UNIVERSIDAD DEL AZUAY

Facultad de Ciencias de la Administración

Carrera de Ingeniería en Ciencias de la Computación

ENRIQUECIMIENTO ESTADÍSTICO DE UN CONJUNTO DE DATOS ESTRUCTURADO. CASO DE ESTUDIO: COMERCIO EXTERIOR DEL AZUAY 2022

Trabajo de titulación previo a la obtención del grado de Ingeniero en Ciencias de la Computación

Autor:
Marco Javier Játiva Bravo

Director:
Marcos Patricio Orellana Cordero

Cuenca – Ecuador
2025

DEDICATORIA

Dedico esta tesis a Dios y a la Virgen, como forma de agradecimiento por todos y cada uno de los favores concedidos a lo largo de mi vida universitaria.

De igualmente a mi familia, que han sido un pilar fundamental en mi vida estudiantil.

AGRADECIMIENTO

En primer lugar, agradezco a Dios, por acompañarme todos estos años, día a día, fue mi soporte y mi luz en cada momento de mi vida universitaria.

En segundo lugar, agradezco a mi familia, cuyos consejos y aliento fueron el combustible para no detenerme hasta ahora.

Finalmente, agradezco al Dr. Barton Zwiebach cuyas enseñanzas, cambiaron mi forma de ver el aprendizaje y el esfuerzo científico.

Índice de Contenidos

Índice de Figuras

Índice de Tablas

Índice de Anexos

Enriquecimiento estadístico de un conjunto de datos estructurado. Caso de estudio: comercio exterior del Azuay 2022

RESUMEN

Este estudio tuvo como propósito desarrollar una aplicación para el enriquecimiento estadístico de conjuntos de datos estructurados, aplicándose al caso del comercio exterior del Azuay en el año 2022. El trabajo se sustentó en fundamentos de análisis descriptivo, metadatos y clasificación de variables. Metodológicamente, se estableció un modelo de detección automática de tipos de variables y valores atípicos, incorporando medidas de tendencia central, dispersión y cardinalidad. La aplicación generó visualizaciones dinámicas y gráficos de relaciones entre columnas, mejorando la interpretación y la calidad de los datos. Los resultados evidenciaron que el enriquecimiento estadístico automatizado constituye una herramienta eficaz para optimizar la exploración y el análisis de datos, incluso en dominios con volúmenes moderados de información.

Palabras clave: análisis estadístico, comercio exterior, enriquecimiento de datos, estadística descriptiva, visualización de datos

Statistical enrichment of a structured data set. Case study: foreign trade of Azuay 2022

ABSTRACT

The purpose of this study was to develop an application for the statistical enrichment of structured data sets, applying it to the case of foreign trade in Azuay in 2022. The work was based on descriptive analysis, metadata, and variable classification. Methodologically, a model was established for the automatic detection of variable types and outliers, incorporating measures of central tendency, dispersion, and cardinality. The application generated dynamic visualizations and graphs of relationships between columns, improving the interpretation and quality of the data. The results showed that automated statistical enrichment is an effective tool for optimizing data exploration and analysis, even in domains with moderate volumes of information.

Keywords: statistical analysis, foreign trade, data enrichment, descriptive statistics, data visualization

1. Introducción

El enriquecimiento de datos en aplicaciones que procesan conjuntos de datos complejos presenta desafíos relevantes relacionados con la integración, la consistencia y la calidad de la información, lo que ocasiona una limitación en la capacidad de generar estadísticas descriptivas útiles desde la primera carga de datos. Existen casos, donde los sistemas no logran proporcionar información estadística clave, como lo es la distribución de variables y las medidas de tendencia central, además de que no muestran cómo se relacionan dichas variables, dificultando la comprensión inicial de un conjunto de datos y paralelamente introduciendo inconsistencias o errores en los metadatos. Problemas de esta índole afectan principalmente a la calidad de los datos disponibles para análisis y a la generación de sobrecarga cognitiva para los usuarios al momento de intentar interpretar información cruda o incompleta. Lo que impacta negativamente en el proceso de toma de decisiones en contextos donde la precisión es clave.

En el ámbito del comercio exterior, estos desafíos se hacen aún más evidentes debido a la existencia de alta cardinalidad y heterogeneidad de los datos involucrados. El laboratorio LIDI (Laboratorio de Investigación y Desarrollo en Informática) de la Universidad del Azuay ha trabajado con una base de datos proporcionada por el SENAE, que contiene aproximadamente 20.000 registros de exportaciones e importaciones realizadas por la provincia del Azuay en el año 2022, donde se identifica de manera clara, la necesidad de un enriquecimiento efectivo, con el objetivo de garantizar la calidad de los datos antes de cualquier acción de análisis.

La ausencia de un enriquecimiento estadístico eficiente en este tipo de aplicaciones recae en la necesidad de desarrollar herramientas que automaticen y optimicen este proceso, integrando técnicas que permitan generar estadísticas descriptivas desde el inicio, como la distribución de variables y las medidas de tendencia central, facilitando la comprensión de las relaciones entre ellas. En el caso del comercio exterior del Azuay, una solución que priorice el enriquecimiento de datos mejoraría la calidad de la información disponible, y proporcionaría una base sólida para la toma de decisiones estratégicas, además de potencia el uso efectivo de los datos en diferentes contextos.

1.1 Objetivos

Objetivo General

Desarrollar una aplicación para enriquecer estadísticamente un conjunto de datos estructurado, incluyendo gráficos de distribución y tendencia central, considerando como caso de estudio, el comercio exterior del Azuay del año 2022.

Objetivos Específicos

1. Examinar técnicas usadas para el enriquecimiento de datos a partir de un conjunto de datos en base al contexto del dataset, de la columna y a los datos de la columna.
2. Definir un modelo de clasificación de variables y enriquecimiento de datos mediante la identificación de criterios estadísticos y estructurales, estableciendo reglas de detección de patrones, relaciones entre variables y niveles de cardinalidad.
3. Desarrollar un repositorio que incorpora un sistema de clasificación automática de variables, generación de metadatos, incluyendo detección de valores atípicos y estructuras de datos consistentes.
4. Implementar mecanismo de visualización, que represente gráficos de relaciones entre variables.

1.2 Marco teórico

1.2.1 Enriquecimiento de Datos

El enriquecimiento de datos es un proceso clave en labores de gestión de la información, planteando como objetivo la mejora de calidad y utilidad de los datos existentes. Mariani et al., (2024) menciona que el enriquecimiento de datos ayuda a describir procedimientos que validan o integran información a un determinado conjunto de datos, pero a través de diferentes fuentes o reglas estadísticas. Mientras que Pyle et al., (1999) lo conceptualiza como el proceso de complementar o aumentar un conjunto de datos con información relevante, obtenida de fuentes internas o externas, para hacerlo más significativo. Aunque en palabras de Azad et al., (2018) el objetivo del enriquecimiento es mejorar, refinar y organizar datos, permitiendo extraer información valiosa que potencie su uso en análisis posteriores. Este proceso no solo implica añadir datos, sino garantizar la coherencia, relevancia y precisión de la información incorporada (Herzog et al., 2007), lo que lo

convierte en una práctica esencial para enfrentar los desafíos de calidad como en el ámbito del Big Data (Shirley Martillo-Alchundia & Alvarado-Zabala, 2025).

1.2.2 Tipos de Enriquecimiento de Datos

El enriquecimiento puede clasificarse de manera estructural donde se organiza y estandariza datos para mejorar su consistencia, como la normalización o integración de metadatos (Maharana et al., 2022). También puede darse de manera contextual, mediante la incorporación de información adicional para dar mayor profundidad, como agregar datos históricos o variables externas relevantes (Azad et al., 2018). Por otra parte, el enriquecimiento estadístico, sobre el cual se focaliza el presente estudio, se orienta a la generación de estadísticas descriptivas —como medidas de tendencia central o distribuciones— con el fin de facilitar el análisis (Kosikov et al., 2021). Estos tipos pueden combinarse dependiendo de las necesidades del conjunto de datos y los objetivos del análisis.

1.2.3 Medidas de Tendencia Central y Dispersión

1.2.3.1 Medidas de Tendencia Central

Las medidas de tendencia central permiten identificar valores alrededor de los cuales se concentran los datos. Estas métricas ofrecen una visión resumida de la distribución, facilitando así, la interpretación y la comparación entre diferentes grupos de datos. La Tabla 1, muestra las definiciones y fórmulas de cada una.

Tabla 1
Definiciones y Fórmulas de Medidas de Tendencia Central

Medida	Concepto	Fórmula
Media	Es la suma de todos los valores dividida por el número de observaciones. (Vetter, 2017)	$\boldsymbol{x} = \frac{\sum_{i=1}^{n} x_i}{n}$
Mediana	Es el número medio de un conjunto ordenado. Si $\boldsymbol{n}$ es impar: posición $\frac{n+1}{2}$ Si $\boldsymbol{n}$ es par: promedio de las observaciones centrales	
Moda	Valor que aparece con mayor frecuencia en el conjunto de datos, llegando a ser unimodal, bimodal, o multimodal. (Malakar, 2023)	$\boldsymbol{Mo} = El\ valor\ más\ repetido$
Rango Medio	Corresponde al punto central entre el valor mínimo y el máximo de la distribución. Aunque ofrece una medida sencilla de tendencia central alternativa a la media, es extremadamente sensible a valores atípicos, lo que limita su utilidad en distribuciones con alta dispersión (Bland, 2015).	$\boldsymbol{Rango\ medio} = \frac{x_{máx} - x_{mín}}{2}$

1.2.3.2 Medidas de Dispersión

Las medidas de dispersión cuantifican el grado de variabilidad de los datos respecto a su tendencia central. Permiten evaluar qué tan dispersos o concentrados están los valores, detectar posibles valores atípicos y comprender la distribución del conjunto. Entre las principales se encuentran la varianza, la desviación estándar, el rango y los cuartiles, cada una proporcionando información complementaria sobre la amplitud y la consistencia de los datos, la Tabla 2 detalla conceptos y fórmulas de las medidas.

Tabla 2
Definiciones y Fórmulas de Medidas de Dispersión

Medida	Concepto	Fórmula
Varianza	Representa el promedio de los cuadrados de las desviaciones de cada valor respecto a la media.	$\boldsymbol{s^2} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$
Desviación Estándar (s)	Corresponde a la raíz cuadrada de la varianza y expresa la dispersión de datos distribuidos normalmente Siendo una medida que representa la media de los datos de una muestra. (Darling, 2022)	$\boldsymbol{s} = \sqrt{s^2}$
Rango (R)	Diferencia entre el valor máximo y el mínimo de la distribución.	$\boldsymbol{R} = x_{máx} - x_{mín}$
Cuartiles	Dividen la distribución en cuatro partes iguales, lo que permite identificar concentración, forma y posibles sesgos (Groth & Bergner, 2006).	$\boldsymbol{Q1} = \frac{n+1}{4}$ $\boldsymbol{Q2} = \frac{n+1}{2}$; $\boldsymbol{Q3} = \frac{3(n+1)}{4}$
Rango Intercuartil (IQR)	Expresa la amplitud de los valores centrales del conjunto de datos, calculada como la diferencia entre el tercer y el primer cuartil.	$\boldsymbol{IQR} = Q_3 - Q_1$

1.2.4 Pruebas Estadísticas de Supuestos

En los estudios que utilizan datos numéricos, es crucial asegurarse de que la información cumpla con ciertas condiciones básicas antes de aplicar métodos estadísticos avanzados. Estas condiciones, llamadas supuestos, incluyen aspectos como que los datos sigan una distribución normal, que las variaciones entre grupos sean similares o la independencia entre variables. Verificar su cumplimiento resulta esencial porque de ello depende la validez de los resultados y la confianza en las conclusiones obtenidas.

En el caso de la normalidad, existen pruebas que permiten comprobar si los datos se ajustan a este patrón. La más utilizada es la prueba de Shapiro-Wilk, que destaca por su capacidad para detectar desviaciones en muestras pequeñas y por su fiabilidad en contextos donde otras técnicas pueden perder sensibilidad (González-Estrada et al., 2022). Su estadístico se expresa como:

$$W = \frac{\left(\sum_{i=1}^{n} a_i X_{(i)}\right)^2}{\sum_{i=1}^{n} (X_i - X)^2}$$

Donde los coeficientes a_i dependen de la media y la varianza de la muestra, y valores de W cercanos a uno sugieren normalidad. Para muestras más grandes, esta verificación puede complementarse con la prueba de D´Agostino-Pearson, la cual combina medidas de asimetría y curtosis para ofrecer una visión más completa sobre la forma de la distribución (Korkmaz & Demir, 2023). Su lógica se sintetiza en la estadística:

$$\chi^2 = Z(\sqrt{b_1})^2 + Z(b_2)^2$$

que sigue aproximadamente una distribución χ^2 con dos grados de libertad.

