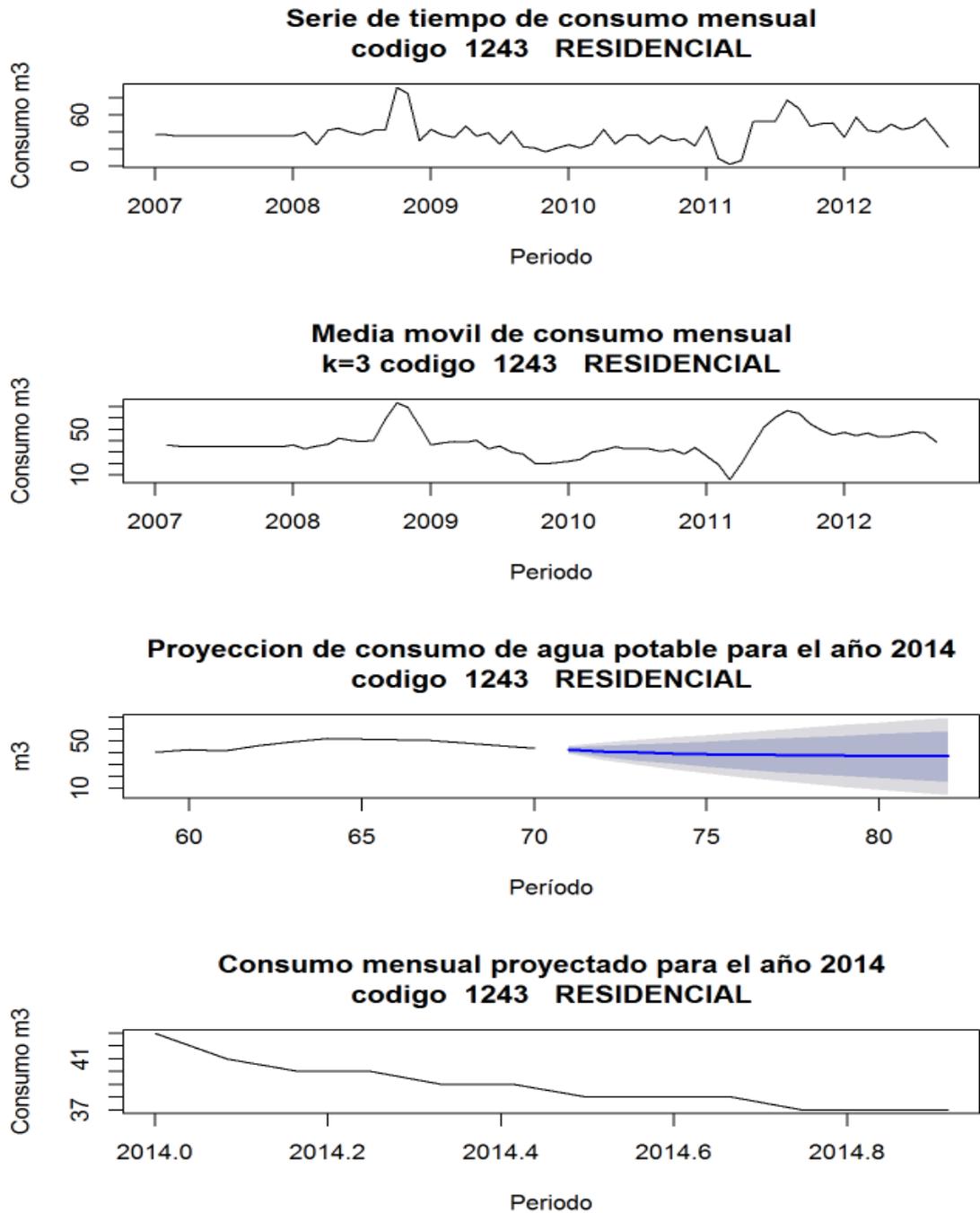


Figura 81 Resultados de modelo gráfico de estimación de consumo de agua potable para un usuario residencial
Fuente. Análisis del investigador

```

## [1] "Valores proyectados para los últimos meses"
##   Point Forecast  Lo 80  Hi 80  Lo 95  Hi 95
## 80      37.43638 18.89186 55.98090 9.074983 65.79778
## 81      37.27439 17.34327 57.20552 6.792365 67.75642
## 82      37.14480 15.88418 58.40541 4.629491 69.66010
## [1] "Valores estadísticos principales"
##   Min. 1st Qu. Median Mean 3rd Qu. Max.
## 2.00 30.50 36.50 38.17 43.00 93.00
## [1] "Desviación estandar: 15.4328573143838"
## [1] "Coeficiente de variación: 40 %"
## [1] "Datos encontrados para 279 84"
## [1] "Datos sin valores extraños para 279 83"

```

Figura 82 Valores estimados de consumo de agua potable para un usuario residencial
Fuente. Análisis del investigador

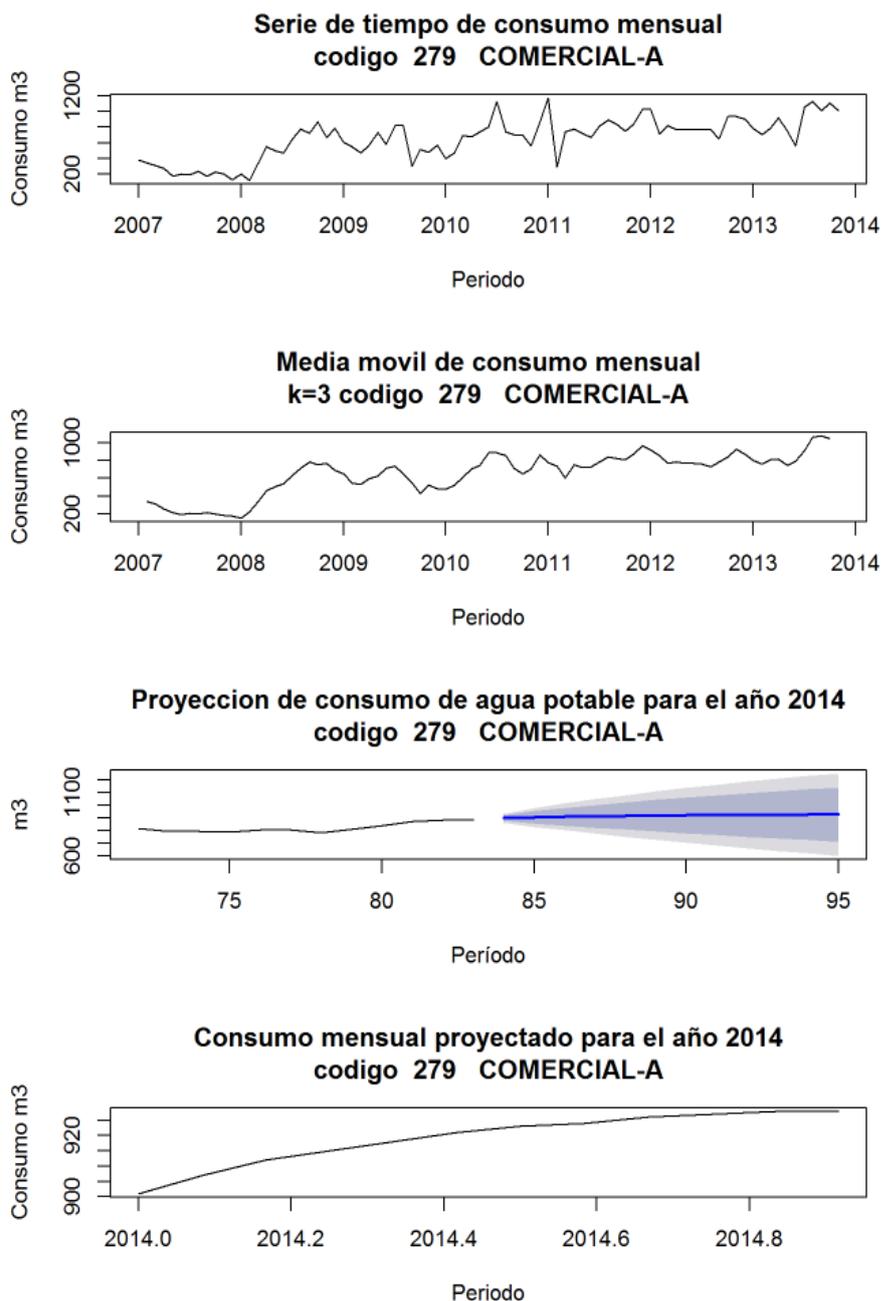


Figura 83 Resultados de modelo gráfico de estimación de consumo de agua potable para un usuario comercial
Fuente. Análisis del investigador

```

## [1] "Valores proyectados para los últimos meses"
##   Point Forecast   Lo 80   Hi 80   Lo 95   Hi 95
## 93      926.7392 740.9664 1112.512 642.6243 1210.854
## 94      927.5517 728.0033 1127.100 622.3688 1232.735
## 95      928.2024 715.4442 1140.961 602.8168 1253.588
## [1] "Valores estadísticos principales"
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 122.0  477.0   714.0   652.1  813.5  1173.0
## [1] "Desviacion estandar: 264.545912494992"
## [1] "Coeficiente de variacion: 41 %"

```

Figura 84 Valores estimados de consumo de agua potable para un usuario residencial
Fuente. Análisis del investigador

La primera parte del modelo indica con cuantos datos se están trabajando, luego se realiza el proceso de eliminación de valores atípicos superiores a 2 desviaciones estándar y los consumos en cero. Se imprime a continuación la cantidad de datos disponibles para realizar la proyección.

En el resultado gráfico se observa la serie de tiempo de datos originales, luego la misma serie de tiempo, pero aplicado un factor de suavización.

El siguiente gráfico muestra una parte de los datos finales de la serie más los datos pronosticados, en donde se aprecia el intervalo de confianza.

En la última imagen están los valores pronosticados para cada periodo, en este caso los 12 meses del año 2014.

El modelo imprime al final una tabla con los valores numéricos del pronóstico e intervalos de confianza de los últimos 3 meses; también se imprime otras variables estadísticas generales como la media, los cuartiles, la desviación estándar y el coeficiente de variación.

Modelo general

El modelo general no presenta resultados intermedios ni gráficos, sino que va formando un conjunto de datos (dataframe) en donde se almacena el valor pronosticado para cada mes, junto con los valores de promedio y desviación estándar de cada resultado.

```

#modelo de predicción de consumo mediante series de tiempo para un conjunto de usuarios
proyeccionesIndividuales=function(codigo,consumos,anioP)
{
  #extracción de datos de interés del dataset
  df1 = data.frame(codigo=character(),anio=numeric(),mes=numeric(),proyeccion=numeric(), media=numeric(),desvS=
numeric() )
  for(i in 1:length(codigo) )
  {
    if( (i/length(codigo)*100)%5==0 )
    {
      print(paste("procesado: ",i/length(codigo)*100))
    }
    consumosU=consumos[consumos$codigo==codigo[i]&consumos$consumo>0,names(consumos)[c(1,2,6,8,9,12,13,15,16,17
,18,19)]]
    filas=NROW(consumosU)
    if(filas>0)
    {
      if(filas>3)
      {
        maximo=mean(consumosU$consumo)+2*sd(consumosU$consumo)
        minimo=mean(consumosU$consumo)-2*sd(consumosU$consumo)
        consumosU=consumosU[consumosU$consumo>minimo&consumosU$consumo<maximo,]
        filas=NROW(consumosU)
      }
      if(filas>0)
      {
        anio=as.integer(format(consumosU[1,1],"%Y"))
        mes=as.integer(format(consumosU[1,1],"%m"))
        consumoTS=ts(consumosU$consumo,start=c(anio,1),frequency=12)
        datos=NROW(consumoTS)
        if(datos>11)
        {
          datos=11
        }
        media = round(mean(consumosU$consumo),0)
        des = round(sd(consumosU$consumo),0)
        if(datos>=3)
        {
          smaConsumo=SMA(consumoTS,datos)
          consumoPry=forecast(smaConsumo,12)
          dfp=as.data.frame( consumoPry)
          consumoTS=ts(round(dfp$`Point Forecast`,0),start=c(anioP,1),frequency=12)
          for(j in 1:length(consumoTS) )
          {
            df = data.frame(codigo=codigo[i],anio=anioP,mes=j,proyeccion=dfp$`Point Forecast`[j], media=media,d
esvS=des )
            df1=rbind(df1,df)
          }
        }
      }
      else
    }
  }
  print(paste("Solo media para " , codigo[i]))
  for(j in 1:12 )
  {
    df = data.frame(codigo=codigo[i],anio=anioP,mes=j,proyeccion=media, media=media,desvS=des )
    df1=rbind(df1,df)
  }
}
else
{
  print(paste("no hay datos luego de filtro para",codigo[i]))
}
}
else
{
  print(paste("no hay datos para",codigo[i]))
}
}
#print(summary(consumosU$consumo))
#print(paste("Desviacion estandar: ",sd(consumosU$consumo) ) )
#print(paste("Coeficiente de variacion: ",round( sd(consumosU$consumo)/mean(consumosU$consumo)*100,0),"%" )
)
}
return(df1)
}
#fin funcion de prediccion

#aplicación del modelo de predicción a un par usuarios
dfp=proyeccionesIndividuales(c(1243,279),consumos,2014)

```

Figura 85 Modelo general de estimación de consumo de agua potable para un listado de usuarios
Fuente. Análisis del investigador

Para ejemplificar, se ha aplicado el modelo para los 2 casos utilizados en el modelo gráfico descrito anteriormente. Los resultados se muestran a continuación.

```
print("Valores proyectados individualmente para un conjunto de 2 usuarios")

## [1] "Valores proyectados individualmente para un conjunto de 2 usuarios"

print(head(dfp))

##   codigo  anio  mes  proyeccion  media  desvS
## 1  1243 2014   1  42.22451    38   15
## 2  1243 2014   2  41.09483    38   15
## 3  1243 2014   3  40.19105    38   15
## 4  1243 2014   4  39.46800    38   15
## 5  1243 2014   5  38.88954    38   15
## 6  1243 2014   6  38.42675    38   15

print(tail(dfp))

##   codigo  anio  mes  proyeccion  media  desvS
## 19  279 2014   7  986.3766    652  265
## 20  279 2014   8  999.7690    652  265
## 21  279 2014   9 1013.1614    652  265
## 22  279 2014  10 1026.5537    652  265
## 23  279 2014  11 1039.9461    652  265
## 24  279 2014  12 1053.3385    652  265
```

Figura 86 Resultados de aplicar el modelo de estimación de consumo de agua potable para un listado de usuarios
Fuente. Análisis del investigador

Modelo de análisis de ingresos

Este modelo analiza los elementos más importantes que proveen de ingresos a la entidad, que son el pago por consumo de agua y la venta de derechos.