Otro supuesto clave es la homogeneidad de varianzas, indispensable cuando se comparan distintos grupos de datos. Para ello, la prueba de Levene se considera una de las opciones más robustas, ya que evalúa si las dispersiones son equivalentes y mantiene un buen desempeño incluso cuando los datos no cumplen estrictamente con la normalidad (Sritan & Phuenaree, 2021).

Cuando se trabaja con variables categóricas, la atención se centra en la independencia entre ellas. Aquí se recurre a la prueba de χ^2 de independencia, que contrasta las frecuencias observadas con las esperadas bajo la hipótesis de no relación El estadístico se define como:

$$\chi^2 = \sum \frac{(O_{rc} - E_{rc})^2}{E_{rc}}$$

donde O_{ij} son las frecuencias observadas y E_{ij} las esperadas. Si los resultados muestran diferencias significativas, puede inferirse que las variables están asociadas de manera estadísticamente relevante (Nihan, 2020).

Finalmente, una vez que se han verificado estos supuestos, es posible aplicar análisis comparativos más exigentes, como el ANOVA de un factor. Esta prueba permite establecer si existen diferencias significativas entre las medias de tres o más grupos. Su validez, depende directamente de la normalidad y la homogeneidad de varianzas, lo que evidencia la importancia de los pasos previos (Alassaf & Qamar, 2022)

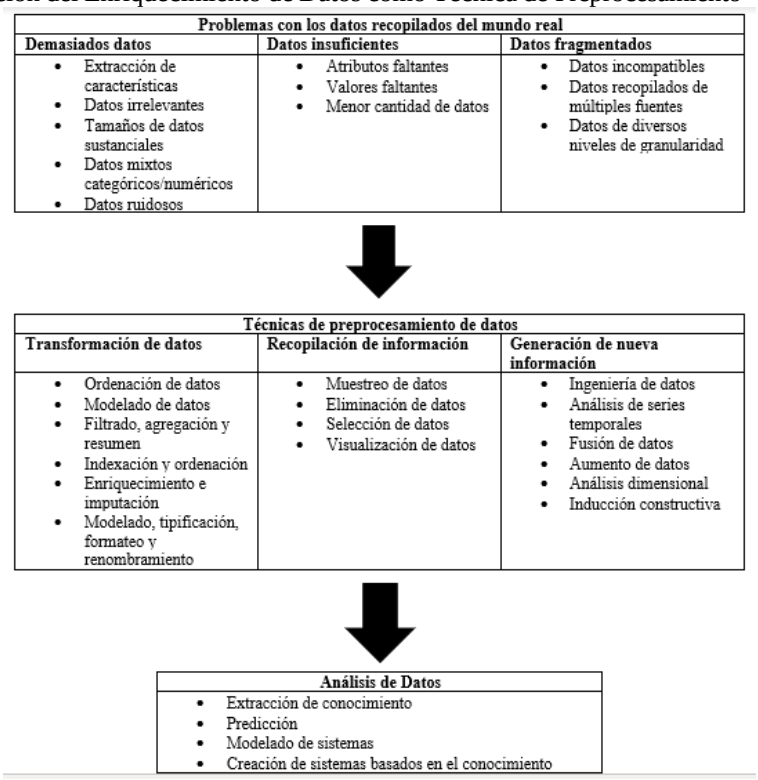
En conjunto, estas herramientas metodológicas no son un requisito accesorio, sino un pilar para asegurar que el análisis estadístico sea consistente y que las conclusiones reflejen de manera fiel los casos estudiados.

1.2.5 Estructura del Proceso de Enriquecimiento

El enriquecimiento de datos sigue una estructura sistemática que forma parte del preprocesamiento de datos. Según Maharana et al., (2022), este proceso incluye etapas como la limpieza de datos, la integración de fuentes externas, la transformación y validación de información incorporada. En el caso de enriquecimiento estadístico, el cálculo de métricas e identificación de relaciones entre variables se añaden al resto de procesos. Ayudando así la coherencia y utilidad de los datos resultantes. En la Figura 1 se observan todas las fases y procesos planteados por el autor.

Figura 1

Estructura y Ubicación del Enriquecimiento de Datos como Técnica de Preprocesamiento



Nota: Basado en "A review: Data pre-processing and data augmentation techniques", por K. Maharana, S. Mondal, B. Nemade, 2022, Global Transitions Proceedings, Volume 3, p. 91-99, https://doi.org/10.1016/j.gltp.2022.04.020.

1.2.6 Aplicaciones y Caso de Estudio

El enriquecimiento se aplica en diversos contextos, como en Big Data, donde ayuda a manejar grandes volúmenes de datos mejorando su utilidad para identificar grupos y relaciones (Kosikov et al., 2021). De igual manera, en el aprendizaje automático, donde

junto con procesos como el aumento de datos, el enriquecimiento mejora el poder de generalización de modelos al añadir profundidad y variabilidad a los datos (Mumuni & Mumuni, 2024). Por otro lado, en sistemas de transporte inteligente (ITS, por sus siglas en inglés), donde se fusionan con datos heterogéneos de fuentes como redes sociales para enriquecer descripción de rutas, mejorando la gestión del tráfico (Rettore et al., 2020).

Asimismo, en gestión de datos personales y ontologías, el enriquecimiento de datos se aplica en temas demográficos y geográficos, permitiendo a empresas optimizar ventas y extraer conocimiento semántico de datos heterogéneos, aumentado el valor y precisión de la información, acción clave para decisiones comerciales (Malyavko et al., 2022).

Paralelamente, ofrece ventajas significativas como la mejora de la calidad y relevancia de la información, permitiendo extracción de conocimiento más profundo y útil (Azad et al., 2018), además de facilitar la integración de datos actualizados desde fuentes externas en tiempo real, ayudando en contextos como la carga de red (Ritze & Eckert, 2014). Sin embargo, presenta desafíos como la dificultad de garantizar la coherencia y precisión de datos incorporados, especialmente cuando se integran fuentes heterogéneas. Además, se debe ser cuidadoso con este proceso para evitar introducir ruido o sesgos en el conjunto de datos.

La importancia del enriquecimiento de datos reposa en el hecho que permite ampliar un conjunto de datos con información relevante procedente de repositorios externos, a menudo desconocidos. Además, implica alinear e integrar datos con el conjunto de datos principal para rellenar con detalles faltantes y enriquecer su contenido (Hristov et al., 2024). Su relevancia a llegado a tal punto, que a lo largo de estos últimos años se han planteado numerosas iniciativas para apoyar la preparación de datos, enfatizando tareas de enriquecimiento, con la finalidad de dar más profundidad a los datos y obtener los mejores resultados.

El estudio de caso aplica al comercio exterior de la provincia del Azuay en el año 2022, gestionados por el SENAE (Servicio Nacional de Aduana del Ecuador). El SENAE publica periódicamente datos relacionados con importaciones y exportaciones realizadas por las provincias del Ecuador, con objetivos que incluyen la facilitación del comercio

exterior y la recaudación de obligaciones tributarias, según su plan estratégico (SENAE, 2015). Entre sus funciones, definidas en la Ley Orgánica de Aduanas, (2006) está el almacenamiento, verificación, valoración, recaudación y vigilancia de mercancías, lo que genera datos detallados cruciales para el análisis de comercio exterior, que Jankulovski et al., (2017) define como transacciones planificadas con antelación, que abordan las fronteras nacionales para la satisfacción de necesidades de todo tipo de individuo y empresa.

El dataset analizado incluye 20,000 registros con 77 variables, entre las que destacan el código arancelario, basado en el Sistema Armonizado (Harmonized System, HS por sus siglas en inglés), que Shumba, (2024) define como un estándar internacional para clasificar mercancías mediante códigos numéricos de seis dígitos; la descripción del producto; el valor FOB, definido por (Soni, 2014) como un término que engloba las responsabilidades de comprador y vendedor en un contrato de comercio internacional; el país de origen y destino; y el tipo de flujo comercial. En este contexto, el enriquecimiento de datos se vuelve esencial para mejorar la calidad de la información, permitiendo un análisis más preciso de las actividades comerciales del Azuay.

2. Revisión de literatura

El enriquecimiento se ha convertido en un eje central en la investigación contemporánea, especialmente por la capacidad que ofrece potenciando procesos de análisis y visualización de grandes volúmenes de información, adhiriéndose a disciplinas como Big Data y técnicas enriquecimiento y aumento de datos. En este contexto, se han identificado estudios relevantes que abordan desde enfoques teóricos hasta implementaciones prácticas en sistemas reales. Ofreciendo así, un panorama sobre técnicas, desafíos y oportunidades del enriquecimiento, construyendo una base sólida para explorar su aplicación en análisis de datos de comercio exterior. A continuación, se presenta una serie de análisis de estos estudios, sus técnicas, aportes y resultados.

Entre las diversas formas en que los estudios abordan el enriquecimiento de datos, destaca el trabajo de Scheinert et al., (2023) donde se analizan diversas técnicas de enriquecimiento de datos aplicadas a sistemas de procesamiento distribuido en tiempo

real. El foco del estudio es comparar métodos como el uso de joins en tiempo de ejecución, caching y replicación de datos, evaluando impacto en latencia, rendimiento y escalabilidad del sistema. Todo esto probado en entornos simulados y reales, donde presentan pros y contras dependiendo del volumen de datos y la arquitectura del sistema.

Zhou et al., (2023) proponen al enriquecimiento de datos como un paso crucial en la etapa de preprocesamiento, con el objetivo de mejorar la predicción a corto plazo de energía eólica. A lo largo del estudio se integran variables meteorológicas tales como: temperatura, humedad, presión, con datos históricos de generación de energía obtenida con técnicas de combinación de fuentes heterogéneas, interpolación de datos faltantes y normalización. Los resultados muestran que el uso de datos enriquecidos permite no solo capturar mejores patrones temporales sino mejorar significativamente la precisión de modelos de aprendizaje automático.

El enriquecimiento de datos no es solo proceso de interés para las ingenierías, desde el mundo de la manufactura y la producción, Shyalika et al., (2024) realiza una examinación a la manera en que el enriquecimiento de datos puede mejorar el análisis de eventos poco frecuentes dentro del sector manufacturero. Los autores investigan técnicas de enriquecimiento como la fusión de datos de múltiples sensores, integración de información contextual del entorno de producción e incorporación de datos históricos de mantenimiento. Haciendo uso de máquinas de soporte vectorial, el estudio evalúa el impacto de técnicas de detección y predicción de eventos raros, arrojando como resultados la mejora significativa que aporta el enriquecimiento de datos a la precisión de modelos predictivos.

La creación de herramientas para el enriquecimiento de datos, es una práctica que hoy en día recolecta datos de múltiples fuentes, como lo menciona Sánchez et al., (2023) que presenta Data Enrichment Toolchain (DET), una plataforma diseñada para armonizar y enriquecer datos provenientes de fuentes heterogéneas como dispositivos IoT, portales de datos abiertos y redes sociales. La plataforma permite organizar y combinar la información obtenida de manera estructurada, facilitando su comprensión y por ende su análisis, haciendo uso de técnicas de procesamiento de datos e inteligencia artificial (IA) con el objetivo de mejorar la calidad y el contexto de los datos, los resultados evidencian

que DET no mejora solo la eficiencia, sino que es precisa en la integración de la información en cualquier tipo de entorno.

Incluso la implementación del enriquecimiento de datos ha sido añadida en nuevos marcos de trabajo de Big Data como lo detalla Wang & Carey, (2018) específicamente en Apache AsterixDB, un marco que permite incorporar información adicional a los datos de entrada, mejorando su valor en análisis posteriores. Para conseguirlo se usan funciones definidas por el usuario, mismas que pueden ser implementadas en Java o SQL++, facilitando así, operaciones de enriquecimiento (integración de datos de referencia o aplicación de modelos de AA), las evaluaciones de rendimiento muestran que la solución mejora temas de eficiencia y escalabilidad en la introducción de datos enriquecidos en entornos Big Data.

Se da, de la mano de Ghosh, Gupta, Mehrotra, & Sharma, (2022) EnrichDB, un sistema de gestión de bases de datos (SGBD) diseñado para integrar funciones de enriquecimiento de datos dentro del flujo de procesamiento, EnrichDB incorpora funciones dentro del sistema, con el objetivo de permitir que el enriquecimiento ocurra al momento de almacenar datos o ejecutar consultas, mejorando la eficiencia y reduciendo el tiempo necesario para obtener datos enriquecidos, una de las grandes ventajas es la puesta en segundo plano del enriquecimiento al mismo tiempo que se procesan consultas, adaptándose al intenso análisis de datos en tiempo real.