El proceso comienza con la carga, limpieza, selección y transformación de datos obtenidos del sistema de facturación de la Junta.

En el script de la figura 87 se observa la aplicación de funciones similares a las aplicadas cuando se analizó el consumo de agua potable. Ha sido necesario cambiar el símbolo de decimales, que en los datos originales utilizaban la coma, por el punto; para ello se utiliza la función `gsub()`. Posteriormente la descripción de los cobros se agrupa dentro de 3 categorías generales, esto debido a que para una misma categoría el documento de facturación tiene denominaciones diferentes.

En la parte final del script se aplica una agrupación de datos basados en las categorías, lo que permite graficar tanto la evolución de los cobros y la tendencia de cada elemento.

```

pagos=pagos[pagos$fecha < "2014-01-01",]
pagos$periodo = paste(format(pagos$fecha,"%Y"),format(pagos$fecha,"%m"),"01",sep="/")
pagos$pagoMateriales=ifelse(is.null(pagos$pagoMateriales),0,pagos$pagoMateriales)
pagos$interes=ifelse(is.null(pagos$interes),0,pagos$interes)
pagos$otros=ifelse(is.null(pagos$otros),0,pagos$otros)
pagos$pagoMateriales = gsub(",",".",pagos$pagoMateriales)
pagos$interes = gsub(",",".",pagos$interes)
pagos$otros = gsub(",",".",pagos$otros)
pagos$pagoMateriales=as.numeric(pagos$pagoMateriales)
pagos$interes=as.numeric(pagos$interes)
pagos$otros=as.numeric(pagos$otros)
pagos$valo = pagos$total -pagos$pagoMejoras - pagos$pagoMateriales - pagos$interes - pagos$descuento
pagos$categoria = substr(pagos$estatus,1,15)
pagos=mutate_each(pagos, funs(toupper))
pagos$categoria=ifelse(pagos$categoria=="PAGO INTEGRADO ","CONSUMO DE AGUA",pagos$categoria)
pagos$categoria=ifelse(pagos$categoria=="PAGO DE PLANILL","CONSUMO DE AGUA",pagos$categoria)
pagos$categoria=ifelse(pagos$categoria=="DERECHO: CAMBIO","CAMBIO DE NOMBR",pagos$categoria)
pagos$categoria=ifelse(pagos$categoria=="TRASPASO DE USU","CAMBIO DE NOMBR",pagos$categoria)
pagos$categoria=ifelse(pagos$categoria=="DERECHO DE INST","DERECHO DE AGUA",pagos$categoria)
pagos$valo=as.numeric(pagos$valo)
resumenPagos = pagos %>% group_by(categoria, periodo) %>% summarize(valor=sum(valor))
resumenPagos$periodo = as.Date(resumenPagos$periodo)
resumenPagos=resumenPagos[resumenPagos$categoria!="FACTURA POR VEN" & resumenPagos$categoria!="COBRO MEJORAS -"
& resumenPagos$categoria!="NOTA DE DEBITO ",]
ggplot(resumenPagos,aes(periodo,valor,group=categoria,colour=categoria)) + geom_line() +geom_point() + ggtitle("Pagos Mensuales") + geom_smooth()

```

Figura 87 Script para la carga y tratamiento de datos de cobros
Fuente. Análisis del investigador

La figura 88 muestra que mensualmente el pago por consumo de agua se ha incrementado desde el año 2001 de \$ 5.000 a \$35.000 en 2014. Se puede observar además que desde el 2009 existe una tendencia relativamente estable, pero inferior a los períodos anteriores.

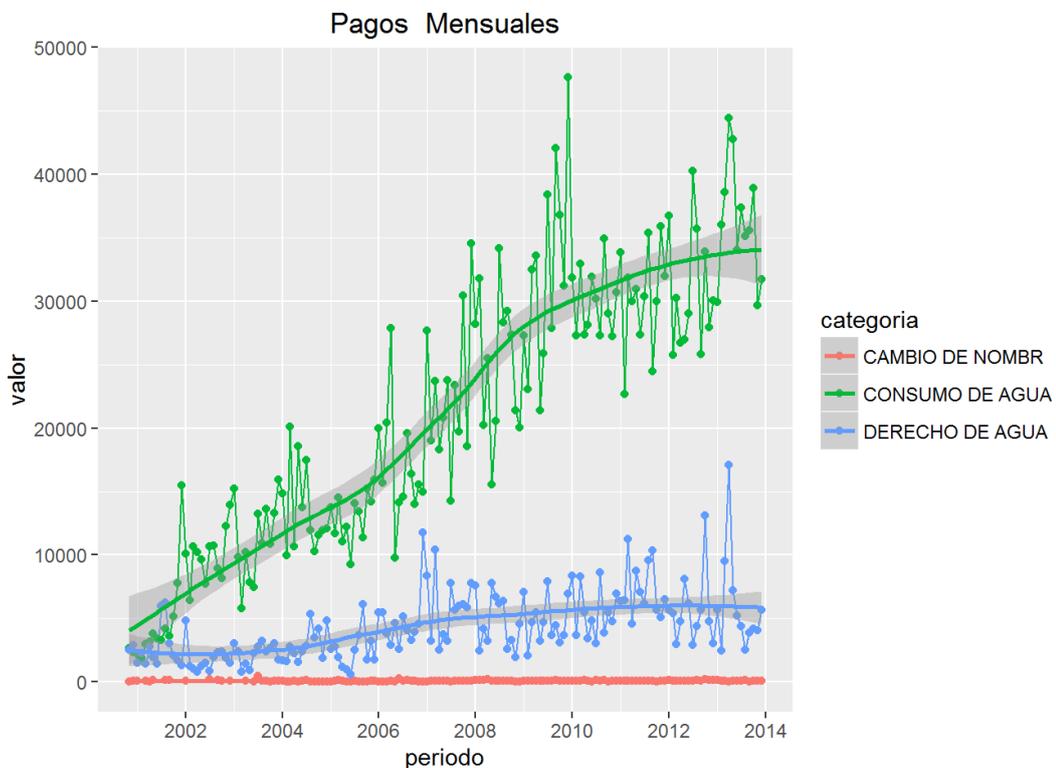


Figura 88 Evolución y tendencia de cobros agrupados por categoría

Fuente. Análisis del investigador

En cuanto al comportamiento de las ventas de derechos, se observa que desde el año 2007 prácticamente se estabiliza la tendencia, con ingresos cercanos a los \$5.000 mensuales.

Es importante desgajar la información anterior de acuerdo al tipo de los usuarios para saber la contribución que hace cada grupo, recordemos que los usuarios comerciales pagan un costo elevado por cada m³; para ello se filtran los datos correspondientes al consumo de agua y se grafica agrupándolo según el tipo de usuario.

```
#pagos mensuales de agua por tipo de usuario
resumenPagos = pagos %>% group_by(categoria, periodo, tipo) %>% summarize(valor=sum(valor))
resumenPagos=resumenPagos[resumenPagos$categoria=="CONSUMO DE AGUA",]
resumenPagos$periodo = as.Date(resumenPagos$periodo)
ggplot(resumenPagos, aes(periodo, valor, group=tipo, colour=tipo)) + geom_line() + geom_point() + ggtitle("Pagos Mensuales por tipo de usuario") + geom_smooth()
```

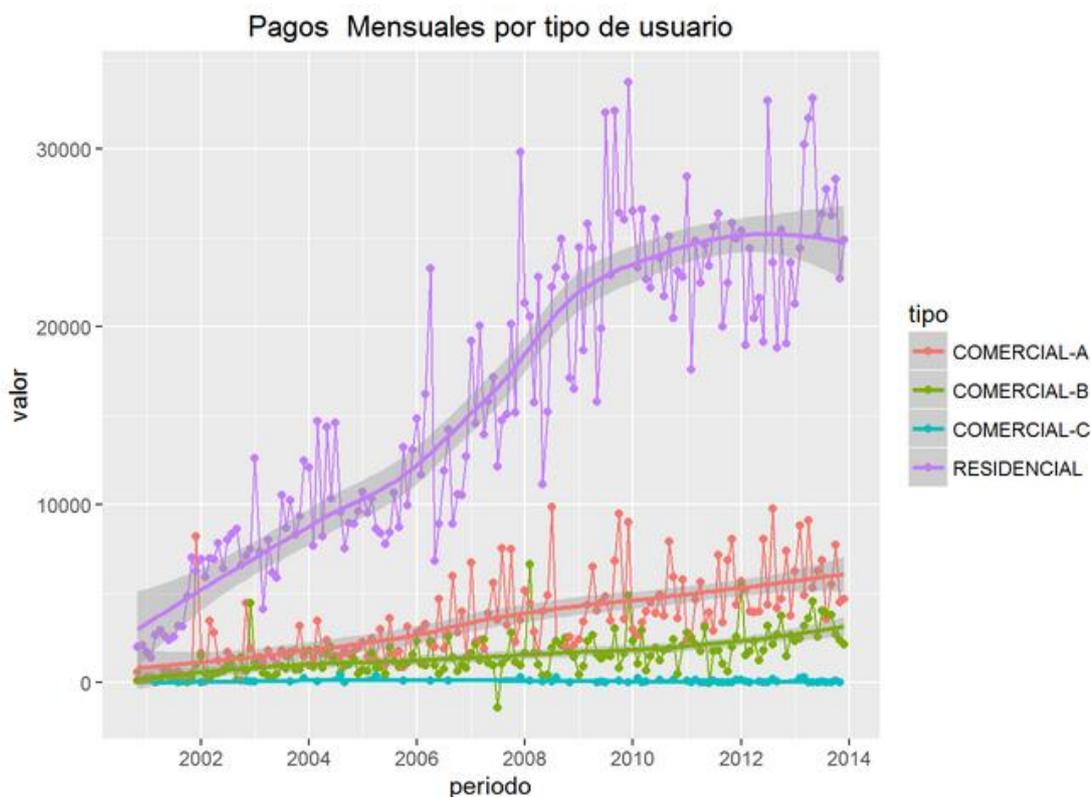


Figura 89 Script y gráfica de pagos por consumo de agua agrupados por tipo de usuario
Fuente. Análisis del investigador

La figura 89 evidencia que los usuarios residenciales son los que pagan más por el consumo de agua, con una tendencia estable horizontal desde 2011. Durante los años 2006 al 2008 se muestra un gran incremento en los valores pagados; la explicación es que en ese período se realizó una gran campaña de cobros de valores

atrasados. Actualmente los usuarios pagan a tiempo, teniendo como máximo 3 meses de gracia, caso contrario se procede con la suspensión del servicio.

Conociendo la cantidad de usuarios por cada categoría, es posible pronosticar los ingresos al multiplicarlos por los valores medios de pago por categoría.

Esto último se logra con el siguiente script que selecciona la categoría residencial y grafica su evolución. Si bien aparece que los pagos disminuyen, en realidad no ha sido preocupante este hecho, debido a que se trata de que antes los usuarios dejaban acumular varios meses y pagaban valores altos, pero ahora al pagar puntualmente, el pago medio ha disminuido; sin embargo, el ingreso total crece de acuerdo a la cantidad de usuarios.

```
#pagos promedio mensual de categoria residencial
resumenPagosPromedio=resumenPagosPromedio[resumenPagosPromedio$tipo=="RESIDENCIAL" ,]
ggplot(resumenPagosPromedio,aes(periodo,valor)) + geom_line() +geom_point() + ggtitle("Pagos mensuales promedi
os por categoria residencial") + geom_smooth()
```

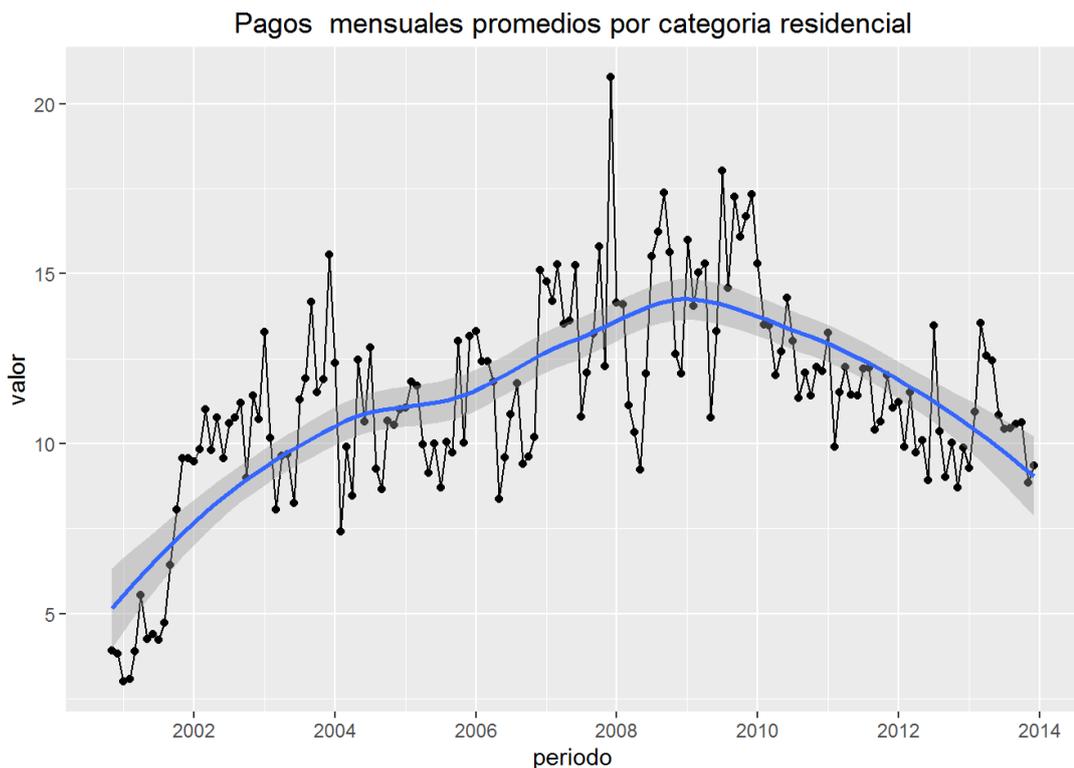


Figura 90 Script y gráfica de pagos mensuales promedio por consumo de agua tipo residencial
Fuente. Análisis del investigador

4.7.4. Informe

La generación del informe del proceso y aplicación de los diversos análisis, así como del modelo propuesto, se ha realizado como se indicó al inicio, utilizando RMarkdown, cuya compilación produce el informe que se muestra a continuación.

The image shows a screenshot of an RMarkdown report. At the top, the title is "Análisis de Datos de Consumo de agua potable" and the author is "Ing. Hernan Quiño MAE". The date is "10 de diciembre de 2016". The report is divided into sections, with the first being "1. Carga de datos". This section contains R code for loading CSV files: "usuarios", "consumos", "pagos", "consumoPeriodo", and "consumoPeriodoTipo". It also includes code for preparing the data, such as converting columns to factors and dates. The code is followed by a warning message: "Warning: package 'ggplot2' was built under R version 3.2.3". The report also shows the loading of the "scales" and "dplyr" libraries. The bottom left corner of the report shows "1 de 64" and the bottom right corner shows "13/6/17 14:32".

Figura 91 Informe general en HTML generado con RMarkdown
Fuente. Análisis del investigador

La figura 91 muestra el informe generado en formato HTML que consta de 64 páginas. En este informe se detallan análisis adicionales tanto exploratorios como de tendencia con una gran variedad de gráficos y el código que lo genera.

Es importante la posibilidad de incluir el código, puesto que esto permitirá que personas interesadas en el tema puedan validar esta investigación. Este es uno de los requerimientos cuando se aplica el análisis científico de datos, pues es necesario evitar que falsos experimentos muestren descubrimientos aparentemente brillantes, como ya ha sucedido en algunos sonados casos.

4.8. Elaboración del presupuesto anual

4.8.1. Análisis y esquema de elaboración de presupuesto

Para la realización del presupuesto, se debe considerar las dependencias entre los distintos presupuestos:

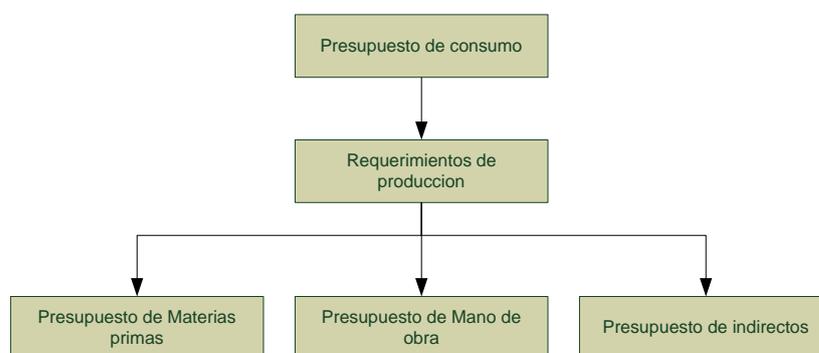


Figura 92 Diagrama de presupuesto de consumo y producción

Fuente. Diseñado por el investigador

De acuerdo al esquema anterior, el presupuesto se basa en determinar el consumo esperado de agua potable, con el que es posible determinar los requerimientos de materias primas necesarias para el proceso de potabilización.

En la realización del pronóstico de consumo de agua potable puede aplicarse las técnicas basadas en grandes cantidades de datos, como en las secciones anteriores se ha demostrado,

Además de lo anterior, un presupuesto depende de las actividades que cada una de las áreas que conforman una empresa planeen implementar en el siguiente período. Esta parte no puede ser pronosticada, y se requiere de reuniones con los responsables para poder determinar y presupuestar estas actividades.

4.8.2. Incidencia del Plan Operativo

Una gestión eficiente involucra definir las actividades que se van a desarrollar en un determinado período, así como los medios y recursos necesarios para alcanzarlas.

Lo anterior implica que el presupuesto estará influenciado por el costo de la realización de las actividades durante el período planificado, de tal manera que se requiere realizar un proceso de obtención de este tipo de datos.

El presupuesto financiero parte del análisis de los periodos anteriores para identificar la tasa de variación interanual, esto se logra con el análisis horizontal.

El presupuesto proyectado necesita además de la aplicación de un análisis vertical, esto para poder establecer la relación de proporcionalidad entre las variables clave como ingreso y cantidad de consumo, que pueden ser pronosticadas, con otras variables como el costo de materia prima, que no se puede pronosticarse directamente.

Los gastos de la JAAPB se han clasificado de acuerdo a su situación de recurrencia debida a las actividades normales de la Junta y de acuerdo a las nuevas actividades planificadas. Esto se puede observar en el presupuesto que se encuentra en la siguiente sección.

El presupuesto considera ingresos y egresos, tanto recurrentes como nuevos e incluye las inversiones planificadas para este año.

Análisis vertical de los informes financieros: años 2012- 2013

A continuación, se presenta la información financiera de la JAAPB, para lo cual se utilizará en este inciso los Estados Financieros evaluados en porcentajes y valores de los períodos 2012 y 2013. Estos datos nos sirven de base para el presupuesto del año 2014. El análisis de variación interanual permitirá estimar los valores totales esperados para el siguiente período y utilizando la composición porcentual del análisis vertical de cada cuenta, se logra determinar el detalle de cuentas del informe financiero proforma del año 2014 a través de mantener la relación porcentual de cada cuenta sobre el total.

AÑOS	2012	ANA.VER 2012	2013	ANA.VER 2013
ACTIVO CORRIENTE				
Caja, Bancos	137.451,66	14,47	280.964,88	21,15
Caja	4.240,66	0,45	811,71	0,06
Banco del Pichincha	3.200,10	0,34	5.362,07	0,40
Cooperativas.	130.010,90	13,69	274.791,10	20,69
Otras Cuentas x Cobrar	26.944,43	2,84	27.178,59	2,05
Ventas Bienes y Servicios	23.291,22	2,45	18.785,93	1,41
Anticipos	3.653,21	0,38	8.392,66	0,63
Créd. Tribut. a Favor de la Empresa (I.Renta)	147.044,64	15,48	-	-
IVA Crédito Fiscal en Compras	147.044,64	15,48	-	-
Retenciones en la Fuente Imp.Renta		-	-	-
Invent. Repuest., Herramient. y Accesorios	15.155,57	1,60	27.613,63	2,08
Accesorios Recuperables	1.664,40	0,18	2.775,01	0,21
Accesorios No Recuperables	10.050,24	1,06	21.101,01	1,59
Químicos	2.756,77	0,29	3.266,77	0,25
Herramientas	684,16	0,07	470,84	0,04
TOTAL ACTIVO CORRIENTE	326.596,30		335.757,10	
ACTIVO FIJO				
BIENES ADMINISTRACION	161.002,71	16,95	413.250,97	31,11
Mobiliarios	22.055,81	2,32	7.682,50	0,58
Maq. y Equipos	25.370,17	2,67	26.531,69	2,00
Eq.Sist. Y Paq.Informat.	23.731,04	2,50	8.527,60	0,64
Terrenos	65.075,00	6,85	266.759,38	20,08
Edif. Locales y Resid.	24.770,69	2,61	103.749,80	7,81
BIENES PRODUCCION	801.505,48	48,67	914.370,84	43,60
Maquinaria y Equipos	147.545,33	15,53	136.239,44	10,26
Vehículos	54.823,88	5,77	34.468,31	2,60
Planta Tratam. Y Redes	599.136,27	63,07	743.663,09	55,99
(-)Depreciac. Acumulada Activo Fijo	- 339.100,56	- 35,69	- 335.297,13	- 25,24
OBRAS INFRAESTRUCTURA	-	-	-	-
INVERSIONES LARGO PLAZO			112,42	0,01
Certif. Aportación Cooperativas		-	112,42	0,01
TOTAL ACTIVO FIJO	623.407,63		992.437,10	
TOTAL ACTIVO	950.003,93	100,00	1.328.194,20	100,00

Tabla 9 Análisis comparativo vertical de activos
Fuente. Análisis del investigador

AÑOS	2012	ANA. VER 2012	2013	ANA. VER 2013
PASIVO CORRIENTE				
Obligaciones con la Administ. Tributaria	809,60	0,09	2.261,69	0,17
Retenciones en la Fie.Imp.Renta	265,40	0,03	799,06	0,06
Retenciones IVA	544,20	0,06	1.462,63	0,11
Obligaciones con el IESS	1.693,92	0,18	2.185,67	0,16
Aporte Patronal	893,95	0,09	1.194,04	0,09
Aporte Individual	799,97	0,08	991,63	0,07
Fondo de Reserva	-	-	-	-
Obligaciones con Terceros	4.048,17	0,43	2.478,77	0,19
Diets	3.630,60	0,38	2.061,20	0,16
Ctas. x Pagar Proveedores	417,57	0,04	417,57	0,03
Obligaciones con Empleados	9.048,80	0,95	8.764,51	0,66
Liquidación Luis Lucero	-	-	-	-
Liquidación Ec. Pablo Alemán	-	-	-	-
XIII Sueldo	662,08	0,07	829,42	0,06
XIV Sueldo	1.824,75	0,19	2.449,48	0,18
Vacaciones	6.561,97	0,69	5.485,61	0,41
OTROS PASIVOS				
Ingresos Anticipados	15.661,73	1,65	15.438,01	1,16
Anticipos x Bienes y Servicios	15.661,73	1,65	15.438,01	1,16
PASIVO NO CORRIENTE				
Obligaciones Largo Plazo			63.078,46	4,75
Bonificación por Deshaucio			17.336,50	1,31
Jubilación Patronal			45.741,96	3,44
TOTAL PASIVO	31.262,22		94.207,11	
PATRIMONIO		96,71		92,91
Patrimonio Jta. Admin. A.P. Baños	93.554,66	9,85	93.554,66	7,04
Aportes de Socios Futura Capitalizacion	29.225,97	3,08	29.225,97	2,20
Resultados Ejercicios Anteriores	726.740,07	76,50	589.348,75	44,37
Superavit por Valuación			391.014,36	29,44
Utilidad del ejercicio	69.221,01	7,29	130.843,35	9,85
TOTAL PATRIMONIO	918.741,71		1.233.987,09	
TOTAL PASIVO Y PATRIMONIO	950.003,93	100,00	1.328.194,20	100,00

Tabla 10 Análisis comparativo vertical de pasivo y patrimonio
Fuente. Análisis del investigador

Los porcentajes que se muestran en el Estado de Situación Financiera están relacionados con el activo total para los años 2012 y 2013.