Trabajo más directo, pero no por eso menos preciso viene de parte de Dong & Oyamada, (2022) que presentan un sistema diseñado para enriquecer tablas de consulta con columnas adicionales provenientes de fuentes externas, con el objetivo de mejorar el rendimiento de modelos de aprendizaje automático. Todo el sistema se basa en cuatro etapas: búsqueda de filas para unir (tabla de consulta y tablas relacionadas), selección de tablas relevantes (para tarea específica de aprendizaje, identificando las que aportan atributos útiles para el modelo), alineación de filas y columnas (asegurando coherencia y calidad de datos combinados) y por último selección y evaluación de características (determinando variables más impactantes para el rendimiento del modelo). Los resultados muestran que este enfoque mejora la precisión de modelos predictivos, aprovechando riqueza de información disponible en lagos de datos, paralelamente, el sistema ofrece una

interfaz web fácil de usar, permitiendo así, la fácil integración de tablas en flujos de trabajo.

De la mano de Gong et al., (2024) viene Delta, un sistema diseñado para el mejorar el aprendizaje continuo dentro de dispositivos móviles (teléfonos o sensores), que en su mayoría poseen pocos datos y recursos limitados, Delta permite un enriquecimiento de datos locales haciendo uso de información almacenada en la nube de carácter: relevante. Además, se propone un marco de trabajo de cuatro etapas: dividir el proceso entre dispositivo y nube (protegiendo privacidad), usar una estrategia flexible para emparejamiento de datos del dispositivo con ejemplos en la nube, seguido de, seleccionar que datos se devolverán al dispositivo para mejorar el aprendizaje, finalmente, equilibrar el uso de datos antiguos, con el fin de que el modelo no desconozca lo aprendido. Toda esta labor demostró que Delta es una alta solución efectiva en el entrenamiento de modelos inteligentes de pequeños dispositivos.

La importancia del enriquecimiento influye incluso a nivel de información geoespacial, Hristov et al., (2024) abordan y comparan dos enfoques de enriquecimiento SemTui (marco) y QGIS (plataforma), haciendo uso del servicio HERE Geocoder, SemTui emplea una técnica de enriquecimiento semántico interactivo añadiendo información suplementaria (duración y longitud de rutas) y logrando una precisión del 95% en direcciones residenciales. Mientras que QGIS, hace uso de MMQGIS (plugin con Open Street Map y Google) como modelo de enriquecimiento. Em ambos casos el enriquecimiento de datos demuestra ser altamente efectivo para mejorar no solo calidad sino la utilidad de la información, siendo SemTUI la de mayor precisión.

En cuanto a técnicas más detalladas se encuentra JENNER (Ghosh, Gupta, Mehrotra, Yus, et al., 2022) que enriquece datos justo cuando se deben responder consultas, haciendo uso de herramientas de IA (SVM,KNN) para analizar sentimientos en tuits o algoritmos (LOC_n) para ubicar personas con datos WiFi, ajustando precisión según el tiempo disponible. Aquí se divide el trabajo en “épocas” (pequeños pasos), en cada una, decide la importancia de los datos con el objetivo de mejorar y aplicar herramientas más adecuadas, utilizando como guía RelativeBenefit cuya labor es calcular cuánto se ganará, en términos de respuesta, por el esfuerzo invertido.

JENNER paralelamente reduce la incertidumbre de datos mediante un enfoque matemático específicamente la entropía, mismo que ha sido probado con millones de registros (10 millones de WiFi, y 11M de tuits) y ofrece un 90% de calidad en poco tiempo (25s). JENNER es una brillante herramienta en términos de eficiencia y mejora constante, siendo idóneo para el análisis de datos sin momentos de espera.

Una solución con una técnica de enriquecimiento más específico está en el trabajo de (Jegierski & Saganowski, 2020) en la resolución de problemas de datos desbalanceados; Cuando se intentan detectar fraudes o enfermedades, en lugar de, crear datos artificiales o eliminarlos, se usan datos reales de un conjunto externo. Se proponen tres formas de hacerlo: siendo un enriquecimiento randómico (RanE), donde se agregan ejemplos al azar, Semi-greedy Enrichment (SemE), que prueba uno a uno y solo se queda con los que mejoran el clasificador (usa como base un Random Forest) y finalmente un enriquecimiento supervisado (SupE), que elige exclusivamente solo ejemplos clave para separar mejor las clases.

Esta solución se probó en 10 conjuntos reales, siendo Twitter y el MIT grandes referencias, obteniendo como resultados, la importancia del enriquecimiento de datos y su posición clave para el problema del desbalance de clases (CIP) siendo crucial en escenarios críticos donde los datos son escasos o desbalanceados, mejorando calidad de predicciones con la gran ventaja de no introducir ruido artificial.

Singh et al., (2021) presentan un modelo de aprendizaje automático para predecir el consumo energético de edificios que están en etapas tempranas de diseño, en donde el problema principal son el alto costo de las simulaciones tradicionales. Los autores proponen dos enfoques para la recolección de datos de entrenamiento: En primer lugar, está el Incremento cuya labor es agregar más muestras de un solo modelo de edificio variando parámetros como el área y la altura y, en segundo lugar, está el Enriquecimiento, que usa múltiples modelos con diferentes formas para obtener diversas muestras. El modelo de machine learning, se entrena con datos simulados por EnergyPlus y usa un análisis de componentes principales (PCA, por sus siglas en inglés) para el manejo de formas complejas.

Otorgándole un papel más protagónico al enriquecimiento de datos (Cutrona, 2019), presenta ASIA una herramienta que enriquece tablas a gran escala (en proyectos de ciencia de datos, donde la tarea de preparación consume un 80% del tiempo) conectándolas con bases de conocimiento externas como GeoNames, esto , puesto a prueba, con un dataset de 2227 nombres de ciudades, ASIA tomó estos valores y los extendió con datos climáticos, logrando un tiempo de ejecución de 0.0168 ms por fila. Otro ejemplo se dio con un dataset masivo de 100 GB (alrededor de 500M filas), mantuvo una escalabilidad lineal (R2=0.998) en un clúster de 8 nodos, enriqueciendo datos de manera eficiente. Comparado con OpenRefine, ASIA destaca en términos de velocidad y automatización en entornos Big Data, haciendo del enriquecimiento una labor más fácil y sin la necesidad de intervención manual.

Creando una relación más cercana al mundo del internet de las cosas (IoT, por sus siglas en inglés) Park et al., (2021) propone ThingsMetadata, un esquema basado en XML para enriquecer datos en entornos de IoT, permitiendo describir sensores y robots y procesar sus datos para servicios inteligentes. La técnica organiza el enriquecimiento en cuatro niveles: 1) ThingsMetadata, que define perfiles y tipos de datos; 2) contexto de bajo nivel, que convierte datos crudos a unidades legibles; 3) contexto de alto nivel, que transforma valores numéricos en estados descriptivos y 4) relaciones situacionales, que genera tripletas (sujeto, predicado, valor) para representar estados comprensibles. Probado en un escenario con un robot Turtlebot, enriqueció datos de sensores para detectar si tazas de té fueron tomadas, generando tripletas en tiempo real que activaron servicios de forma precisa.

Comparado con enfoques SOSA y SSN, ThingsMetadata destaca por flexibilidad y capacidad para integrar datos heterogéneos en IoT, mejorando legibilidad para sistemas conscientes del contexto. Los estudios revisados destacan el enriquecimiento de datos como un proceso crucial para transformar información cruda en conocimiento útil, siendo parte de varios tipos de aplicaciones, como modelos predictivos de energía y manufactura hasta la optimización de sistemas distribuidos y entornos Big Data. De igual manera se evidencia que existe un interés constante en el desarrollo de herramientas y plataformas automatizadas para la integración de datos heterogéneos, así como en técnicas para contextos específicos como IoT, geoespacialidad y dispositivos móviles. Las investigaciones subrayan la importancia de adaptar estrategias de enriquecimiento de

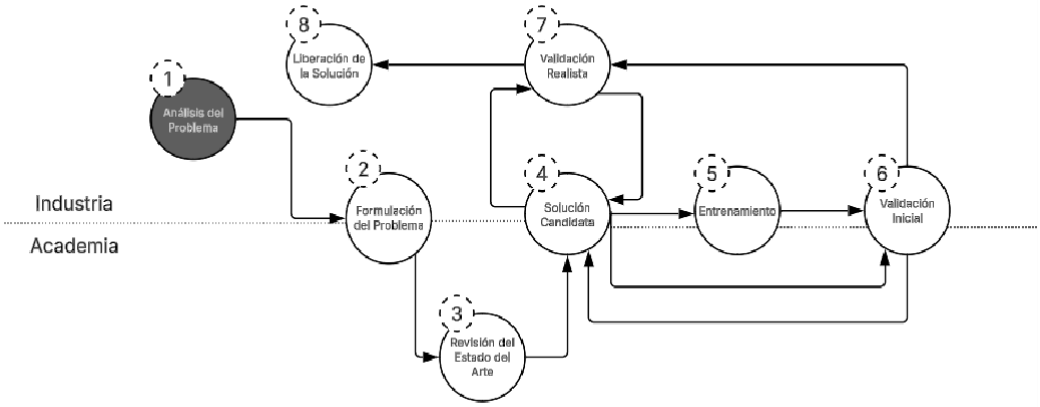
datos a las necesidades del análisis, como mejorar escalabilidad o potenciar interpretación de datos complejos, aspectos que son claves en dominios donde calidad y contexto son esenciales, como el comercio exterior.

La presente investigación, centra al enriquecimiento de datos como un núcleo y en la creación de un visualizador como resultado, comparte con los estudios revisados la meta de generar valor a partir de datos crudos, apoyándose en enfoques de enriquecimiento estadístico. Sin embargo, se diferencia al aplicar estas técnicas ya no en dominios IoT o Big Data sino en el análisis de comercio exterior, específicamente al de la provincia del Azuay en 2022, un ámbito menos explorado que requiere un enfoque claro. El trabajo se fundamenta en la necesidad de enriquecer datos a través de métodos estadísticos, permitiendo, como un resultado secundario, la creación de gráficos de distribución y tendencia central, facilitando así su interpretación. El aporte principal de esta investigación reposa en desarrollar una aplicación que optimiza el enriquecimiento estadístico de datos, ofreciendo así, una solución práctica para el comercio exterior ecuatoriano, que puede ser aplicado a distintos contextos y contribuir a la generación de información útil.

3. Métodos

Este proyecto está guiado bajo la metodología propuesta por Gorschek & Larsson (2006), la cual integra un enfoque iterativo que articula la investigación teórica con validación empírica, de esta manera, asegura la aplicabilidad práctica de la solución tecnológica. Se muestra a continuación, en la Figura 2, el diagrama de las fases de la metodología Gorschek.

Figura 2
Metodología de Gorschek et al. (2006)



Nota: Basado en "A Model for Technology Transfer in Practice", por T. Gorschek, P. Garre, S. Larsson y C. Wohlin, 2006, IEEE Software,23 (6), p. 88-95, https://doi.org/10.1109/MS.2006.147.

3.1 Tipo de Investigación

La presente investigación es de tipo exploratoria-descriptiva, busca comprender el entorno actual en torno al manejo y análisis de datos tabulares, específicamente en cuanto a sus limitaciones estadísticas, estructurales y de visualización. A partir de ello, se describen las características y aportes de una solución informática diseñada para enriqueces estadísticamente este tipo de datos, mejorando su utilidad analítica.

3.2 Enfoque

El estudio adopta un enfoque cuantitativo, debido a que se trabaja directamente con datos estructurados, en especial en formato tabular. Estos datos permiten aplicar medidas estadísticas, generar visualizaciones e identificar las características estructurales de cada variable, como lo sería su tipo de dato, cardinalidad o distribución. Como caso de aplicación se emplea un dataset real relacionado al comercio exterior de la provincia del Azuay durante el año 2022.

3.3 Universo, Población y Muestra

El universo de esta investigación está compuesto por conjuntos de datos tabulares utilizados en entornos académicos, técnicos o administrativos, que requieren un análisis

estadístico o enriquecimiento de metadatos. La población está representada por datos públicos y estructurados relacionados con el comercio exterior de la provincia del Azuay. Para la muestra, se han considerado lo registros correspondientes al año 2022, obtenidos de fuentes oficiales como el Servicio Nacional de Aduana del Ecuador (SENAE).