En el activo corriente podemos observar que el año base era 34,38% y el año 2013, 25,28%, existiendo un incremento del 9,10% observándose que en la cuenta Caja-Bancos radicó esta diferencia.

Las cuentas por cobrar representan el 2,84% y el 2,05% para los años 2012 y 2013 respectivamente, observándose una diferencia de 0,79% siendo un porcentaje mínimo que no representa inconvenientes.

Un renglón relevante de los activos es la cuenta activos fijos, bienes que cambian del año base desde un 16,95% al año 2013 con un 31,11%, dándose un incremento de 14,16%. Esta diferencia se debe a las inversiones que realizó la JAAPB durante ese período, así como a la actualización del valor de sus activos.

En los pasivos corrientes podemos observar que en la cuenta de obligaciones tributarias se incrementa en un 0.08%; este incremento se da en la cuenta de IVA retenciones a los clientes.

AÑOS	2012	ANA. VER 2012	2013	ANA. VER 2013
INGRESOS				
Vtas. Netas Gravadas con Tarifa Cero (Suministro de Agua)	461.798,96		531.767,71	
Derechos de Instalacion	66.960,00	14,23	71.825,00	13,25
Consumo de Agua	393.558,96	83,64	458.982,71	84,65
Traspaso de Usuario	1.280,00	0,27	960,00	0,18
Otros (Multas)	1.033,20	0,22	430,76	0,08
Devoluciones en Ventas	0,00	0,00	0,00	0,00
OTROS INGRESOS		0,00		0,00
Otros Ingresos no Especificados	0,00	0,00	0,00	0,00
Rendimientos Financieros (Intereses)	7.703,95	1,64	10.028,66	1,85
TOTAL INGRESOS	470.536,11	100,00	542.227,13	100,00
COSTOS Y GASTOS				
Gastos en Trabajadores	175.337,72	43,69	195.651,01	47,56
Sueldos, Horas Extras y Dietas	95.811,59	23,87	110.705,56	26,91
Remuneraciones Complementarias	26.050,29	6,49	39.121,93	9,51
Aportes IESS	19.496,64	4,86	21.774,80	5,29
Indemnizaciones	28.779,02	7,17	18.615,70	4,53
Vestuario y Prendas Protección	3.534,24	0,88	3.108,67	0,76
Alimentos y Bebidas	1.665,94	0,42	2.324,35	0,57
Contratacion de Servicios	54.332,42	13,54	54.871,42	13,34
Servicios de Profesionales	27.451,24	6,84	34.591,07	8,41
Servic.en Sistemas Informat.	4.200,00	1,05	3.500,00	0,85
Servicios de Auditoria	1.900,00	0,47	4.668,00	1,13
Capacitaciones	2.102,50	0,52	0,00	0,00
Energia Elect. Y Telecom.	5.998,31	1,49	6.015,23	1,46
Eventos de la Junta de Agua	3.882,91	0,97	50,00	0,01
Transporte	4.628,75	1,15	1.868,00	0,45
Difusion, Publicidad, Impres. Public.	3.262,63	0,81	3.673,45	0,89
Vigilancia, Aseo y Otros	906,08	0,23	505,67	0,12

Producción y Distrib. de Agua	106.697,66	26,59	51.535,98	12,53
Químicos	27.175,26	6,77	31.573,13	7,67
Accesorios Instalación	9.344,66	2,33	2.974,39	0,72
Accesorios Reparación	29.369,27	7,32	1.050,22	0,26
Alquiler Maquinaria y Equip.	4.792,34	1,19	2.811,02	0,68
Reparación de Tuberías	18.634,56	4,64	4.136,81	1,01
Materiales para Laboratorio	6.964,94	1,74	3.601,78	0,88
Combustibles y Lubricantes	4.110,78	1,02	4.517,26	1,10
Repuestos y Herramientas	6.305,85	1,57	871,37	0,21
Mantenimientos	9.567,41	2,38	9.910,54	2,41
Edificios	6.250,36	1,56	5.154,95	1,25
Maquinaria, Equip. Vehículos	3.317,05	0,83	4.755,59	1,16
Gastos en Oficinas y Otros	9.247,59	13,80	39.942,43	24,17
Materiales de Oficina	3.593,24	0,90	3.524,00	0,86
Insumos para Aseo	1.062,42	0,26	1.389,01	0,34
Bienes No Depreciables	1.536,52	0,38	7.359,25	1,79
Gastos Financieros Impuestos y Otros	746,99	0,19	24.220,74	5,89
Seguros	1.375,57	0,34	2.948,67	0,72
Arriendo de Mobiliarios e Inmuebles	932,85	0,23	500,76	0,12
Depreciación Activos Fijos	46.132,30	11,50	59.472,40	14,46
TOTAL COSTOS Y GASTOS	401.315,10	100,00	411.383,78	100,00
UTILIDAD DEL EJERCICIO ECONOMICO	69.221,01		130.843,35	

Tabla 11 Análisis vertical de ingresos y costos
Fuente. Análisis del investigador

Todos los porcentajes que se analizan están relacionados con las ventas netas que representan el total para los años 2012 y 2013. Como resultado de la comparación de los dos años se observa en el cuadro que las ventas para el año 2012, en comparación con el año 2013 se incrementaron en \$ 69.968,75.

En lo que respecta a los gastos de operación los mismos tuvieron un incremento en los gastos de oficina del 13,80 al 24,17% y una reducción en los gastos de producción y distribución del agua del 26,59% al 12,53% en comparación con el año 2012 al 2013 sin embargo al ser comparados en relación al incremento de ventas representa un aumento porcentual aceptable.

Las ventas del año 2012 fueron de \$ 461.798,96 y en el año 2013 fueron \$ 531.767,71 dándose un incremento de un 13%.

El superávit que se refleja en el año 2012 fue de \$69.221,01 y el del año 2013, \$ 130.843,35 observándose un incremento cercano al 47% con respecto al año base.

4.8.3. Actividades y plan presupuestario

A continuación, se detallan las actividades adicionales a realizar durante el año 2014, organizadas por cada departamento.

Los indicadores de gestión son los que permitirán a los jefes departamentales dar seguimiento al cumplimiento de cada actividad, que será evaluada de acuerdo a la meta establecida. Es importante recordar que las metas deben ser evaluadas periódicamente, para ajustar a la realidad en la que se encuentre la Junta, pudiendo por tanto ser modificada de presentarse argumentos valederos, en cuyo caso, es necesario comunicar los nuevos valores a la dirección administrativa.

En la siguiente hoja se encuentra el formato de levantamiento de datos que cada jefe departamental debe llenar. Este formato es una versión simplificada que permite considerar los costos que involucran llevar a cabo una determinada actividad del plan, de tal manera que se pueda estimar de forma más precisa los requerimientos financieros para ejecutar el plan.

El encabezado relaciona el objetivo estratégico con la acción específica a realizar, para luego determinar las inversiones, gastos y costos de operación que son necesarios para llevar a cabo esa acción.

El resto de campos permitirá elaborar las tablas de depreciaciones, que es parte importante de los balances proyectados.

Las siguientes tablas muestran la planificación de gastos por cada trimestre y las fuentes de financiamiento

MEDIOAMBIENTE										
ACCIONES	INDICADORES DE GESTION	META	PROGRAMACION				TOTAL PRESUPUESTO	FINANCIAMIENTO O JAAPB	FINANCIAMIENTO OTROS	OBSERVACIONES
			TRIMESTRAL (avance)							
			I	II	III	IV				
Campañas de concientización y capacitación sobre usos alternativos del suelo	CANTIDAD DE TALLERES A SER DICTADOS	4 TALLERES DICTADOS			50%	100%	\$ 644,00	\$ 429,33	\$ 214,67	FINANCIA SENAGUA
Ejecutar proyectos de conservación, forestación reforestación	PROYECTO EJECUTADOS	50% DE EJECUCION			20%	50%	\$ 1.983,33	\$ 991,67	\$ 991,67	FINANCIA SENAGUA
Ensayos (Insitu) de adaptación de especies nativas	NUMERO DE PLANTAS A LAS QUE SE APLICARON EL ENSAYO	200 PLANTAS			20%	100%	\$ 408,33	\$ 408,33	\$ -	
						TOTAL	\$ 3.035,67	\$ 1.829,33	\$ 1.206,33	

Tabla 12 Costos de implementación de actividades medioambientales

Fuente. Análisis del investigador

PRODUCCION										
ACCIONES	INDICADORES DE GESTION	META	PROGRAMACION				TOTAL PRESUPUESTO	FINANCIAMIENTO O JAAPB	FINANCIAMIENTO OTROS	OBSERVACIONES
			I	II	III	IV				
Adquisicion e Instalacion de Equipos (Macromedidores)	MACROMEDIDORES INSTALADOS	3 MACROMEDIDORES INSTALADOS				100%	\$ 4.159,88	\$ 4.159,88	\$ -	
Instalar el sistema de retrolavado	AVANCE DE LA CONSTRUCCION	20% DE LA CONSTRUCCION				20%	\$ 5.833,33	\$ 5.833,33	\$ -	
Reparacion de desarenador	AVANCE DE LA REPARACION	60% DE LA REPARACION				60%	\$ 875,00	\$ 875,00	\$ -	
Construccion del Tanque rompepresion	CONSTRUCCION DEL TANQUE	100% DE CONSTRUCCION			100%		\$ 3.500,00	\$ 3.500,00	\$ -	
Implementar y renovar equipos e instrumentos acordes a tecnología actual	PORCENTAJES EQUIPOS OBSOLETOS	20% DE EQUIPOS OBSOLETOS			40%	20%	\$ 1.219,17	\$ 1.219,17	\$ -	
Estudios de nueva planta - convenio Ecuador Estratégico	AVANCE DEL ESTUDIO	100% CUMPLIDO			50%	100%	\$ 112.000,00	\$ -	\$ 112.000,00	FINANCIA ECUADOR ESTRATÉGICO
						TOTAL	\$ 127.587,38	\$ 15.587,38	\$ 112.000,00	

Tabla 13 Costos de implementación de actividades de producción
Fuente. Análisis del investigador

DISTRIBUCION										
ACCIONES	INDICADORES DE GESTION	META	PROGRAMACION				TOTAL PRESUPUESTO	FINANCIAMIENTO O JAAPB	FINANCIAMIENTO OTROS	OBSERVACIONES
			I	II	III	IV				
Delimitar el area de cobertura	ESTUDIO DEL AREA DE COBERTURA	ESTUDIO CONCLUIDO			50%	50%	\$ 3.500,00	\$ 3.500,00	\$ -	
Diagnostico de las redes de agua potable	CONTRATAR UN ESTUDIO DE REDES	70% DE AVANCE			30%	70%	\$ 17.675,00	\$ 17.675,00	\$ -	
Reactivacion de los equipos de lecturacion	BRINDAR UN SERVICIO AGIL	100% DE LECTURAS CON EQUIPOS			20%	60%	\$ 175,00	\$ 175,00	\$ -	
						TOTAL	\$ 21.350,00	\$ 21.350,00	\$ -	

Tabla 14 Costos de implementación de actividades para distribución de agua
Fuente. Análisis del investigador

ADMINISTRACION											
ACCIONES	INDICADORES DE GESTION	META	PROGRAMACION				TOTAL PRESUPUESTO	FINANCIAMIENTO O JAAPB	FINANCIAMIENTO OTROS	OBSERVACIONES	
			I	II	III	IV					
Asesoría Legal permanente	EXISTENCIA DE REGLAMENTOS APROBADOS	REGLAMENTOS APROBADOS			60%	100%	\$ 1.166,67	\$ 1.166,67	\$ -		
Actualización de los reglamentos y políticas internas	EXISTENCIA DE REGLAMENTOS ACTUALIZADOS	REGLAMENTOS ACTUALIZADOS			40%	100%	\$ 583,33	\$ 583,33	\$ -		
Creación de un manual de funciones y procesos	EXISTENCIA DE MANUAL DE FUNCIONES	MANUAL APROBADO			60%	100%	\$ 583,33	\$ 583,33	\$ -		
Crear un plan continuo de capacitación del personal	TALLERES DE CAPACITACION DICTADOS	2 TALLERES DE TRATO A LOS CLIENTES			80%	100%	\$ 2.333,33	\$ 2.333,33	\$ -		
Incrementar la asociatividad entre los sistemas de agua	EXISTENCIA DE ACTAS DE COMPROMISO	2 ACTAS			40%	80%	\$ 262,50	\$ 262,50	\$ -		
Establecer convenios de cooperación interinstitucional	EXISTENCIA DE CONVENIOS INTERINSTITUCIONALES	2 CONVENIOS FIRMADOS			30%	70%	\$ 2.216,67	\$ 2.216,67	\$ -		
Campaña para posicionar la imagen institucional (Marketing y Publicidad)	EXISTENCIA DE PLAN DE MARKETING	PLAN APROBADO			50%	100%	\$ 1.633,33	\$ 1.633,33	\$ -		
Gestionar incremento de caudal de aprovechamiento del río minas	PERMISO DE CAUDAL GESTIONADO	AUTORIZACIÓN PARA 100%			20%	40%	\$ 525,00	\$ 525,00	\$ -		
Compra de equipo caminero	existencia de equipo caminero	equipo adquirido			100%	100%	\$ 60.000,00	\$ 60.000,00	\$ -		
Reactivación de concesión Yunguillayacu	EXISTENCIA DE CONCESION ACTUALIZADA	CONCESION AUTORIZADA			40%	80%	\$ 583,33	\$ 583,33	\$ -		
						TOTAL	\$ 69.887,50	\$ 69.887,50	\$ -		

Tabla 15 Costos de implementación de actividades administrativas
Fuente. Análisis del investigador

CONTABILIDAD										
ACCIONES	INDICADORES DE GESTION	META	PROGRAMACION				TOTAL PRESUPUESTO	FINANCIAMIENTO O JAAPB	FINANCIAMIENTO OTROS	OBSERVACIONES
			I	II	III	IV				
Estudio tecnico de produccion y costo real	EXISTENCIA DE ESTUDIO	ESTUDIO APROBADO			100%	100%	\$ 1.500,00	\$ 300,00	\$ 1.200,00	FINANCIA ADCON - BIZMAN
Actualizacion del sistema contable	SISTEMA CONTABLE ACTUALIZADO	SISTEMA INSTALADO Y FUNCIONANDO			100%	100%	\$ 1.500,00	\$ 1.500,00		
						TOTAL	\$ 3.000,00	\$ 1.800,00	\$ 1.200,00	

Tabla 16 Costos de implementación de actividades en contabilidad
Fuente. Análisis del investigador

SISTEMAS										
ACCIONES	INDICADORES DE GESTION	META	PROGRAMACION				TOTAL PRESUPUESTO	FINANCIAMIENTO O JAAPB	FINANCIAMIENTO OTROS	OBSERVACIONES
			I	II	III	IV				
Implementacion de seguridad cuarto de equipos informaticos	EXISTENCIA DE SEGURIDAD	100%			100%	100%	\$ 1.200,00	\$ 1.200,00	\$ -	
Mantenimiento de sistemas de informacion actualizados	solicitudes de cambios	100% DE CUMPLIMIENTO DE LAS AUTORIZADAS SEGÚN PLAN				100%	\$ 350,00	\$ 350,00	\$ -	
Mantenimiento de equipos actualizados	cantidad de equipos renovados	10% anual			5%	10%	\$ 2.000,00	\$ 2.000,00		
Implementar un sistema de facturacion aplicando el metodo de devengo	SISTEMA TRABAJANDO CON EL METODO DE DEVENGO	100%			50%	100%	\$ 1.500,00	\$ 1.500,00	\$ -	
Sistemas automaticos de respaldo	sistema en funcionamiento	100%				100%	\$ 1.200,00	\$ 1.200,00	\$ -	
						TOTAL	\$ 6.250,00	\$ 6.250,00	\$ -	

Tabla 17 Costos de implementación de actividades en departamento de sistemas
Fuente. Análisis del investigador

4.8.4. Presupuesto año 2014

Condiciones: El presupuesto se ha elaborado tomando en cuenta los siguientes aspectos:

El presupuesto está desarrollado en base al análisis de las tendencias de los informes financieros, que da como resultado, en el caso ingresos, que en los últimos 3 períodos hay una tendencia de incremento promedio de un 7.1%, mientras que, en el caso de los gastos, la tendencia es de un incremento del 4%. La diferencia radica principalmente en que el ingreso de socios significa un incremento proporcional en dinero de acuerdo a los m³ consumidos, sin embargo, en el caso de los costos, lo único que cambia son los costos variables, que son reducidos.

El incremento de costos se ve afectado en gran medida por los efectos inflacionarios que para este año se espera sean de alrededor del 3% según datos del BCE.

La primera parte del presupuesto, compuesta por los datos financieros de ingresos y egresos, considera los datos anteriormente indicados para proyectar el comportamiento a finales de este año, mientras que la segunda parte, que abarcan los nuevos gastos e ingresos, se basa en los costos estimados para cumplir el plan de actividades del año 2014.

Se agrega una sección de inversiones programadas para este año y se complementa el presupuesto con la estimación del flujo de caja, que se ve afectado por las inversiones y los costos de depreciaciones y amortizaciones.

JUNTA ADMINISTRADORA DE AGUA POTABLE DE BAÑOS						
INGRESOS PROMEDIO 7,10%	PRESUPUESTO 2014					
EGRESOS PROMEDIO 0,4%						
	PRESUPUESTO TRIMESTRAL ACUMULADO					
AÑO	2014	MENSUAL	I TRIMESTRE	II TRIMESTRE	III TRIMESTRE	IV TRIMESTRE
INGRESOS			1	2	3	4
Vtas.Netas Gravadas con Tarifa Cero (Suministro de Agua)						
Derechos de Instalacion	77.000	6.417	19.250	38.500	57.750	77.000
Consumo de Agua	491.000	40.917	122.750	245.500	368.250	491.000
Traspaso de Usuario	1.000	83	250	500	750	1.000
Otros (Multas)	-	-	-	-	-	-
Devoluciones en Ventas	-	-	-	-	-	-
OTROS INGRESOS						
Otros Ingresos no Especificados	-	-	-	-	-	-
Rendimientos Financieros (Intereses)	11.000	917	2.750	5.500	8.250	11.000
INGRESOS POR CO-FINANCIAMIENTO						
Campañas de concientizacion y capacitación sobre usos alternativos del suelo	214		-	-	107	107
Ejecutar proyectos de conservacion, forestacion reforestacion	991		-	-	496	496
Estudios de nueva planta - convenio Ecuador Estratégico	112.000		-	-	-	112.000
Estudio tecnico de produccion y costo real	1.200		-	-	600	600
TOTAL INGRESOS	694.405	48.333	145.000	290.000	436.203	693.203

Tabla 18 Resumen de presupuesto de ingresos para el año 2014
Fuente. Análisis del investigador

Los valores proyectados están en miles de dólares, puesto que valores menores tienen poco significado en relación al total agrupado. En el caso de los valores presupuestados para cumplir el plan, se colocan con valores menores debido a que son actividades individuales.

COSTOS Y GASTOS	AÑO 2014	MENSUAL	I TRIM	II TRIM	III TRIM	IV TRIM
Gastos en Trabajadores						
Sueldos, Horas Extras y Dietas	119.000	9.917	29.750	59.500	89.250	119.000
Remuneraciones Complementarias	42.000	3.500	10.500	21.000	31.500	42.000
Aportes IESS	23.000	1.917	5.750	11.500	17.250	23.000
Indemnizaciones	20.000	1.667	5.000	10.000	15.000	20.000
Vestuario y Prendas Protección	3.000	250	750	1.500	2.250	3.000
Alimentos y Bebidas	2.000	167	500	1.000	1.500	2.000
Contratacion de Servicios						
Servicios de Profesionales	37.000	3.083	9.250	18.500	27.750	37.000
Servic.en Sistemas Informat.	4.000	333	1.000	2.000	3.000	4.000
Servicios de Auditoria	5.000	417	1.250	2.500	3.750	5.000
Capacitaciones	-	-	-	-	-	-
Energia Elect. Y Telecom.	6.000	500	1.500	3.000	4.500	6.000
Eventos de la Junta de Agua	-	-	-	-	-	-
Transporte	2.000	167	500	1.000	1.500	2.000
Difusion, Publicidad, Impres. Public.	4.000	333	1.000	2.000	3.000	4.000
Vigilancia, Aseo y Otros	1.000	83	250	500	750	1.000
Produccion y Distrib.de Agua						
Químicos	34.000	2.833	8.500	17.000	25.500	34.000
Accesorios Instalación	3.000	250	750	1.500	2.250	3.000
Accesorios Reparación	1.000	83	250	500	750	1.000
Alquiler Maquinaria y Equip.	3.000	250	750	1.500	2.250	3.000
Reparación de Tuberías	4.000	333	1.000	2.000	3.000	4.000
Materiales para Laboratorio	4.000	333	1.000	2.000	3.000	4.000
Combustibles y Lubricantes	5.000	417	1.250	2.500	3.750	5.000
Repuestos y Herramientas	1.000	83	250	500	750	1.000
Mantenimientos						
Edificios	6.000	500	1.500	3.000	4.500	6.000
Maquinaria, Equip. Vehiculos	5.000	417	1.250	2.500	3.750	5.000
Gastos en Oficinas y Otros						
Materiales de Oficina	4.000	333	1.000	2.000	3.000	4.000
Insumos para Aseo	1.000	83	250	500	750	1.000
Bienes No Depreciables	8.000	667	2.000	4.000	6.000	8.000
Gastos Financieros Impuestos y Otros	26.000	2.167	6.500	13.000	19.500	26.000
Seguros	3.000	250	750	1.500	2.250	3.000
Arriendo de Mobiliarios e Inmuebles	1.000	83	250	500	750	1.000
Depreciacion Activos Fijos	64.000	5.333	16.000	32.000	48.000	64.000

Tabla 19 Resumen de presupuesto de costos y gastos para el año 2014
Fuente. Análisis del investigador

NUEVOS GASTOS PLANIFICADOS	AÑO 2014	MENSUAL	I TRIM	II TRIM	III TRIM	IV TRIM
Campañas de concientización y capacitación sobre usos alternativos del suelo	644	-	-	-	322	322
Ejecutar proyectos de conservación, forestación reforestación	1.983	-	-	-	992	992
Ensayos (Insitu) de adaptación de especies nativas	408	-	-	-	204	204
Depreciaciones y amortizaciones inversiones nuevas	14.291				7.146	7.146
Asesoría Legal permanente	1.000	-	-	-	500	500
Actualización de los reglamentos y políticas internas	500	-	-	-	250	250
Creación de un manual de funciones y procesos	500	-	-	-	250	250
Crear un plan continuo de capacitación del personal	2.000	-	-	-	1.000	1.000
Incrementar la asociatividad entre los sistemas de	225	-	-	-	113	113
Gestionar incremento de caudal de aprovechamiento del río Minas	450	-	-	-	225	225
Reactivación de concesión Yunguillayacu	500	-	-	-	250	250
Estudio técnico de producción y costo real	750	-	-	-	375	375
Mantenimiento de sistemas de información	300	-	-	-	150	150
Diagnóstico de las redes de agua potable	650	-	-	-	325	325
Establecer convenios de cooperación interinstitucional	900	-	-	-	450	450
Campaña para posicionar la imagen institucional (Marketing y Publicidad)	400	-	-	-	200	200
Reparación de desarenador	875	-	-	-	438	438
Establecer convenios de cooperación interinstitucional	2.000	-	-	-	1.000	1.000
TOTAL COSTOS Y GASTOS	465.201	36.750	110.250	220.500	342.851	453.101
SUPERAVIT DEL EJERCICIO ECONOMICO	229.204	19.100	57.301	114.602	171.903	229.204
Depreciaciones y amortizaciones	106.873	8.906	26.718	53.437	80.155	106.873
Inversiones	- 234.658	- 19.555	- 58.665	- 117.329	- 175.994	- 234.658
Flujo neto de efectivo	101.419	8.452	25.355	50.710	76.064	101.419

Tabla 20 Resumen de presupuesto de nuevos costos y gastos para el año 2014
Fuente. Análisis del investigador

INVERSIONES JAAPB - 2014			
Detalle	Valor	vida útil	Depreciacion / Amortizacion anual
Sistema de retrolavado	5.833	10	583
Construccion del Tanque rompe-presión	6.000	10	600
Implementar y renovar equipos e instrumentos acordes a tecnología actual	2.090	5	418
Estudios de nueva planta - convenio Ecuador Estratégico (Fase 1)	112.000	10	11.200
Estudio de delimitación de área de cobertura	6.000	5	1.200
Estudio - Diagnóstico de las redes de agua potable	29.000	5	5.800
equipos para campaña para posicionar la imagen institucional (Marketing y Publicidad)	2.000	10	200
Implementacion de seguridad cuarto de equipos informaticos	1.200	5	240
Implementación de un sistema de facturacion aplicando el metodo de devengo	1.500	3	500
Sistemas automáticos de respaldo	1.200	3	400
Software de contabilidad	1.500	5	300
Adquisicion e Instalacion de Equipos (Macromedidores)	4.160	10	416
Adquisicion de equipo caminero	60.000	10	6.000
Actualizacion de equipos informaticos	2.000	3	667
Reactivacion de los equipos de lecturacion	175	3	58
TOTAL INVERSIONES	234.658	Total	28.582
	Depreciacion período		14.291

Tabla 21 Tabla de depreciaciones presupuestadas para el año 2014
Fuente. Análisis del investigador

Capítulo 5: Comparación y análisis de resultados

5.1. Análisis de resultados

El proceso de aplicar el modelo a todo el conjunto de datos tomó un tiempo de 23 minutos, dando como resultado un valor pronosticado de consumo de 1'555.414 m³ para el año 2014.