3.4 Recolección de Datos

La recolección de datos se llevó a cabo mediante un enfoque secundario, haciendo uso de archivos previamente disponibles en plataformas oficiales. En particular, se accedió a archivos en formato CSV que contienen información estructurada relacionada con el comercio exterior de la provincia del Azuay. Estos datos fueron seleccionados por su pertinencia al contexto del caso de estudio y por estar disponibles públicamente a través de instituciones como el Servicio Nacional de Aduana del Ecuador (SENAE).

3.5 Fuentes de los Datos

Todos los datos empleados en esta investigación fueron obtenidos de entidades oficiales. Específicamente, se ha hecho uso de los registros obtenidos del Servicio Nacional de Aduana del Ecuador (SENAE), correspondientes al segmento de comercio exterior, con cobertura geográfica en la provincia del Azuay y temporal en el año 2022.

3.6 Caracterización de Atributos

Una vez comprendida la terminología básica, se procede con la caracterización de atributos presentes en los datos, utilizando diversas medidas estadísticas, de esta manera se podrá identificar patrones, comportamientos típicos y anomalías. Mediante métricas como mediana, moda, varianza, etc., se obtiene una visión más profunda del comportamiento de cada atributo, facilitando así el diseño de soluciones que respondan a la realidad reflejada en los datos. Las tablas 3,4 y 5 comprenden el proceso de caracterización, mostrando la descripción y el tipo de dato, hasta la obtención de medidas estadísticas.

Tabla 3

Descripción y Tipo de Dato

Atributo	Descripción	Tipo de Dato
UNIDADES	Número total de unidades enviadas	numérico (float64)
KILOS_NETO	Peso neto en kilogramos	numérico (float64)
FOB	Valor Free On Board (USD)	numérico (float64)
FLETE	Costo del transporte (USD)	numérico (float64)
SEGURO	Costo del seguro (USD)	numérico (float64)
CIF	Valor Cost, Insurance & Freight (USD)	numérico (float64)
DIAS_SALIDA	Días entre embarque y llegada	numérico (int64)

Tabla 4

Medidas de Tendencia Central y Valores Nulos y Atípicos

Variable	Media	Mediana	Moda	Rango Medio	Valores Nulos	Valores Atípicos
UNIDADES	4.27 K	6.0	1.0	16.71 M	0	6.99 K
KILOS_NETO	3.96 K	3.64	0	18.63 M	0	6.88 K
FOB	7.42 K	108.47	100.0	24.59 M	0	6.25 K
FLETE	68.59	1.9	0	11.57 K	0	6.44 K
SEGURO	41.56	0.43	0	245.90 K	0	4.92 K
CIF	5.67 K	87.69	0	2.48 M	0	5.89 K
DIAS_SALIDA	69.51	1.0	0	4.66 K	0	3.95 K

Nota: K = miles; M = millones. N/A significa no aplica, no existe medida de tendencia central para esa variable. En la sección de Anexos, se encuentran el resto de atributos de manera más específica.

Tabla 5

Medidas de Dispersión

Atributo	Varianza	Desviación Estándar	Rango	Cuartiles		Rango Intercuartil
				Q1	Q3	
UNIDADES	47.59 M	218.15 K	33.43 M	1	50	49
KILOS_NETO	70.72 M	265.93 K	37.25 M	0.86	35.78	34.927
FOB	167.62 M	409.42 K	49.18 M	33.73	480.69	446.95
FLETE	264.03 K	513.83	23.13 K	0	8.38	8.38
SEGURO	16.58 M	4.07 K	491.80 K	0.05	1.47	1.42
CIF	170.28 M	412.65 K	49.67 M	19.95	325.23	305.27
DIAS_SALIDA	539.29 K	734.36	9.31 K	0	4	4

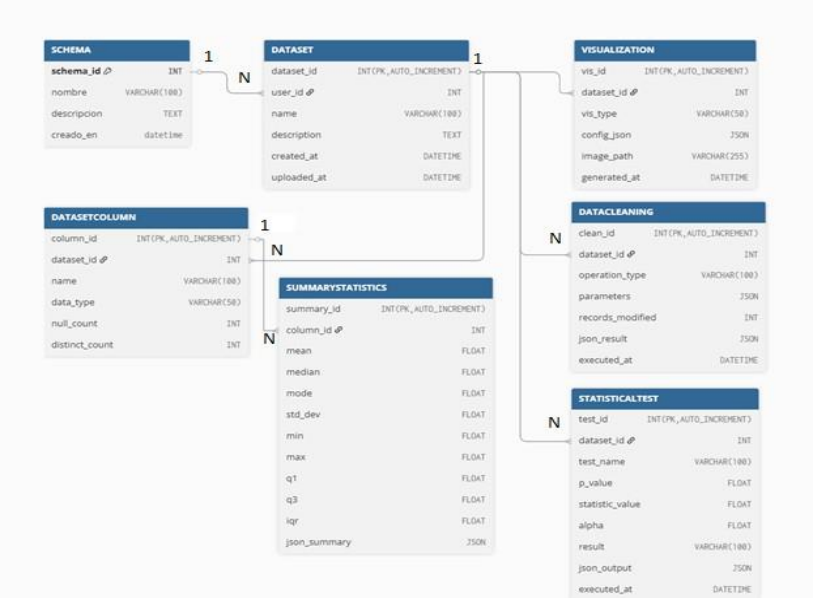
3.7 Diagramación del Modelo Entidad Relación

El modelo entidad-relación (MER) permite representar de forma ordenada y coherente diferentes elementos que intervienen en el procesamiento y análisis de datos. A partir de este diagrama, se definen entidades principales que participan en labores de gestión, limpieza y evaluación estadística de los datos, junto con sus respectivas relaciones e interdependencias. Así, el MER actúa como el marco lógico que organiza la información que el sistema manipula, facilitando su comprensión, seguimiento y la integración con herramientas analíticas o visuales.

Si bien el modelo entidad-relación fue concebido inicialmente como una estructura conceptual, su forma final surgió a partir del comportamiento real del sistema y de los resultados generados durante el desarrollo del programa. A medida que se implementaron procesos de carga, limpieza e imputación de datos, el diagrama fue ajustándose para reflejar fielmente la dinámica observada en los archivos JSON producidos por el sistema;

La Figura 3 refleja el estado en el que se manejan todos los actores involucrados en el enriquecimiento de datasets.

Figura 3
Esbozo de ER (Modelo Entidad-Relación) sobre el Diccionario de Datos



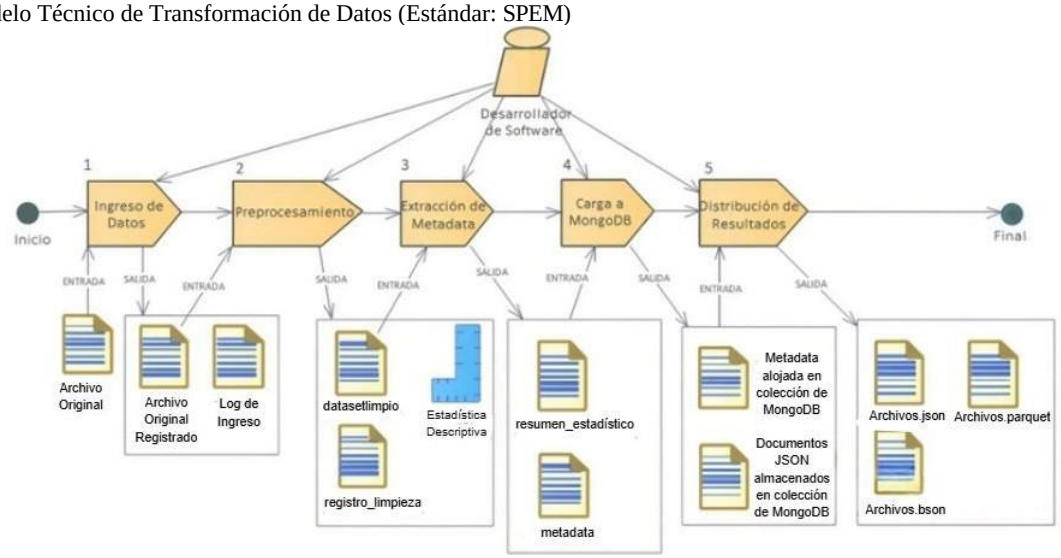
Cada entidad dentro del ER representa un momento del flujo de análisis. Dataset describe el punto de partida, Data Cleaning refleja la depuración y corrección de valores, mientras que StatisticalTest y SummaryStatistics documentan la validación y el enriquecimiento de los datos. Finalmente, Visualization materializa los resultados a través de gráficos y resúmenes visuales. Este recorrido dentro del modelo es más que una estructura, es una narrativa que traduce el viaje de los datos crudos hacia la información interpretada, vinculando lo técnico con lo analítico de manera fluida.

El paso del ER hacia su representación en JSON se apoya en una condición estructural de equivalencia: cada entidad del ER se transforma en un objeto JSON, sus atributos se traducen en pares clave-valor, y las relaciones se materializan como arreglos o referencias anidadas. De esta manera, se logra mantener la coherencia entre el diseño lógico y la implementación real del sistema. A nivel práctico, la conversión funciona como un "puente" entre dos lenguajes –el conceptual y el programático—donde el primero describe las conexiones y el segundo las ejecuta. Así, el modelo no solo se interpreta visualmente en diagramas, sino que se proyecta en estructuras operativas con la capacidad de evolucionar junto con los datos y los procesos estadísticos que los acompañan.

3.8 Modelo Técnico de Procesos

Tras el planteamiento metodológico general, se definió un modelo técnico de transformación de datos, estructurando bajo el estándar SPEM 2.0 (Software Process Engineering Metamodel). Este modelo permite representar actividades, tareas, roles y artefactos involucrados en el flujo de trabajo técnico, desde el ingreso del dataset hasta su análisis estadístico, enriquecimiento y posterior almacenamiento. La Figura 4 presenta todas las etapas que conforman el diagrama y sus artefactos relacionados.

Figura 4
Modelo Técnico de Transformación de Datos (Estándar: SPEM)



3.8.1 Fase 1: Ingreso de Datos

El objetivo de esta etapa es verificar que la integridad y compatibilidad técnica no contenga errores para las fases posteriores. La fase inicial se centró en el ingreso del dataset principal, que constituyó el insumo base de todo el proceso metodológico. El archivo utilizado, denominado *muestra_comex_ec_2022.xlsx* de 13 MB contiene 77 variables, constituye el insumo inicial sin preprocesamiento previo.

Para ello, se desarrollaron microactividades específicas que incluyeron la validación del archivo, comprobando que su extensión y formato fueran adecuados y que no existieran errores en su estructura. Adicionalmente, se revisó la coherencia de las columnas, asegurando que los nombres y tipos de datos se ajustaran a lo esperado y pudieran ser interpretados correctamente en las fases de limpieza y análisis. Los artefactos generados son: El archivo original, preservado como referencia y el registro de ingreso, que tiene la documentación de la fuente, tamaño y estado general del dataset.

3.8.2 Fase 2: Preprocesamiento

Consiste primero de un análisis exploratorio que extrae la naturaleza y estructura de los datos antes de transformarlo, y luego se depura y organiza el dataset, reduciendo redundancias y preparando los datos para análisis posteriores. En esta etapa se realiza una lectura adaptativa de archivos (CSV, Excel). La Tabla 6 contiene las microactividades que se realizan en esta fase y los resultados que aportan al dataset.

Tabla 6
Microactividades del Preprocesamiento

Microactividad	Resultado
Eliminación de columnas vacías o irrelevantes	Reduce ruido y mejora eficiencia.
Eliminación de duplicados	Garantiza unicidad de registros.
Ajuste de tipos de datos (datetime, numérico, texto)	Facilita cálculos y transformaciones.
Control de valores nulos	Evita errores y sesgos en análisis, imputando o eliminando valores nulos según el criterio.
Generación de archivos dataset_limpio y registro_limpieza	Documenta transformaciones aplicadas, brindando un dataset refinado y registro técnico disponible de 37.2 MB.

3.8.3 Fase 3: Enriquecimiento Estadístico y Generación de Metadata

La tercera fase enriquece a la información de un nivel adicional de análisis que permite describir con precisión las características de cada atributo, al mismo tiempo que se creaba un insumo documental en formato JSON para su posterior gestión y consulta.

El proceso comienza con el análisis descriptivo de cada columna, donde se calculan medidas de tendencia central, dispersión y rangos intercuartílicos. Se identifican posibles valores atípicos en variables numéricas y se documenta su frecuencia y rango, permitiendo evaluar la calidad y consistencia de la información. De manera simultánea se clasifican cada atributo según el tipo, a continuación, se enlistan:

1. **Numérico:** variables cantidades o magnitudes.
2. **Categórico:** variables con valores discretos y limitados, incluyendo variables nominales y ordinales.
3. **Fecha:** variables con información temporal, que se agregan según granularidad adecuada (día, semana, mes, año) dependiendo de la cantidad de registros distintos.

De igual manera se implementa un análisis de cardinalidad y discretización que busca una mejor consideración de variables, separando las no relevantes de las que puedan aportar relevancia al dataset, además busca 3 objetivos principales:

- Identificar valores únicos y su frecuencia relativa para detectar atributos poco representativos o con alta dispersión.
- Agrupar categorías menores en un solo grupo denominado “OTROS”, evitando un exceso de granularidad que dificulte la interpretación.
- Dividir variables numéricas en intervalos (bins) para generar histogramas resumidos, respetando la regla de Sturges modificada para no superar un número máximo de intervalos.

La Tabla 7 contiene las actividades realizadas, además del contenido de cada una, indica las medidas descriptivas de cada variable hasta la información del tratamiento de los resultados generados por las actividades de cardinalidad y tratamiento de valores nulos.

Tabla 7
Microactividades del Enriquecimiento Estadístico

Actividad/Artefacto	Contenido
resumen_estadístico	Contiene para cada atributo medidas descriptivas, valores nulos, cantidad de outliers y distribución general.
cardinalidad_columnas	Resume la distribución de categorías, bins de variables numéricas y agregación de valores minoritarios.
resumen_con_modelo	Combina la metadata estadística con el Modelo Entidad-Relación (MER), permitiendo vincular cada atributo a su entidad correspondiente y documentar relaciones entre entidades.
Exportación a formatos Parquet y BSON	Facilita la interoperabilidad con otras plataformas y bases de datos.
Generación de reportes HTML exploratorios	Haciendo uso de ydata-profiling, permite visualizar gráficamente distribuciones, correlaciones y detección de patrones o inconsistencias de manera interactiva.

Las pruebas estadísticas de supuestos, son el complemento a este enriquecimiento, aquí se ejecutan pruebas inferenciales que evalúan supuestos fundamentales para análisis posteriores, desde la normalidad hasta la independencia se detallan a continuación:

1. **Normalidad:** Se aplican pruebas de Shapiro Wilk para datasets pequeños y D´Agostino-Pearson para datasets grandes, evaluando si las distribuciones de variables numéricas siguen un patrón normal. Eso es esencial para seleccionar correctamente tests paramétricos en análisis avanzados.
2. **Homogeneidad de Varianzas:** Mediante la prueba de Levene, se comparan las varianzas de pares de variables numéricas. La homogeneidad garantiza la validez de modelos ANOVA y otras técnicas paramétricas.

3. **Independencia:** Se analiza la relación entre variables categóricas usando Chi-cuadrado. Esta evaluación permite identificar asociaciones significativas o dependencias que podrían afectar interpretaciones estadísticas.
4. **ANOVA (Análisis de Varianza):** Para cada variable numérica frente a una categórica, se comparan las medias de los grupos. Se genera un boxplot de apoyo cuando las diferencias no son significativas, para visualizar la distribución de los datos.

Esta fase asegura que todas las decisiones posteriores en el análisis de datos estén estrictamente fundamentadas en supuestos estadísticos validados, evitando así, interpretaciones incorrectas o sesgadas.

3.8.4 Fase 4: Almacenamiento en Base de Datos NoSQL

La cuarta fase se centra en garantizar la persistencia, disponibilidad y estructuración de la información, permitiendo su reutilización en análisis futuros o integraciones con otras plataformas. Se selecciona MongoDB por su capacidad de almacenar documentos en formato JSON, su flexibilidad para manejar estructuras de datos dinámicas y su compatibilidad con consultas analíticas avanzadas. En la Tabla 8, mostrada a continuación, se muestran las tareas realizadas en esta fase a mayor detalle.

Tabla 8
Microactividades de Almacenamiento

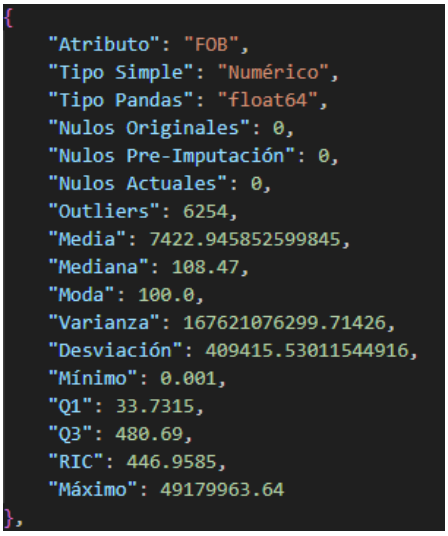
Microactividad	Resultado
Configuración de conexión	Se estableció la conexión entre Python y MongoDB mediante la librería pymongo, asegurando autenticación segura y acceso a la base de datos correspondiente.
Creación de Colecciones	Se definieron dos colecciones principales: **metadata:** contiene descripción completa de los atributos del dataset, incluyendo tipso de datos, valores nulos, cardinalidad, outliers y medidas estadísticas. **dataset_enriquecido:** almacena los registros procesados y enriquecidos, listos para consultas analíticas o integración con herramientas de visualización.
Inserción de documentos	Cada artefacto generado en la fase anterior se transformó en documentos JSON, asegurando que la información fuera accesible y vinculada a su respectiva entidad dentro del Modelo Entidad-Relación.
Validación de Integridad	Se ejecutaron pruebas de inserción y consulta para garantizar que los documentos se almacenaran correctamente, que los campos fueran consistentes y que la estructura jerárquica reflejara fielmente las relaciones definidas en el MER.

Esta fase asegura que toda la información se encuentra centralizada, segura y lista para ser consultada por procesos analíticos, generando un entorno escalable que permite actualizar la metadata o ejecutar análisis adicionales sin comprometer la calidad de los registros.

En esta etapa, la colección DatasetColumn se considera el núcleo funcional del sistema, debido a que representa la aplicación directa del modelo. Cada documento JSON dentro de esta colección describe los atributos de un conjunto de datos incorporando su nombre, tipo de dato, conteo de valores nulos, cardinalidad y medidas estadísticas. Así, DatasetColumn materializa la estructura analítica del sistema, convirtiéndose en el punto de acceso principal para la interpretación de resultados y la vinculación con los módulos de visualización.

La traducción del ER hacia una estructura documental en formato JSON permite visualizar como las relaciones lógicas se transforman en jerarquías anidadas de datos. En esta representación las entidades del modelo ER se convierten en colecciones, y sus atributos en campos dentro de cada documento. Por ejemplo, la entidad DatasetColumn agrupa internamente las estadísticas calculadas y la metadata de cada atributo, reflejando el paso de un modelo relacional a uno documental. La Figura 5 muestra las medidas de tendencia central y dispersión generadas automáticamente para la columna “FOB”, almacenadas en MongoDB como parte de la persistencia de metadatos.

Figura 5
Traducción del Modelo ER hacia JSON



El resto de colecciones cumplen funciones de soporte dentro del modelo general garantizando la coherencia entre la metadata, las operaciones estadísticas y los resultados gráficos. De esta manera, se logra un entorno escalable y modular, en el que cada componente del modelo mantiene independencia funcional, pero continúa vinculado mediante relaciones lógicas definidas en el ER, asegurando consistencia entre la representación conceptual y la implementación documental en MongoDB.

3.8.5 Fase 5: Exportación y Distribución de Resultados

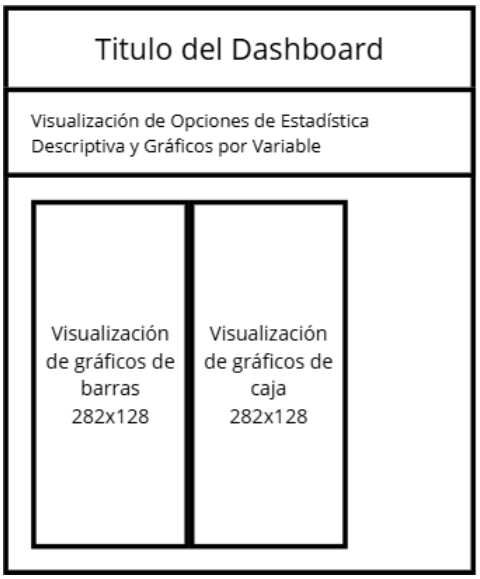
En esta fase se desarrollan mecanismos de salida que permiten transformar resultados del procesamiento en insumos útiles para el análisis y la toma de decisiones. Por un lado, se habilita la exportación de archivos en diferentes formatos –como CSV, Excel o JSON— con el objetivo de garantizar la reutilización de la información en otros entornos y herramientas estadísticas. De manera complementaria, se construyen visualizaciones interactivas que convierten los datos en representaciones gráficas dinámicas, estas visualizaciones se diseñan de acuerdo a la naturaleza de las variables.

El valor de esta fase reside en integrar tanto la exportación de resultados como su representación visual en un solo entorno. De esta manera, los usuarios no solo obtienen archivos listos para análisis externos, sino que también acceden a representaciones gráficas que favorecen la interpretación inmediata de los datos, potencian la identificación de patrones y facilitan la comunicación de hallazgos a distintos públicos.

En las figuras 6 y 7 se muestran mockups de los dashboards de presentación de resultados, mismos que poseen las dimensiones iniciales para cada sección, tanto de estadística descriptiva como de gráficos interactivos para cada una de las variables, en la Figura 6 se encuentra la versión de inicio de la herramienta mostrando las medidas de tendencia central y dispersión de cada variable, mientras que en la Figura 7, se muestran los gráficos de barras y gráficos de cajón, además al final de esta página se encuentra el mapa de correlaciones entre variables.

Figura 6
Mockup I de visualización de resultados

Figura 7
Mockup II de visualización de gráficos



4. Resultados

Los resultados obtenidos reflejan la eficiencia del sistema de enriquecimiento estadístico aplicado al conjunto de datos de comercio exterior del Azuay (2022), compuesto originalmente por 37,484 registros y 77 variables. Tras la fase de validación y limpieza, se eliminaron seis columnas vacías y ningún registro duplicado, quedando 71 variables válidas (100% reducción de vacíos) que constituyen la base del análisis posterior, además no se obtuvo valores duplicados.

Durante el proceso de imputación, se identificaron un total de 500,466 valores nulos repartidos entre variables, pasando a 275,568 después de la eliminación de columnas vacías, imputándose según el tipo de variable al que pertenecieran, es importante recalcar que se han imputado los valores tanto numéricos como categóricos según el criterio del desarrollador, los valores tipo fecha se remplazan con una fecha base (1/1/1900), a pesar de ser columnas que estén vacías. En la Tabla 9 se observa con mayor detalle los resultados obtenidos de todo el proyecto.

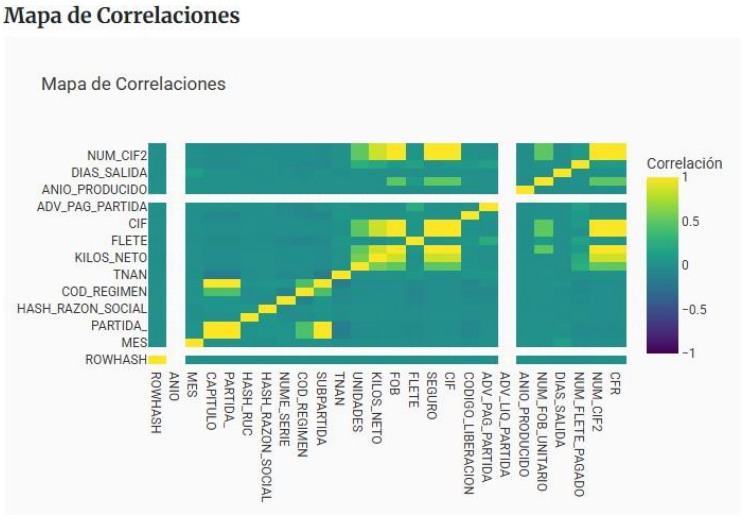
Tabla 9
Resultados de Procesamiento y Limpieza

Proceso	Variables Afectadas	Valores Procesados
Limpieza	6 columnas eliminadas	77 -> 71
Limpieza	71 variables restantes	50,0466 -> 27,5568 valores nulos
Imputación	71 variables restantes	27,5568 -> 0, todos los valores se imputaron
Enriquecimiento	71 variables	37,484 valores procesados para generación de metadata.
Exportación	71 variables	5 colecciones JSON, con cada uno de los procesamientos planteados.

De forma complementaria, se generaron los mecanismos de visualización automática, entre ellos: histogramas, gráficos de caja y mapas de calor, los cuales representan la distribución y correlación entre variables, dichas visualizaciones permiten al usuario interpretar con facilidad la estructura y comportamiento de los datos sin necesidad de

recurrir a herramientas externas. Un ejemplo se visualiza en la Figura 8, que representa mediante un mapa de correlaciones, el grado de relación entre las variables del dataset.