Los valores consumidos durante el año 2014 según los registros de la base de datos fueron de 1'598.473m³, que implica una diferencia de 43.000m³, esto es un 3% debajo del valor real.

Lograr resultados con una precisión del 97% es muy alentador acerca de las capacidades de la modelización con la herramienta R. Al analizar si es posible realizar mejoras al modelo de pronósticos, se encontró un elemento que podría ser mejorado, y es que el modelo necesitaba de al menos 3 datos con consumos mayores a cero para poder realizar el proceso de suavizado preliminar, y cuando no contaba con suficientes datos, simplemente descartaba a ese usuario.

Las causas posibles de usuarios sin datos es que tengan solamente comprado el derecho, pero no tengan una instalación física, por lo que en realidad no consumen agua y no deben ser considerados en el modelo.

También existe la posibilidad de que hayan menos de 3 datos de consumos, en el caso de usuarios que recién comienzan a utilizar el servicio de agua, por lo que se realizó una consideración adicional al modelo, en donde, a aquellos datos que tienen una o dos observaciones, se considere como consumo probable el promedio de sus consumos, o el único valor que tiene de ser el caso.

La aplicación de este modelo mejorado arrojó un pronóstico de consumo de 1'625.730 m³, lo que significa que la diferencia fue de 27.257 m³ por encima del valor real, esto es un 2% superior.

Este resultado del modelo mejorado eleva la precisión a un 98%, mayor al que normalmente se acostumbra en estudios científicos, que es del 95%.

5.2. Conclusiones y recomendaciones

La aplicación de modelos de pronóstico mediante el uso de herramientas informáticas tiene la gran ventaja de que permiten mejorar la gestión empresarial, puesto que los

factores sobre los cuales se basan los administradores normalmente dependen únicamente de proyecciones sobre datos globales, sumada a la experiencia y algunas suposiciones subjetivas, lo que eleva el riesgo de obtener resultados inadecuados.

La dificultad de realizar análisis detallados es una gran limitante para aplicar modelos a grandes cantidades de datos, pero se ha demostrado en el presente proyecto, que, con la selección adecuada de herramientas, este proceso puede realizarse con cierta facilidad.

El hecho de aplicar el análisis científico de datos, permite que los procesos de análisis sean documentados y, sobre todo, que sea posible de replicarse al tener organizados los scripts en un solo archivo. Esto permite además aplicar los mismos procesos a otro conjunto de datos únicamente cambiando los archivos con las observaciones, por lo que el gran esfuerzo que significa elaborar el estudio una vez, se ve totalmente simplificado para las siguientes ocasiones.

El analista o el equipo administrativo invertirá mejor su tiempo interpretando los datos y no tratando de lidiar cada vez con demorados procesos de carga, limpieza y generación de diversos gráficos adecuados a cada tema.

En este momento existen diversas técnicas de aprendizaje de máquina (machine learning), como el análisis de clúster, clasificación por vecinos cercanos, árboles de decisión, regresión múltiple, redes neuronales y máquina de vectores de soporte, entre otros, que pueden explorarse e implementarse en modelos desarrollados con R.

Para que las empresas puedan progresar en esta era del conocimiento, será necesario que éstas definan el nivel de precisión requerido y la adecuación de cada técnica de pronóstico de acuerdo a las características de los datos y a la evolución de la industria, a fin de elaborar un banco de aplicaciones óptimas para cada caso.

Lo anterior permitirá en última instancia, realizar pronósticos con base científica, que complementen la capacidad holística de toma de decisiones de los administradores.

Glosario

Analítica de datos. Proceso de tomar datos y convertirlos en conocimiento que permita tomar decisiones más inteligentes.

Aprendizaje automático. Tipo de inteligencia artificial que permite a las computadoras aprender en base a un entrenamiento con datos de ejemplo.

Big data. Se refiere a una cantidad de datos tan grande que supera las capacidades de las técnicas tradicionales de captura, almacenamiento y tratamiento.

BPM. Del inglés Business Performance Management, es un conjunto de procesos de gestión del rendimiento que permite a una empresa lograr sus objetivos.

Ciencia de datos. Es un campo interdisciplinario que involucra métodos científicos, procesos y sistemas para extraer conocimiento o un mejor entendimiento de los datos.

Computación en la nube. Del inglés *Cloud computing* y conocido también como servicios en la nube, es un paradigma que permite ofrecer servicios de computación a través de Internet.

CRAN. Es una red de servidores web alrededor del mundo que almacenan versiones idénticas y actualizadas de código y documentación para R.

CSV. Del inglés *comma-separated values*, es un tipo de documento en formato de texto, abierto y sencillo para representar datos en forma de tabla, en las que las columnas se separan por comas.

DataFrame. Conjunto de datos en R, similares a una tabla, que se cargan en memoria para ser analizados.

Dataset. Conjunto de datos en donde cada elemento contiene las mismas propiedades que los demás, el similar a una tabla.

Estocástico. Proceso sometido al azar por lo que no puede ser analizado de forma determinística.

Floculación. Proceso mediante el cual se añaden sustancias químicas al agua para que se aglutinen sustancias allí presentes, de modo que faciliten su decantación y filtrado.

Html. Sigla de *HiperText Markup Language* (Lenguaje de Marcación de Hipertexto). Es un lenguaje que se utiliza para establecer la estructura y contenido de una página web.

inteligencia Artificial. Ciencia e ingenio de hacer máquinas inteligentes, especialmente programas de cómputo inteligentes.

Inteligencia de Negocios (*Business Intelligence*). Capacidad para transformar los datos en conocimiento, para optimizar el proceso de toma de decisiones en los negocios

Json. Del inglés *JavaScript Object Notation*, es un formato ligero de intercambio de datos basado en texto simple.

KPI. Del inglés *Key performance Indicator*, y traducido como Indicadores Clave de Desempeño, consisten en métricas que ayudan a medir el progreso en relación de metas y objetivos planteados dentro de la empresa.

Matriz de agua potable. Tuberías que conforman el núcleo de una red de distribución de agua potable.

Metadatos. Información de características de los datos.

Modelo determinístico. Modelo que puede determinar con certeza el valor de una variable, mediante la aplicación de relaciones matemáticas.

NOAA. Siglas de *National Oceanic and Atmospheric Administration*, una agencia científica de los Estados Unidos, que se encarga de monitorizar las condiciones de los océanos y la atmósfera.

RMarkdown. Implementación para R del lenguaje de marcado ligero Markdown usando texto plano.

RStudio. Entorno de desarrollo integrado para R.

Script. Es un programa normalmente simple, que se almacena en un archivo de texto.

Sedimentación. Proceso por el cual los materiales transportados en el agua son depositados en el fondo de un contenedor.

XML. Siglas de *eXtensible Markup Language*, traducido como "Lenguaje de Marcado Extensible", es un meta-lenguaje que permite definir lenguajes para almacenar datos en forma legible para las computadoras.

Referencia bibliográfica

- Baca Urbina, Gabriel. 2013. *Evaluación de proyectos*. México, D.F.: McGraw-Hill.
- Bonini, Charles P, Warren H Hausman, Harold Bierman, y Cecilia Avila de Barón. 2000. *Análisis cuantitativo para los negocios*. Santafé de Bogotá: McGraw-Hill.
- Brito Alemán, Emma. 2005. *Baños, ensueños del Alma*. 1a ed. Cuenca, Ecuador.
- Burbano Ruíz, Jorge E. 2011. *Presupuestos: un enfoque de direccionamiento estratégico, gestión, y control de recursos*. Bogotá: McGraw-Hill.
- Chandrasekaran. 2015. "TCS Annual Report 2014-2015".
- Chapman, Chris. 2015. *R for marketing research and analytics*. New York, NY: Springer Science+Business Media, LLC.
- Chase, Charles. 2009. *Demand-driven forecasting: a structured approach to forecasting*. Wiley & SAS business series. Hoboken, N.J: John Wiley & Sons.
- Dietrich, David, Barry Heller, Beibei Yang, y EMC Education Services, eds. 2015. *Data science & big data analytics: discovering, analyzing, visualizing and presenting data*. Indianapolis, IN: Wiley.
- Douglas A Lind, William G Marchal, Samuel A Wathen, y Vizcaino Díaz, Ofelia. 2012. *Estadística aplicada a los negocios y la economía*. México D. F.: McGraw-Hill. 2012.
- Dropkin, Murray, James Halpin, y Bill La Touche. 2007. *The budget-building book for nonprofits: a step-by-step guide for managers and boards*. 2nd ed. San Francisco: Jossey-Bass.
- Erl, Thomas, Wajid Khattak, y Paul Buhler. 2016. *Big data fundamentals: concepts, drivers & techniques*. The Prentice Hall service technology series from Thomas Erl. Boston : [Vancouver, BC]: Prentice Hall ; Service Tech Press.
- Ford, Martin. 2015. *Rise of the robots: technology and the threat of a jobless future*. New York: Basic Books, a member of the Perseus Books Group.
- Heizer, Jay H, Barry Render, Jesús Elmer Murrieta Murrieta, y Guillermo Haaz Díaz. 2009. *Principios de administración de operaciones*. México: Pearson Educación.
- Horngren, Charles T, Srikant M Datar, Madhav V Rajan, y Jaime Gómez Mont Araiza. 2012. *Contabilidad de costos: un enfoque gerencial*. México, D.F.: Pearson Educación.
- Hoshmand, A. Reza. 2010. *Business forecasting: a practical approach*. 2nd ed. New York: Routledge.
- Keller, Gerald. 2014. *Statistics for management and economics*. 10e ed. Stamford, CT, USA: Cengage Learning.
- Kelly, John E., y Steve Hamm. 2013. *Smart machines: IBM's Watson and the era of cognitive computing*. New York: Columbia Business School Publishing.
- Kudyba, Stephan. 2014. *Big Data, Mining, and Analytics: Components of Strategic Decision Making*. Boca Raton: CRC Press, Taylor and Francis. <http://www.books24x7.com/marc.asp?bookid=61774>.
- Manyika, James, Chui, Michael, Brown, Brad, Bughin, Jacques, Dobbs, Richard, Roxburgh, Charles, y Hung Byers, Angela. 2011. "Big data: The next frontier for innovation, competition, and productivity". Washington, DC: McKinsey & Company.
- Matloff, Norman S. 2011. *The art of R programming: tour of statistical software design*. San Francisco: No Starch Press.

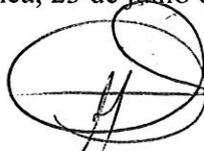
- Mun, Johnathan. 2010. *Modeling risk: applying Monte Carlo risk simulation, strategic real options, stochastic forecasting, and portfolio optimization*. 2nd ed. Wiley finance. Hoboken, N.J: Wiley.
- Mutanov, Galimkair. 2015. *Mathematical methods and models in economic planning, management and budgeting*. 2nd edition. Berlin, Heidelberg, NY: Springer Berlin Heidelberg.
- Nugus, Sue. 2009. *Financial planning using Excel: forecasting, planning and budgeting techniques*. 2nd ed. Oxford: CIMA.
- Ohlhorst, Frank. 2013. *Big data analytics: turning big data into big money*. Wiley & SAS business series. Hoboken, N.J: Wiley.
- Pries, Kim H, y Robert Dunnigan. 2015. *Big Data Analytics: A Practical Guide for Managers*.
<http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=957542>.
- Quito B., Manuel Hernan, y Fabiola Marisol Cárdenas V. 2015. *Análisis de factibilidad socio-económica para la construcción de una planta de agua potable para la parroquia Baños*. Tesis Posgrado. Cuenca, Ecuador: Universidad Politécnica Salesiana. <http://dspace.ups.edu.ec/handle/123456789/8432>.
- . 2014. "Plan estratégico de la junta administradora de agua potable de la parroquia Baños." JAAPB.
- Render, Barry, Ralph M Stair, Michael E Hanna, y Marcía Aída González Osuna. 2012. *Métodos cuantitativos para los negocios*. México, D.F.: Pearson Educación.
- Shiavi, Richard. 2007. *Introduction to applied statistical signal analysis: guide to biomedical and electrical engineering applications*. 3rd ed. Burlington, MA: Elsevier/Academic Press.
- Shim, Jae K., Joel G. Siegel, y Allison I. Shim. 2012. *Budgeting basics and beyond*. 4th ed. Wiley corporate F&A series. Hoboken, N.J: Wiley.
- Turban, Efraim. 2011. *Decision support and business intelligence systems*. 9th ed. Boston: Prentice Hall.
- Vaikunth , Pai. 2016. "BIG DATA NEW CHALLENGES, TOOLS AND TECHNIQUES", International Journal of Engineering Research and Modern Education, 1 (1).
- Vu, Tam Bang. 2015. *Seeing the Future: How to Build Basic Forecasting Models*. <http://www.businessexpertpress.com/>.
- Warden, Pete. 2011. *Big Data Glossary: [a Guide to the New Generation of Dta Tools]*. Sebastopol, CA: O'Reilly Media.
- Welsch, Glenn A, y Julián Vázquez Prada Gutiérrez. 2005. *Presupuestos: planificación y control*. México: Pearson Educación.