Figura 8
Mapa de Correlaciones de Comercio Exterior

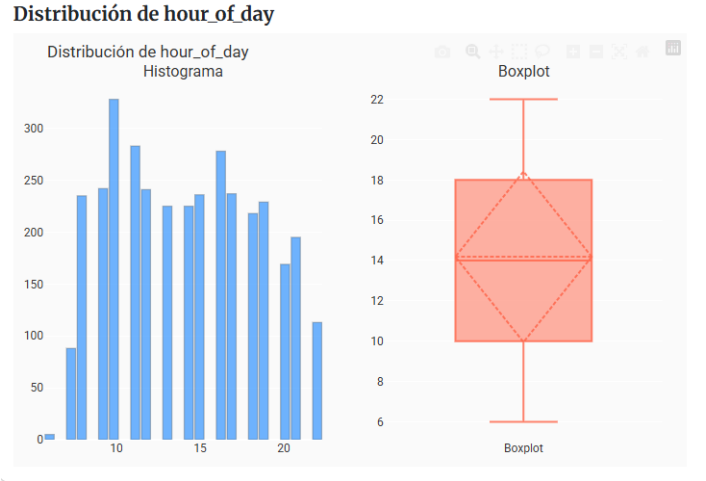


Se ejecutaron pruebas con varios datasets adicionales para validar la generalidad y robustez del sistema. Un conjunto de prueba (3,500 registros en formato CSV) y un dataset de mayor tamaño (37,484 registros). En ambos casos, los resultados confirman que la aplicación ejecuta correctamente todas las fases del proceso –limpieza, imputación, enriquecimiento y exportación de metadatos–, adaptándose al tamaño y formato sin requerir modificaciones en el código base. Observar la Figura 9 y Figura 10 para comprobar la funcionalidad de la herramienta, en otro dataset (relacionado a la venta de café, y en el cual se han eliminado datos de la variable money a propósito para verificar la funcionalidad de identificar valores nulos):

Figura 9
Tabla de Estadística Descriptiva por Variable

	Variable	Cantidad	Únicos	Moda	Frecuencia	Media	Desviación	Mínimo	Q1	Mediana	Q3	Máximo	missing
0	hour_of_day	3547.0	NaN	NaN	NaN	14.185791	4.23401	6.0	10.0	14.0	18.0	22.0	0
1	cash_type	3547	1	card	3547	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0
2	coffee_name	3547	8	Americano with Milk	809	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0
3	Time_of_Day	3547	3	Afternoon	1205	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0
4	Weekday	3547	7	Tue	572	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0
5	Month_name	3547	12	Mar	494	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0
6	Weekdaysort	3547.0	NaN	NaN	NaN	3.845785	1.971501	1.0	2.0	4.0	6.0	7.0	0
7	Monthsort	3547.0	NaN	NaN	NaN	6.453905	3.500754	1.0	3.0	7.0	10.0	12.0	0
8	Date	3547	381	11/10/2024	26	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0
9	Time	3547	3373	08:18.8	3	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0

Figura 10
Gráficos de la variable hour_of_day



En conjunto, los resultados muestran que la aplicación automatiza tareas clave del análisis de datos y amplía el alcance de la técnica de enriquecimiento desde una perspectiva descriptiva y no exclusivamente algorítmica. La integración de los reportes estadísticos, la generación de dashboards y la creación sistemática de metadatos, permiten escalar su uso hacia futuras líneas de investigación basadas en modelos de lenguaje enriquecidos por estadística, donde la estructura y la calidad del dataset constituyen insumos fundamentales.

5. Discusión

Los resultados obtenidos en el desarrollo del sistema de enriquecimiento estadístico de datos demuestran que es posible integrar técnicas de preprocesamiento, imputación y generación de metadatos dentro de una misma plataforma orientada al análisis de información estructurada. Este enfoque coincide con las tendencias reportadas por Sánchez et al. (2023) y Ghosh et al. (2022), quienes destacan la necesidad de automatizar los procesos de integración y enriquecimiento para reducir la intervención manual y mejorar la calidad de los datos.

Sin embargo, a diferencia de las plataformas diseñadas para entornos distribuidos o flujos de datos en tiempo real, como las propuestas de Scheinert et al. (2023) o Wang & Carey (2018). El sistema desarrollado en esta investigación se centra en contextos locales de análisis, priorizando la consistencia, la accesibilidad y la interpretación visual de los resultados. De igual modo, mientras Zhou et al. (2023) orientan la visualización hacia fines predictivos, el sistema desarrollado adopta un enfoque exploratorio y descriptivo, donde el objetivo es facilitar la comprensión estadística de los datos antes que la optimización de modelos.

En relación con Ghosh et al. (2022), su propuesta de enriquecimiento se orienta a la generación y gestión de metadatos técnicos, donde el valor añadido proviene de describir la estructura y procedencia de los datos para optimizar su interoperabilidad en sistemas distribuidos. En contraste, en este estudio los metadatos se reinterpretan desde una función analítica, ya que no solo documentan el origen o formato, sino que actúan como indicadores de calidad, dispersión y variabilidad estadística de cada atributo. Esta diferencia transforma los metadatos de un componente descriptivo a un instrumento activo de evaluación del dataset.

Estas diferencias permiten reconocer aportes específicos de este estudio. En primer lugar, el sistema demuestra que el enriquecimiento de datos no requiere necesariamente infraestructuras distribuidas para ser efectivo, ya que puede implementarse en entornos locales manteniendo estándares de calidad y eficiencia. En segundo lugar, el estudio propone una reinterpretación metodológica del enriquecimiento, orientándolo hacia la generación de conocimiento estadístico más que hacia la mera integración técnica. En tercer lugar, introduce un componente educativo y exploratorio como herramientas de

interpretación analítica, potenciado la comprensión del comportamiento de las variables y sus relaciones. Si bien el caso de estudio es el comercio exterior de la provincia del Azuay en el año 2022, el sistema desarrollado se ejecuta sobre cualquier tipo de dataset que se considere, descartando así, archivos corrompidos o dañados.

En síntesis, el presente estudio coincide con la literatura en su propósito de mejorar la calidad y automatización del enriquecimiento de datos; se diferencia por su enfoque descriptivo, local y analítico; y aporta una visión complementaria que amplía el campo de aplicación del enriquecimiento hacia contextos donde la interpretabilidad y la generación de conocimiento son prioritarias frente al volumen o la complejidad tecnológica.

Más allá del ámbito estadístico, el sistema presenta un potencial de integración con modelos de lenguaje de gran escala (LLMs). Los datos enriquecidos, al incorporar contexto numérico y metadatos estructurados, pueden ser utilizados por estos modelos a través de técnicas como Retrieval-Augmented Generation (RAG) o mediante la creación de embeddings híbridos, en los que se combinan representaciones textuales y estadísticas. Esto permite que los LLMs –intrínsecamente cualitativos– incorporen información cuantitativa dentro de sus procesos de generación y razonamiento, fortaleciendo la precisión de sus respuestas y la comprensión semántica de los datos.

En ese sentido, la herramienta desarrollada no solo aporta a la gestión automatizada de datos estructurados, sino que sienta las bases para la convergencia entre analítica descriptiva y los modelos de inteligencia artificial. Este tipo de sinergia proyecta nuevas posibilidades para el aprovechamiento del conocimiento estadístico en tareas de minería de información, análisis predictivo y generación de conocimiento contextualizado.

6. Conclusión

Se desarrolló una herramienta capaz de enriquecer estadísticamente un conjunto de datos estructurado, incorporando gráficos de distribución y medidas de tendencia central. A través del caso de estudio del comercio exterior del Azuay (2022), se comprobó la eficacia del sistema para automatizar el proceso completo de análisis descriptivo, desde la carga del dataset hasta la generación de resultados visuales interpretables. Se examinaron las principales técnicas de enriquecimiento de datos aplicadas al contexto del dataset. Este análisis diferenció las variables numéricas, categóricas y temporales, y definir los criterios más adecuados para su tratamiento según la naturaleza de los valores y la estructura de la información.

También, se estableció un modelo de clasificación automática de variables que integra reglas de detección de patrones, relaciones entre columnas y niveles de cardinalidad. Este modelo, estandarizó el proceso de reconocimiento de tipos de datos, garantizando una interpretación coherente y adaptable a diferentes conjuntos tabulares.

Asimismo, se construyó un repositorio con capacidad para almacenar los resultados del enriquecimiento, incluyendo metadatos, medidas estadísticas y detección de valores atípicos. La integración con MongoDB aseguró la persistencia de la información y la posibilidad de consultas analíticas futuras, consolidando un entorno escalable y reutilizable.

Finalmente, se implementaron mecanismos de visualización que exploraron de forma dinámica las relaciones entre variables mediante gráficos estadísticos. Estas visualizaciones facilitan la interpretación de los resultados y fortalecen la comprensión del comportamiento del conjunto de datos sin necesidad de herramientas externas.

En síntesis, la aplicación desarrollada no solo responde a los objetivos planteados, sino que también amplía el alcance del análisis estadístico hacia entornos más accesibles y flexibles. Su diseño modular, abierto y basado en JSON, posibilita su integración con sistemas inteligentes o modelos de lenguaje, contribuyendo a la democratización del análisis de datos y al aprovechamiento del conocimiento estadístico como base para nuevas líneas de investigación.

7. Referencias

Alassaf, M., & Qamar, A. M. (2022). Improving Sentiment Analysis of Arabic Tweets by One-way ANOVA. *Journal of King Saud University - Computer and Information Sciences*, *34*(6), 2849–2859. https://doi.org/10.1016/j.jksuci.2020.10.023

Azad, S. A., Wasimi, S., & Ali, A. B. M. S. (2018). Business Data Enrichment: Issues and Challenges. *Proceedings - 2018 5th Asia-Pacific World Congress on Computer Science and Engineering, APWC on CSE 2018*, 98–102. https://doi.org/10.1109/APWConCSE.2018.00024

Bland, M. (2015). Estimating Mean and Standard Deviation from the Sample Size, Three Quartiles, Minimum, and Maximum. In *International Journal of Statistics in Medical Research* (Vol. 4).

Cutrona, V. (2019). *Semantic Enrichment for Large-Scale Data Analytics*. https://agate.readthedocs.org/en/1.3.1

Darling, H. S. (2022). Do you have a standard way of interpreting the standard deviation? A narrative review. *Cancer Research, Statistics, and Treatment*, *5*(4), 728–733. https://doi.org/10.4103/crst.crst_284_22

Dong, Y., & Oyamada, M. (2022). Table Enrichment System for Machine Learning. *SIGIR 2022 - Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 3267–3271. https://doi.org/10.1145/3477495.3531678

Ghosh, D., Gupta, P., Mehrotra, S., & Sharma, S. (2022). *A Case for Enrichment in Data Management Systems*.

Ghosh, D., Gupta, P., Mehrotra, S., Yus, R., & Altowim, Y. (2022). JENNER: Just-in-time Enrichment in Query Processing. *Proceedings of the VLDB Endowment*, *15*(11), 2666–2678. https://doi.org/10.14778/3551793.3551822

Gong, C., Zheng, Z., Wu, F., Jia, X., & Chen, G. (2024). Delta: A Cloud-assisted Data Enrichment Framework for On-Device Continual Learning. *ACM MobiCom 2024 - Proceedings of the 30th International Conference on Mobile Computing and Networking*, 1408–1423. https://doi.org/10.1145/3636534.3690701

González-Estrada, E., Villaseñor, J. A., & Acosta-Pech, R. (2022). Shapiro-Wilk test for multivariate skew-normality. *Computational Statistics*, *37*(4), 1985–2001. https://doi.org/10.1007/s00180-021-01188-y

Gorschek, T., & Larsson, S. B. (2006). *A Model for Technology Transfer in Practice Per Garre i.*

Groth, R. E., & Bergner, J. A. (2006). Preservice Elementary Teachers' Conceptual and Procedural Knowledge of Mean, Median, and Mode. *Mathematical Thinking and Learning*, *8*(1), 37–63. https://doi.org/10.1207/s15327833mtl0801_3

Herzog, T. N., Scheuren, F. J., & Winkler, W. E. (2007). Data quality and record linkage techniques. In *Data Quality and Record Linkage Techniques*. Springer New York. https://doi.org/10.1007/0-387-69505-2

Hristov, E., Petrova-Antonova, D., De Paoli, F., Krasteva, I., Ciavotta, M., & Avogadro, R. (2024). Geospatial Data Enrichment through Address Geocoding: Challenges and Solutions. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, *48*(4), 239–245. https://doi.org/10.5194/isprs-archives-XLVIII-4-2024-239-2024

Jankulovski, N., Bojkovska, K., & Grozdanovska, V. (2017). International Business and Trade. *International Journal of Sciences: Basic and Applied Research*, *31*(3), 105–114. http://gssrr.org/index.php?journal=JournalOfBasicAndApplied

Jegierski, H., & Saganowski, S. (2020). An "Outside the Box" Solution for Imbalanced Data Classification. *IEEE Access*, *8*, 125191–125209. https://doi.org/10.1109/ACCESS.2020.3007801

Korkmaz, S., & Demir, Y. (2023). *INVESTIGATION OF SOME UNIVARIATE NORMALITY TESTS IN TERMS OF TYPE-I ERRORS AND TEST POWER*.

Kosikov, S., Ismailova, L., & Wolfengagen, V. (2021). Data Enrichment in the Information Graphs Environment Based on a Specialized Architecture of Information Channels. *Procedia Computer Science*, *190*, 492–499. https://doi.org/10.1016/j.procs.2021.07.001

Ley Organica de Aduanas. (2006). *LEY ORGANICA DE ADUANAS, CODIFICACION*.

Maharana, K., Mondal, S., & Nemade, B. (2022). A review: Data pre-processing and data augmentation techniques. *Global Transitions Proceedings*, *3*(1), 91–99. https://doi.org/10.1016/j.gltp.2022.04.020

Malakar, I. M. (2023). Conceptualizing Central Tendency and Dispersion in Applied Statistics. *Cognition*, *5*(1), 50–62. https://doi.org/10.3126/cognition.v5i1.55408

Malyavko, A. A., Reutov, V. V., Korotkikh, I. V., & Shperling, V. K. (2022). Algorithms, methods and approaches to details and enrich data, including personal data. *Digital Technology Security*, *4*, 9–26. https://doi.org/10.17212/2782-2230-2022-4-9-26

Mariani, P., Marletta, A., & Locci, M. (2024). Missing values and data enrichment: an application to social media liking. *Computational Statistics*, *39*(1), 217–237. https://doi.org/10.1007/s00180-022-01261-0

Mumuni, A., & Mumuni, F. (2024). Automated data processing and feature engineering for deep learning and big data applications: A survey. *Journal of Information and Intelligence*. https://doi.org/10.1016/j.jiixd.2024.01.002

Nihan, S. T. (2020). Karl Pearsons chi-square tests. *Educational Research and Reviews*, *15*(9), 575–580. https://doi.org/10.5897/err2019.3817

Park, Y., Choi, J., & Choi, J. (2021). An Extensible Data Enrichment Scheme for Providing Intelligent Services in Internet of Things Environments. *Mobile Information Systems*, *2021*. https://doi.org/10.1155/2021/5535231

Pyle, D., Cerra, D. D., & Kaufmann, M. (1999). *Data Preparation for Data Mining*.

Rettore, P. H. L., Santos, B. P., Lopes, R. R. F., Maia, G., Villas, L. A., & Loureiro, A. A. F. (2020). Road Data Enrichment Framework Based on Heterogeneous Data Fusion for ITS. *IEEE Transactions on Intelligent Transportation Systems*, *21*(4), 1751–1766. https://doi.org/10.1109/TITS.2020.2971111

Ritze, D., & Eckert, K. (2014). *Data Enrichment in Discovery Systems Using Linked Data*. https://doi.org/10.1007/978-3-319-01595-8_49

Sanchez, L., Lanza, J., Santana, J. R., Sotres, P., Gonzalez, V., Martin, L., Solmaz, G., Kovacs, E., Dietzel, M., Summa, A., Jafari, A. R., Minerva, R., & Crespi, N. (2023). Data Enrichment Toolchain: A Data Linking and Enrichment Platform for Heterogeneous Data. *IEEE Access*, *11*, 103079–103091. https://doi.org/10.1109/ACCESS.2023.3317705

Scheinert, D., Casares, F., Geldenhuys, M. K., Styp-Rekowski, K., & Kao, O. (2023). Evaluation of Data Enrichment Methods for Distributed Stream Processing Systems. *Proceedings - 2023 IEEE International Conference on Cloud Engineering, IC2E 2023*, 202–211. https://doi.org/10.1109/IC2E59103.2023.00030

SENAE. (2015). *SERVICIO NACIONAL DE ADUANA DEL ECUADOR*.

Shirley Martillo-Alchundia, I. I., & Alvarado-Zabala, J. I. (2025). *Fortalecimiento de la Gestión de Datos a través de Big Data Analytics para la Transformación y Mejora de los Servicios en Instituciones Financieras Strengthening Data Management through Big Data Analytics for the Transformation and Improvement of Services in Financial Institutions Fortalecendo a Gestão de Dados por meio de Big Data Analytics para a Transformação e Melhoria dos Serviços nas Instituições Financeiras*. *102*, 2042–2059. https://doi.org/10.23857/pc.v10i1.8796

Shumba, W. (2024). *The International Convention on the Harmonized Commodity Description and Coding System: Legal Pillar Behind the Harmonized System*. https://www.wto.org/english/thewto_e/thewto_e.htm

Shyalika, C., Wickramarachchi, R., El Kalach, F., Harik, R., & Sheth, A. (2024). Evaluating the Role of Data Enrichment Approaches towards Rare Event Analysis in Manufacturing. *Sensors*, *24*(15). https://doi.org/10.3390/s24155009

Singh, M. M., Singaravel, S., & Geyer, P. (2021). Machine learning for early stage building energy prediction: Increment and enrichment. *Applied Energy*, *304*. https://doi.org/10.1016/j.apenergy.2021.117787

Soni, P. (2014). *The Rights and Duties of the Transacting Parties under FOB International Sales Contract*. http://ssrn.com/abstract=2423707

Sritan, K., & Phuenaree, B. (2021). A Comparison of Efficiency for Homogeneity of Variance Tests under Log-normal Distribution. In *Asian Journal of Applied Sciences* (Vol. 9, Issue 4). www.ajouronline.com

Vetter, T. R. (2017). Descriptive Statistics: Reporting the Answers to the 5 Basic Questions of Who, What, Why, When, Where, and a Sixth, so What? *Anesthesia and Analgesia*, *125*(5), 1797–1802. https://doi.org/10.1213/ANE.0000000000002471

Wang, X., & Carey, M. J. (2018). An IDEA: An ingestion framework for data enrichment in AsterixDB. *Proceedings of the VLDB Endowment*, *12*(11), 1485–198. https://doi.org/10.14778/3342263.3342628

Zhou, Y., Ma, L., Ni, W., & Yu, C. (2023). Data Enrichment as a Method of Data Preprocessing to Enhance Short-Term Wind Power Forecasting. *Energies*, *16*(5). https://doi.org/10.3390/en16052094

8. Anexos

Anexo 1
Tablas de Variables y Medidas

Atributo	ROWHASH
Tipo de Dato	Numérico
Tipo de Dato	uint64
Valores Nulos	0
Valores Atípicos	0
Media	9.214266391890612e+18
Mediana	9.178508454138885e+18
Moda	106085383637909
Rango Medio	9.223026079824881e+18
Varianza	2.8549120741894836e+37
Desviación Estándar	5.343137724398917e+18
Rango	18445839988882486064
Cuartiles	Q1: 4.573065662247307e+18, Q3: 1.3879046239029807e+19
Rango Intercuartil	9.3059805767825e+18

Atributo	TIPO
Tipo de Dato	Nominal no categórico
Tipo de Dato	object
Valores Nulos	0
Valores Atípicos	No aplica
Media	No aplica
Mediana	No aplica
Moda	IMP
Rango Medio	No aplica
Varianza	No aplica
Desviación Estándar	No aplica
Rango	No aplica
Cuartiles	No aplica
Rango Intercuartil	No aplica

Atributo	CAPITULO
Tipo de Dato	Numérico
Tipo de Dato	int64
Valores Nulos	0
Valores Atípicos	1011
Media	74.441533
Mediana	84.0
Moda	98
Rango Medio	49.5
Varianza	628.280044
Desviación Estándar	25.065515
Rango	97
Cuartiles	Q1: 62.0, Q3: 98.0
Rango Intercuartil	36.0

Atributo	DESCRIPCION_CAPITULO
Tipo de Dato	Nominal no categórico
Tipo de Dato	object
Valores Nulos	0
Valores Atípicos	No aplica
Media	No aplica
Mediana	No aplica
Moda	MERCANCIAS CON TRATAMIENTO ESPECIAL
Rango Medio	No aplica
Varianza	No aplica
Desviación Estándar	No aplica
Rango	No aplica
Cuartiles	No aplica
Rango Intercuartil	No aplica

Atributo	DESCRIPCION_PARTIDA
Tipo de Dato	Nominal no categórico
Tipo de Dato	object
Valores Nulos	3548
Valores Atípicos	No aplica
Media	No aplica
Mediana	No aplica
Moda	Tráfico Postal Internacional y Correos Rápidos
Rango Medio	No aplica
Varianza	No aplica
Desviación Estándar	No aplica
Rango	No aplica
Cuartiles	No aplica
Rango Intercuartil	No aplica

Atributo	RUC
Tipo de Dato	Nominal no categórico
Tipo de Dato	object
Valores Nulos	0
Valores Atípicos	No aplica
Media	No aplica
Mediana	No aplica
Moda	1791359240001
Rango Medio	No aplica
Varianza	No aplica
Desviación Estándar	No aplica
Rango	No aplica
Cuartiles	No aplica
Rango Intercuartil	No aplica

Atributo	HASH_RUC
Tipo de Dato	Numérico
Tipo de Dato	int64
Valores Nulos	0
Valores Atípicos	0
Media	2107349020.9404
Mediana	1973645851.0
Moda	422539839
Rango Medio	2147178675.0
Varianza	1.4647965245407224e+18
Desviación Estándar	1210287785.834726
Rango	4294220428
Cuartiles	Q1: 1073854119.0, Q3: 3130164385.0
Rango Intercuartil	2056310266.0

Atributo	HASH_RAZON_SOCIAL
Tipo de Dato	Numérico
Tipo de Dato	int64
Valores Nulos	0
Valores Atípicos	0

Media	2260263021.848678
Mediana	2305936190.0
Moda	1766245315
Rango Medio	2147773054.5
Varianza	1.5384766361094072e+18
Desviación Estándar	1240353431.933579
Rango	4294373741
Cuartiles	Q1: 1226000818.5, Q3: 3345048306.0
Rango Intercuartil	2119047487.5

Atributo	REMITENTE
Tipo de Dato	Nominal no categórico
Tipo de Dato	object
Valores Nulos	1872
Valores Atípicos	No aplica
Media	No aplica
Mediana	No aplica
Moda	TRANS EXPRESS INC
Rango Medio	No aplica
Varianza	No aplica
Desviación Estándar	No aplica
Rango	No aplica
Cuartiles	No aplica
Rango Intercuartil	No aplica

Atributo	NOTIFY
Tipo de Dato	Nominal no categórico
Tipo de Dato	object
Valores Nulos	25
Valores Atípicos	No aplica
Media	No aplica
Mediana	No aplica
Moda	SIN INTERMEDIARIO---
Rango Medio	No aplica
Varianza	No aplica
Desviación Estándar	No aplica
Rango	No aplica
Cuartiles	No aplica
Rango Intercuartil	No aplica

Atributo	EMBARCADOR_CONSIGNEE
Tipo de Dato	Nominal no categórico
Tipo de Dato	object
Valores Nulos	616
Valores Atípicos	No aplica
Media	No aplica
Mediana	No aplica
Moda	EMPTY FIELD
Rango Medio	No aplica
Varianza	No aplica
Desviación Estándar	No aplica
Rango	No aplica
Cuartiles	No aplica
Rango Intercuartil	No aplica

Atributo	EMBARCADOR_CONSIGNEE_ADDRESS
Tipo de Dato	Nominal no categórico
Tipo de Dato	object

Valores Nulos	25
Valores Atípicos	No aplica
Media	No aplica
Mediana	No aplica
Moda	EMPTY FIELD
Rango Medio	No aplica
Varianza	No aplica
Desviación Estándar	No aplica
Rango	No aplica
Cuartiles	No aplica
Rango Intercuartil	No aplica

Atributo	REFRENDO
Tipo de Dato	Nominal no categórico
Tipo de Dato	object
Valores Nulos	0
Valores Atípicos	No aplica
Media	No aplica
Mediana	No aplica
Moda	028-2022-10-01057580
Rango Medio	No aplica
Varianza	No aplica
Desviación Estándar	No aplica
Rango	No aplica
Cuartiles	No aplica
Rango Intercuartil	No aplica