Doctora Jenny Ríos Coello, Secretaria de la Facultad de Ciencias de la Administración de la Universidad del Azuay,

CERTIFICA:

Que, el H. Consejo de Facultad en sesión realizada el 18 de junio de 2015, conoció la petición del estudiante **MANUEL HERNAN QUITO BACULIMA** con código 28501, que denuncia su trabajo de titulación denominado: **“PRESUPUESTO FINANCIERO ANUAL DE LA JUNTA ADMINISTRADORA DE AGUA POTABLE DE BAÑOS EMPLEANDO PROYECCIONES CON LA APLICACIÓN INFORMÁTICA R Y TÉCNICAS DE BIG DATA.”**, presentado previa a la obtención del título de Ingeniero en Comercial.- El Consejo de Facultad acoge el informe de la Junta Académica y aprueba la denuncia del trabajo de titulación. Designa como **Director al Ing. Marcos Orellana Cordero** y como miembros del Tribunal Examinador al Ing. Iván Orellana Osorio y al Eco. Econ. René Niveló Cabrera .- De conformidad con el Instructivo para la Creación y Funcionamiento de la Unidad de Titulación Especial y las disposiciones superiores Of. No 660-V-15-UDA del Vicerrectorado, el peticionario, para desarrollar y presentar su trabajo de titulación, tienen un plazo de **DIECIOCHO MESES**, contados a partir de la fecha de aprobación del Consejo de Facultad, estos es **hasta el 18 de diciembre de 2016** .

Cuenca, 23 de junio de 2015



Dra. Jenny Ríos Coello
Secretaria de la Facultad

UNIV 0854
FACULTAD DE
CIENCIAS DE LA ADMINISTRACIÓN
SECRETARÍA

CONCONVOCATORIA

Por disposición de la Junta Académica de Administración de Empresas, se convoca a los Miembros del Tribunal Examinador, a la sustentación del Protocolo del Trabajo de Titulación "**Elaboración del presupuesto financiero anual de la Junta Administradora de Agua Potable de Baños aplicando proyecciones con la aplicación informática R y técnicas de Big Data**", presentado por el estudiante Manuel Hernán Quito Baculima con código 28501, previa a la obtención del grado de Ingeniero Comercial, para el día JUEVES 07 de MAYO DE 2015 A LAS 18H30

Cuenca, 28 de abril de 2015



Dra. Jenny Ríos Coello
Secretaria de la Facultad

Ing. Marcos Orellana Cordero

Ing. Iván Orellana Osorio

Eco. René Niveló Cabrera



Comunicado



ACTA

SUSTENTACIÓN DE PROTOCOLO/DENUNCIA DEL TRABAJO DE TITULACIÓN

- 1.1 Nombre del estudiante: Manuel Hernán Quito Baculima
- 1.2 Código 28501
- 1.3 Director sugerido: Ing. Orellana Cordero Marcos
- 1.4 Codirector (opcional): _____
- 1.5 Tribunal: Ing. Iván Orellana Osorio / Eco. René Niveló Cabrera
- 1.6 Título propuesto: "Elaboración del presupuesto financiero anual de la Junta Administradora de Agua Potable de Baños aplicando proyecciones con la aplicación informática R y técnicas de Big Data"
- 1.7 Resolución:

1.7.1 Aceptado sin modificaciones _____

1.7.2 Aceptado con las siguientes modificaciones:

Eliminar "Elaboración" del título.
Replantear la pregunta de Investigación.

- Responsable de dar seguimiento a las modificaciones: Ing. Marcos Orellana Cordero

1.7.3 No aceptado

- Justificación:

.....
Ing. Marcos Orellana Cordero

.....
Tribunal
Ing. Iván Orellana Osorio

.....
Eco. René Niveló Cabrera

.....
Sr. Manuel Hernán Quito Baculima

.....
Dra. Jenny Ríos Coello
Secretario de Facultad

Fecha de sustentación: Jueves 7 de mayo de 2015 al as 18h30



RÚBRICA PARA LA EVALUACIÓN DEL PROTOCOLO DE TRABAJO DE TITULACIÓN

1.1 Nombre del estudiante: Manuel Hernán Quito Baculima

1.1.1 Código 28501

1.2 Director sugerido: Ing. Orellana Cordero Marcos

1.3 Codirector (opcional):

1.4 Título propuesto: : "Elaboración del presupuesto financiero anual de la Junta Administradora de Agua Potable de Baños aplicando proyecciones con la aplicación informática R y técnicas de Big Data"

1.5 Revisores (tribunal): Ing. Iván Orellana Osorio / Eco. René Niveló Cabrera

1.6 Recomendaciones generales de la revisión:

	Cumple totalmente	Cumple parcialmente	No cumple	Observaciones (*)
Línea de investigación	✓			
1. ¿El contenido se enmarca en la línea de investigación seleccionada?	✓			
Título Propuesto				
2. ¿Es informativo?	✓			
3. ¿Es conciso?		✓		
Estado del arte				
4. ¿Identifica claramente el contexto histórico, científico, global y regional del tema del trabajo?	✓			
5. ¿Describe la teoría en la que se enmarca el trabajo	✓			
6. ¿Describe los trabajos relacionados más relevantes?	✓			
7. ¿Utiliza citas bibliográficas?	✓			
Problemática y/o pregunta de investigación				
8. ¿Presenta una descripción precisa y clara?		✓		
9. ¿Tiene relevancia profesional y social?	✓			
Hipótesis (opcional)				
10. ¿Se expresa de forma clara?	✓			
11. ¿Es factible de verificación?	✓			
Objetivo general				
12. ¿Concuerda con el problema formulado?	✓			
13. ¿Se encuentra redactado en tiempo verbal infinitivo?	✓			



Objetivos específicos				
14.¿Concuerdan con el objetivo general?	/			
15.¿Son comprobables cualitativa o cuantitativamente?	/			
Metodología				
16.¿Se encuentran disponibles los datos y materiales mencionados?	/			
17.¿Las actividades se presentan siguiendo una secuencia lógica?	/			
18.¿Las actividades permitirán la consecución de los objetivos específicos planteados?	/			
19.¿Los datos, materiales y actividades mencionadas son adecuados para resolver el problema formulado?	/			
Resultados esperados				
20.¿Son relevantes para resolver o contribuir con el problema formulado?	/			
21.¿Concuerdan con los objetivos específicos?	/			
22.¿Se detalla la forma de presentación de los resultados?	/			
23.¿Los resultados esperados son consecuencia, en todos los casos, de las actividades mencionadas?	/			
Supuestos y riesgos				
24.¿Se mencionan los supuestos y riesgos más relevantes?	/			
25.¿Es conveniente llevar a cabo el trabajo dado los supuestos y riesgos mencionados?	/			
Presupuesto				
26.¿El presupuesto es razonable?	/			
27.¿Se consideran los rubros más relevantes?	/			
Cronograma				
28.¿Los plazos para las actividades son realistas?	/			
Referencias				
29.¿Se siguen las recomendaciones de normas internacionales para citar?	/			
Expresión escrita				
30.¿La redacción es clara y fácilmente comprensible?	/			
31.¿El texto se encuentra libre de faltas ortográficas?	/			



(*) Breve justificación, explicación o recomendación.

- Opcional cuando cumple totalmente,
- Obligatorio cuando cumple parcialmente y NO cumple.

.....

.....

.....

Ing. Marcos Orellana Cordero

Ing. Iván Orellana Osorio

Eco. René Niveló Cabrera

Cuenca, 26 de Mayo de 2015

Señor Ingeniero
Xavier Ortega Vázquez
DECANO DE LA FACULTAD DE CIENCIAS DE LA ADMINISTRACIÓN
Presente.-

De mis consideraciones:

Luego de revisar las recomendaciones realizadas por parte del tribunal designado en el diseño del trabajo de graduación: "Presupuesto financiero anual de la Junta Administradora de Agua Potable de Baños empleando proyecciones con la aplicación informática R y técnicas de Big Data"; presentado por el señor Hernán Quito, estudiante de la Escuela de Administración de Empresas, considero que el documento se alinea a los cambios sugeridos y cumple con las normas legales y reglamentarias de la Universidad y de la Facultad de Ciencias de la Administración, por lo que recomiendo su aprobación por parte del Consejo de Facultad.

Atentamente,



Ing. Marcos Orellana Cordero
Director



1. DATOS GENERALES

1.1. Nombre del estudiante: QUITO BACULIMA MANUEL HERNAN

1.1.1. Código 28501

1.1.2. Contacto: 2892198 – 09 95959560 – mbaculima@unazuay.edu.ec

1.2. Director Sugerido: Orellana, Marcos Ing.

1.2.1. Contacto: marcos.orellana@unazuay.edu.ec

1.3. Co-director sugerido:

1.4. Asesor metodológico:

1.5. Tribunal designado:

1.6. Aprobación:

1.7. Línea de investigación de la Carrera: GESTIÓN FINANCIERA

1.7.1. Código Unesco: 5311.02

1.7.2. Tipo de trabajo:

1.7.2.1. a) trabajos experimentales

1.7.2.2. b) Investigación formativa.

1.8. Área de estudio: ESTADÍSTICA, ADMINISTRACIÓN

1.9. Título propuesto: Presupuesto financiero anual de la Junta Administradora de Agua Potable de Baños empleando proyecciones con la aplicación informática R y técnicas de Big Data.

1.10. Subtítulo:

1.11. Estado del proyecto: Trabajo de investigación nuevo.

2. CONTENIDO

2.1 Motivación de la investigación: La investigación está orientada a la descripción de la problemática

2.2 Problemática: La falta de una planificación basada en técnicas estadísticas que aprovechen las nuevas propuestas de análisis científico de datos, tendientes a explotar la gran cantidad de información almacenada en las bases de datos en las empresas, impide un análisis más acertado y técnico en la administración y por ende impide mejorar la efectividad empresarial.

2.3 Pregunta de investigación: ¿Se puede elaborar el presupuesto financiero anual de la Junta Administradora de Agua Potable de Baños utilizando modelos predictivos con el software R y técnicas de análisis de Big Data?.

2.4 Resumen: Se analizarán técnicas administrativas tradicionales de elaboración de presupuestos y toma de decisiones, para luego aplicar los nuevos procesos de análisis basados en Big Data y técnicas estocásticas de proyección mediante el uso del software R. Se utilizará como insumo

principal los datos de consumo de agua potable y de las recaudaciones de la Junta Administradora de Agua Potable de la Parroquia Baños.

2.5 Estado del Arte y marco teórico:

Jorge Burbano (2011) señala que los pronósticos de manejo del efectivo son importantes para evitar situaciones de insolvencia y permitir la valuación científica de la empresa; actualmente, con el avance de la informática, se han desarrollado nuevas metodologías de pronóstico y administración empresarial que hacen uso de herramientas de Inteligencia de Negocios para mejorar su desempeño, como es el caso de BPM, que son las siglas en inglés de la *Administración del desempeño empresarial* (Business Performance Management), la que está compuesta por 3 elementos fundamentales que son:

- Un conjunto de procesos administrativos y analíticos, apoyados por la tecnología, que direccionan actividades tanto financieras como operacionales
- Herramientas de negocio para definir objetivos estratégicos, medirlos y administrar su desempeño en relación a las metas.
- Un conjunto medular de procesos, incluidos planeación financiera y operativa, consolidación y reporte, modelado, análisis y monitoreo de los principales indicadores de desempeño (KPIs), enlazados a la estrategia empresarial (Turban, 2014).

Considerando lo anterior, tenemos que para administrar una empresa y lograr cumplir los objetivos de forma eficiente, es necesario realizar estudios adecuados que involucre los análisis de demanda que permitan determinar la cantidad de bienes o servicios que la población objetivo va a requerir (Baca Urbina, 2013), en este caso, la cantidad de agua potable, la que a su vez incide directamente en los ingresos.

Para proyectar la cantidad demandada, es necesario realizar pronósticos mediante técnicas tradicionales de análisis estadístico y cuantitativo, tales como series de tiempo, promedios móviles, suavizado exponencial y técnicas de regresión (Bonini, Hausman, Hausman & Avila, 2000). En el caso del análisis de series de tiempo, se pueden aplicar análisis de tendencias seculares, variaciones cíclicas y variaciones estacionales en el caso de ser necesarias (Lind, Marchal, Wathen & Vizcaino, 2012).

La influencia de la informática que está ya varios años involucrada directamente en los negocios, ha permitido que se genere gran cantidad de datos, que sin embargo debido a la falta de herramientas adecuadas y a la falta de personal capacitado, no se ha logrado explotar el potencial que tienen tales datos para optimizar la toma de decisiones empresariales. Hoy en día esta problemática ha derivado en la necesidad de herramientas que permitan trabajar con grandes



volúmenes de datos, aplicando técnicas de análisis estadístico basadas en la aplicación de modelos estocásticos y el análisis científico de datos, cuyo aporte va desde la adquisición y limpieza de datos, hasta la generación de conocimiento y su aplicación como soporte en la toma de decisiones.

Se debe tener cuidado con el uso apropiado de los datos, como nos lo alerta Jeffrey Leek, profesor de la especialización Data Scientist de la Universidad Johns Hopkin, al expresar que la combinación de datos y el deseo de una respuesta, no asegura que pueda obtenerse una respuesta razonable de un grupo de datos dado, sin importar cuan grandes sean tales cantidades (What about Big Data. Curso Data Scientist).

Frank Ohlhost (2013) en su libro *Big Data Analytics; turning big data into big money*, expresa que en la actualidad los negocios están recolectando cada vez más información, creando así la denominada "Big Data". Información que va desde registros de transacciones comerciales hasta videos e imágenes, pasando por llamadas telefónicas con referencias de posiciones geográficas, datos de sensores climáticos, consumos de recursos, búsquedas en línea, contenido creado en redes sociales, etc. Esta gran cantidad de datos generó el problema inicial de almacenarlos y procesarlos rápidamente; pero una vez solucionado esto, surge el nuevo problema de explorar múltiples fuentes de datos para generar inteligencia (*Business Intelligence*) que ayude en la toma de decisiones (*Decision Support Systems*) para solucionar problemas y producir un incremento de las utilidades, generando valor para la empresa; la solución se está logrando mediante la aplicación de un proceso denominado *Big Data Analytics*

La importancia de Big Data se puede verificar en las tecnologías requeridas para su análisis y su aporte a la administración; como son: *Business Intelligence* para generar información que mejore la toma de decisiones de los administradores, *Data Mining* para encontrar patrones de comportamiento que permitan crear modelos de datos con fines predictivos, *Aplicaciones estadísticas* para analizar la información tanto de forma descriptiva como predictiva para identificar tanto riesgos como oportunidades de negocios, finalizando con el *Modelamiento de datos* que mediante la aplicación de múltiples escenarios permita analizar y visualizar el efecto que tendrían en los resultados empresariales (Ohlhost, 2013).

R es un software gratuito que combina el poder de análisis de un software de tratamiento estadístico con el poder de un lenguaje de programación, al que continuamente se le agregan paquetes para las más variadas tareas. R permite desde obtener los datos de diversas fuentes como bases de datos, archivos de texto, internet, imágenes y más, hasta realizar análisis de minería de datos y aplicaciones de aprendizaje de máquinas (Machine Learnig), lo que está enlazado en la actualidad a los conceptos de Big Data y hace posible el tratamiento de los grandes volúmenes de datos con el fin de generar conocimiento (Lantz, 2014).

2.6 Hipótesis:

2.7. **Objetivo general:** Elaborar el presupuesto financiero anual de la Junta Administradora de Agua Potable utilizando modelos predictivos con el software R y técnicas de análisis de Big Data.

2.8 Objetivos específicos:

2.8.1 Analizar el estado y problemática de la Junta Administradora de agua potable de Baños debido a la falta de información prospectiva.

2.8.2 Analizar las distintas técnicas administrativas y estadísticas para tratamiento y proyección de datos.

2.8.3. Analizar las nuevas herramientas Big Data para el tratamiento de grandes cantidades de datos y la aplicación del análisis científico de datos en los negocios.

2.8.4 Aplicar los modelos de análisis y proyección de la demanda en la elaboración de presupuestos, utilizando el software R para el tratamiento de Big Data.

2.8.5 Comparar y analizar la precisión del presupuesto tradicional y el propuesto con los resultados reales del año anterior.

2.9 Metodología

El estudio se aplicará a la Junta Administradora de Agua Potable de la Parroquia Baños, mediante la recopilación de información sobre los requerimientos de datos proyectados que se necesite para mejorar la elaboración de presupuestos y en consecuencia la toma de decisiones. Esto se realizará mediante entrevistas con los administradores y en base a esa información se analizarán las técnicas clásicas de proyección aplicadas en el campo empresarial en general y a las empresas de agua potable en particular.

Se recopilarán datos históricos de consumos, pagos, cantidad de usuarios y otros que sean de interés en la aplicación de los modelos más recomendables para el caso.

El siguiente paso será la aplicación del análisis científico de datos, para lo que se procederán a aplicar las técnicas de recopilación, limpieza, análisis descriptivo y modelización de datos, para poder aplicar los modelos estocásticos tendientes a la proyección de la demanda de agua potable.

Para esto se utilizará el software de tratamiento estadístico R, que permite realizar programas orientados a aplicar las técnicas antes mencionadas de análisis científico de datos de manera automatizada. En el caso de la Junta, se tratarán cerca de 1 millón de registros.



Utilizando los datos contables de los años pasados y los datos proyectados obtenidos en los pasos anteriores, se realizará el presupuesto financiero anual del año 2014 para compararlo con el presupuesto elaborado de forma tradicional y analizar el nivel de precisión de los datos más relevantes comparándolos con los resultados reales.

2.10 Alcances y resultados esperados.

El alcance de esta propuesta abarca la realización del presupuesto financiero anual mediante los datos de proyección de consumos de agua potable por parte de los usuarios, los ingresos y costos asociados a la producción y tratamiento de agua potable; para con ello realizar la comparativa y análisis de resultados de precisión del nuevo método de elaboración de presupuestos.

2.11 Supuestos y Riesgos

Un riesgo en este estudio, es el de que no se tengan los suficientes datos para evaluar los modelos aplicados, en cuyo caso se realizarán simulaciones, usando ya sean datos aleatorios o datos de otras empresas si es que se llegan a realizar convenios para tal efecto.

2.12 Presupuesto:

Rubro	Costo	Justificación
Materiales de papelería e impresión	200,00	Para levantamiento de datos, elaboración de informes
Libros y cursos	1.200,00	Para obtener documentación y conocimiento necesarios para aplicar el análisis.
Uso de computador y recursos informáticos	200,00	Para recolección de datos, análisis y desarrollo de informes
Movilización	100,00	Para visitar y tomar datos de la empresa
Estudio y tratamiento de datos	4.800,00	240 horas de trabajo del investigador
Otros varios	200,00	Imprevistos
TOTAL	6.700,00	

2.13 Financiamiento.

Aporte de recursos materiales del estudio por parte de la Junta (\$700)

Aporte del investigador \$ 6.000

2.14 Esquema tentativo

1. Estudio de la Junta Administradora de agua potable de Baños
 - a. Antecedentes
 - b. Situación actual
 - c. Análisis de necesidades de información basada en proyecciones
2. Elaboración de presupuestos y técnicas estadísticas para tratamiento y proyección de datos
 - a. Elaboración de Presupuestos
 - b. Modelos determinísticos aplicables a la administración
 - c. Técnicas de proyección de datos basadas en series temporales
3. Nuevas tendencias de tratamiento de datos para uso en la Administración Empresarial
 - a. Administración del desempeño empresarial
 - b. Big Data
 - c. Análisis científico de datos
 - d. Proyección y tratamiento de datos mediante técnicas estocásticas
 - e. Herramienta R para el tratamiento automatizados de grandes cantidades de datos
4. Aplicación de modelos de proyección de la demanda y elaboración de presupuesto
 - a. Proyecciones basadas en modelos tradicionales
 - b. Aplicación del análisis científico de datos
 - c. Proyecciones basadas en técnicas estocásticas mediante R
 - d. Elaboración de Presupuesto Anual
5. Comparación y análisis de resultados
 - a. Análisis de parámetros a comparar
 - b. Comparación y análisis de resultados
 - c. Conclusiones y recomendaciones



2.15 Cronograma

Objetivo específico	Actividad	Resultado Esperado	Tiempo (semanas)
1. Estudio de la Junta Administradora de agua potable de Baños	recopilación de información, entrevistas y análisis de requerimientos	descripción de la organización y lista de necesidades	2
2. Técnicas administrativas y estadísticas para tratamiento y proyección de datos	investigación	resumen de principales técnicas tradicionales	2
3. Nuevas tendencias de tratamiento de datos para uso en la Administración Empresarial	investigación	resumen de principales herramientas y técnicas actuales	2
4. Aplicación de modelos de proyección de la demanda y elaboración de presupuesto	diseño de modelos, aplicación de técnicas y modelos	Análisis de tendencias y presupuesto	4
5. Análisis y comparación de resultados	Elaboración de cuadro comparativo y Análisis de resultados	Tabla comparativa, conclusiones y recomendaciones	2

2.16 Referencias (APA)

Baca, U. (2013). Evaluación de proyectos (7ma ed): México: McGraw-Hill.

Berk, J., DeMarzo, Enríquez, J., & Valencia H. (2008). Finanzas corporativas. Naucalpan, Estado de México: Pearson Educación.

Bonini, C. (2000). Análisis cuantitativo para los negocios. Santafé de Bogotá: McGraw-Hill.

Burbano, J. (2011). Presupuestos, un enfoque de direccionamiento estratégico, gestión y control de recursos (4ta ed.). México: McGraw Hill.

Douglas, L., Marchal, W., Wathen, S., & Vizcaino, D. (2012). Estadística aplicada a los negocios y la economía: México D. F.: McGraw-Hill.

Horngrén, C., Datar, S., & Foster, G. (2007). Contabilidad de costos: un enfoque gerencial. Mexico, D.E.: Pearson educación.

Lantz, B. (2013). Machine Learning with R. Mumbai: Packt Publishing.

Maisel, L. (2014). Coking, Gary. Predictive Business Analytics. New Jersey: Wiley.

Ohlhorst, F. (2013). Big Data Analytics - Turning Big Data into Big Money. New Jersey: Wiley.

Sauter, V. (2010). Decision support systems for business Intelligence (2nd ed.). New Jersey: Wiley.

Schmidt, D., Ostrouchov, G. (2011). Programming with Big Data in R: Joint Institute for Computational Sciences, University of Tennessee: Oak Ridge National Laboratory.

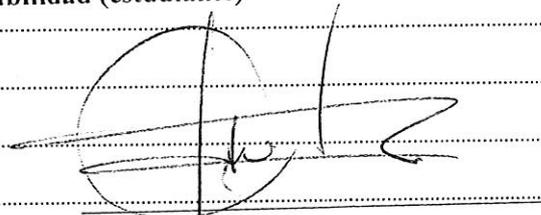
Turban, E. (2013). Decision Support Systems and Business Intelligence (9na ed.): New York, D.F.: Pearson.

Turban, E., Sharda, R., Delen, D., & King, D. (2010). Business Intelligence A Managerial Approach (2da ed.). New Jersey: Prentice Hall.

Yanchang Zhao, R. and Data Mining (2012). Examples and Case Studies: Elsevier.

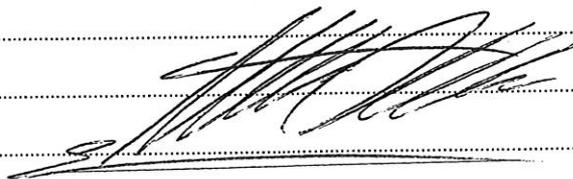
2.17 Anexos:

2.18 Firma de responsabilidad (estudiante)



Ing. Hernán Quito Baculima

2.19 Firma de responsabilidad (Director sugerido)



Ing. Marcos Orellana

2.20 Fecha de entrega:

Cuenca, Marzo 5 del 2015

Señor Ingeniero

Xavier Ortega Vásquez

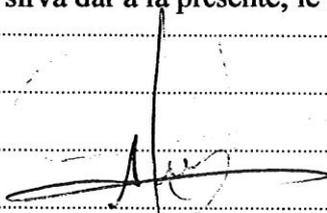
DECANO DE LA FACULTAD DE CIENCIAS DE LA ADMINISTRACION

De mi consideración:

Por medio de la presente, yo, Manuel Hernán Quito Baculima, con C.I. N° 0102748472, solicito de la manera más comedida la aprobación del diseño de tesis titulado **“Presupuesto financiero anual de la Junta Administradora de Agua Potable de Baños empleando proyecciones con la aplicación informática R y técnicas de Big Data ”**, esto debido a que se ha cumplido con los requisitos de actualización académica, necesarios para obtener el título de Ingeniero Comercial.

Por la favorable acogida que se sirva dar a la presente, le anticipo mi agradecimiento,

Atentamente,



Ing. Hernán Quito

C.I. 0102748472

COD. 28501



UNIVERSIDAD DEL
AZUAY

DOCTORA JENNY RIOS COELLO, SECRETARIA DE LA FACULTAD DE
CIENCIAS DE LA ADMINISTRACION DE LA UNIVERSIDAD DEL AZUAY

CERTIFICA:

Que, el señor **Manuel Hernán Quito Baculima**, luego de haber cumplido con todas las
asignaturas del pensum de la Escuela de Administración de Empresas y haber aprobado el
examen de suficiencia de Inglés el 20 de Noviembre de 2002, egresó de la Facultad el 10
de Julio de 2004

Cuenca, 26 de Septiembre de 2014

UNIVERSIDAD DEL
AZUAY
FACULTAD DE
ADMINISTRACION
SECRETARIA

Derecho No. 28501

scv.-