Atributo	NUME_SERIE
Tipo de Dato	Numérico
Tipo de Dato	int64
Valores Nulos	0
Valores Atípicos	3905
Media	168.351679
Mediana	50.0
Moda	1
Rango Medio	2865.0
Varianza	110400.102133
Desviación Estándar	332.265108
Rango	5728
Cuartiles	Q1: 8.0, Q3: 177.0
Rango Intercuartil	169.0

Atributo	TIPO_AFORO
Tipo de Dato	Nominal no categórico
Tipo de Dato	object
Valores Nulos	7070
Valores Atípicos	No aplica
Media	No aplica
Mediana	No aplica
Moda	A
Rango Medio	No aplica
Varianza	No aplica
Desviación Estándar	No aplica
Rango	No aplica
Cuartiles	No aplica
Rango Intercuartil	No aplica

Atributo	COD_REGIMEN
Tipo de Dato	Numérico
Tipo de Dato	int64

Valores Nulos	0
Valores Atípicos	0
Media	36.931329
Mediana	10.0
Moda	10
Rango Medio	53.0
Varianza	1338.27147
Desviación Estándar	36.582393
Rango	86
Cuartiles	Q1: 10.0, Q3: 91.0
Rango Intercuartil	81.0

Atributo	REGIMEN
Tipo de Dato	Nominal no categórico
Tipo de Dato	object
Valores Nulos	0
Valores Atípicos	No aplica
Media	No aplica
Mediana	No aplica
Moda	IMPORTACION A CONSUMO
Rango Medio	No aplica
Varianza	No aplica
Desviación Estándar	No aplica
Rango	No aplica
Cuartiles	No aplica
Rango Intercuartil	No aplica

Atributo	DISTRITO
Tipo de Dato	Nominal no categórico
Tipo de Dato	object
Valores Nulos	0
Valores Atípicos	No aplica
Media	No aplica
Mediana	No aplica
Moda	GUAYAQUIL - MARITIMO
Rango Medio	No aplica
Varianza	No aplica
Desviación Estándar	No aplica
Rango	No aplica
Cuartiles	No aplica
Rango Intercuartil	No aplica

Atributo	AGENTE_AFIANZADO
Tipo de Dato	Nominal no categórico
Tipo de Dato	object
Valores Nulos	0
Valores Atípicos	No aplica
Media	No aplica
Mediana	No aplica
Moda	TORRES & TORRES AGENTES DE ADUANAS TTADAD C.A.
Rango Medio	No aplica
Varianza	No aplica
Desviación Estándar	No aplica
Rango	No aplica
Cuartiles	No aplica
Rango Intercuartil	No aplica

Atributo	AGENCIA
Tipo de Dato	Nominal no categórico
Tipo de Dato	object
Valores Nulos	31374
Valores Atípicos	No aplica
Media	No aplica
Mediana	No aplica
Moda	MAERSK DEL ECUADOR C.A.
Rango Medio	No aplica
Varianza	No aplica
Desviación Estándar	No aplica
Rango	No aplica
Cuartiles	No aplica
Rango Intercuartil	No aplica

Atributo	LINEA
Tipo de Dato	Nominal no categórico
Tipo de Dato	object
Valores Nulos	1903
Valores Atípicos	No aplica
Media	No aplica
Mediana	No aplica
Moda	UNITED PARCEL SERVICE, CO.
Rango Medio	No aplica
Varianza	No aplica
Desviación Estándar	No aplica
Rango	No aplica
Cuartiles	No aplica
Rango Intercuartil	No aplica

Atributo	MANIFIESTO
Tipo de Dato	Nominal no categórico
Tipo de Dato	object
Valores Nulos	1872
Valores Atípicos	No aplica
Media	No aplica
Mediana	No aplica
Moda	1
Rango Medio	No aplica
Varianza	No aplica
Desviación Estándar	No aplica
Rango	No aplica
Cuartiles	No aplica
Rango Intercuartil	No aplica

Atributo	MANIFIESTO_ADUANA
Tipo de Dato	Nominal no categórico
Tipo de Dato	object
Valores Nulos	1568
Valores Atípicos	No aplica
Media	No aplica
Mediana	No aplica
Moda	CEC2022CP00003900010000
Rango Medio	No aplica
Varianza	No aplica
Desviación Estándar	No aplica
Rango	No aplica
Cuartiles	No aplica
Rango Intercuartil	No aplica

Atributo	BL
Tipo de Dato	Nominal no categórico
Tipo de Dato	object
Valores Nulos	1572
Valores Atípicos	No aplica
Media	No aplica
Mediana	No aplica
Moda	CPUIO0000069
Rango Medio	No aplica
Varianza	No aplica
Desviación Estándar	No aplica
Rango	No aplica
Cuartiles	No aplica
Rango Intercuartil	No aplica

Atributo	FECH_EMBAR
Tipo de Dato	Desconocido
Tipo de Dato	datetime64[ns]
Valores Nulos	3366
Valores Atípicos	No aplica
Media	No aplica
Mediana	No aplica
Moda	2022-07-16 00:00:00
Rango Medio	No aplica
Varianza	No aplica
Desviación Estándar	No aplica
Rango	No aplica
Cuartiles	No aplica
Rango Intercuartil	No aplica

Atributo	FECH_LLEGA
Tipo de Dato	Desconocido
Tipo de Dato	datetime64[ns]
Valores Nulos	1569
Valores Atípicos	No aplica
Media	No aplica
Mediana	No aplica
Moda	2022-08-27 00:00:00
Rango Medio	No aplica
Varianza	No aplica
Desviación Estándar	No aplica
Rango	No aplica
Cuartiles	No aplica
Rango Intercuartil	No aplica

Atributo	FECHA_INGRESO
Tipo de Dato	Desconocido
Tipo de Dato	datetime64[ns]
Valores Nulos	0
Valores Atípicos	No aplica
Media	No aplica
Mediana	No aplica
Moda	2022-03-22 00:00:00
Rango Medio	No aplica
Varianza	No aplica
Desviación Estándar	No aplica
Rango	No aplica
Cuartiles	No aplica
Rango Intercuartil	No aplica

Atributo	FECH_PAGO
Tipo de Dato	Desconocido
Tipo de Dato	datetime64[ns]
Valores Nulos	12813
Valores Atípicos	No aplica
Media	No aplica
Mediana	No aplica
Moda	2022-11-09 00:00:00
Rango Medio	No aplica
Varianza	No aplica
Desviación Estándar	No aplica
Rango	No aplica
Cuartiles	No aplica
Rango Intercuartil	No aplica

Atributo	FECH_SALIDA
Tipo de Dato	Desconocido
Tipo de Dato	datetime64[ns]
Valores Nulos	7710
Valores Atípicos	No aplica
Media	No aplica
Mediana	No aplica
Moda	2022-11-16 00:00:00
Rango Medio	No aplica
Varianza	No aplica
Desviación Estándar	No aplica
Rango	No aplica
Cuartiles	No aplica
Rango Intercuartil	No aplica

Atributo	FACTURA
Tipo de Dato	Nominal no categórico
Tipo de Dato	object
Valores Nulos	13145
Valores Atípicos	No aplica
Media	No aplica
Mediana	No aplica
Moda	CFM-2217765
Rango Medio	No aplica
Varianza	No aplica
Desviación Estándar	No aplica
Rango	No aplica
Cuartiles	No aplica
Rango Intercuartil	No aplica

Atributo	NAVE
Tipo de Dato	Nominal no categórico
Tipo de Dato	object
Valores Nulos	3669
Valores Atípicos	No aplica
Media	No aplica
Mediana	No aplica
Moda	HELLE RITSCHER
Rango Medio	No aplica
Varianza	No aplica
Desviación Estándar	No aplica
Rango	No aplica
Cuartiles	No aplica
Rango Intercuartil	No aplica

Atributo	ALMACEN_TEMPORAL
Tipo de Dato	Numérico
Tipo de Dato	float64
Valores Nulos	37483
Valores Atípicos	0
Media	nan
Mediana	nan
Moda	Sin moda
Rango Medio	nan
Varianza	nan
Desviación Estándar	nan
Rango	nan
Cuartiles	Q1: nan, Q3: nan
Rango Intercuartil	nan

Atributo	DEP_COMERCIAL
Tipo de Dato	Nominal no categórico
Tipo de Dato	object
Valores Nulos	20606
Valores Atípicos	No aplica
Media	No aplica
Mediana	No aplica
Moda	INARPI S.A.
Rango Medio	No aplica
Varianza	No aplica
Desviación Estándar	No aplica
Rango	No aplica
Cuartiles	No aplica
Rango Intercuartil	No aplica

Atributo	SUBPARTIDA
Tipo de Dato	Numérico
Tipo de Dato	int64
Valores Nulos	0
Valores Atípicos	1011
Media	7458197053.150042
Mediana	8483309000.0
Moda	9807103000

Rango Medio	4954450500.0
Varianza	6.291236205182306e+18
Desviación Estándar	2508233682.331514
Rango	9706303000
Cuartiles	Q1: 6201160000.0, Q3: 9807102000.0
Rango Intercuartil	3605942000.0

Atributo	TNAN
Tipo de Dato	Numérico
Tipo de Dato	int64
Valores Nulos	0
Valores Atípicos	1884
Media	0.114372
Mediana	0.0
Moda	0
Rango Medio	9.5
Varianza	0.662683
Desviación Estándar	0.814053
Rango	19
Cuartiles	Q1: 0.0, Q3: 0.0
Rango Intercuartil	0.0

Atributo	TASA_ADVALOREM
Tipo de Dato	Numérico
Tipo de Dato	float64
Valores Nulos	37483
Valores Atípicos	0
Media	nan
Mediana	nan
Moda	Sin moda
Rango Medio	nan
Varianza	nan
Desviación Estándar	nan
Rango	nan
Cuartiles	Q1: nan, Q3: nan
Rango Intercuartil	nan

Atributo	DESC_ARAN
Tipo de Dato	Nominal no categórico
Tipo de Dato	object
Valores Nulos	1299
Valores Atípicos	No aplica
Media	No aplica
Mediana	No aplica
Moda	PAQUETES POR CORREOS RÁPIDOS (MENSAJERÍA ACELERADA O COURIER)
Rango Medio	No aplica
Varianza	No aplica
Desviación Estándar	No aplica
Rango	No aplica
Cuartiles	No aplica
Rango Intercuartil	No aplica

Atributo	DESC_COMER
Tipo de Dato	Nominal no categórico
Tipo de Dato	object
Valores Nulos	0
Valores Atípicos	No aplica
Media	No aplica
Mediana	No aplica
Moda	PRENDAS DE VESTIR
Rango Medio	No aplica
Varianza	No aplica
Desviación Estándar	No aplica
Rango	No aplica
Cuartiles	No aplica
Rango Intercuartil	No aplica

Atributo	MARCAS
Tipo de Dato	Nominal no categórico
Tipo de Dato	object
Valores Nulos	12699
Valores Atípicos	No aplica
Media	No aplica
Mediana	No aplica
Moda	SIN MARCA
Rango Medio	No aplica
Varianza	No aplica
Desviación Estándar	No aplica
Rango	No aplica
Cuartiles	No aplica
Rango Intercuartil	No aplica

Atributo	CIUDAD_EMBARQUE
Tipo de Dato	Nominal no categórico
Tipo de Dato	object
Valores Nulos	18437
Valores Atípicos	No aplica
Media	No aplica
Mediana	No aplica
Moda	MIAMI (MIA)-MIAMI INTERNATIONAL AIRPORT
Rango Medio	No aplica
Varianza	No aplica
Desviación Estándar	No aplica
Rango	No aplica
Cuartiles	No aplica
Rango Intercuartil	No aplica

Atributo	TIPO_CARGA
Tipo de Dato	Numérico
Tipo de Dato	float64
Valores Nulos	37483
Valores Atípicos	0
Media	nan
Mediana	nan
Moda	Sin moda
Rango Medio	nan
Varianza	nan
Desviación Estándar	nan
Rango	nan
Cuartiles	Q1: nan, Q3: nan
Rango Intercuartil	nan

Atributo	TIPO_UNIDAD
Tipo de Dato	Nominal no categórico
Tipo de Dato	object
Valores Nulos	0
Valores Atípicos	No aplica
Media	No aplica
Mediana	No aplica
Moda	11-NUMERO DE UNIDADES
Rango Medio	No aplica
Varianza	No aplica
Desviación Estándar	No aplica
Rango	No aplica
Cuartiles	No aplica
Rango Intercuartil	No aplica

Atributo	CODIGO_LIBERACION
Tipo de Dato	Numérico
Tipo de Dato	int64
Valores Nulos	0
Valores Atípicos	1158
Media	18.535656

Mediana	0.0
Moda	0
Rango Medio	2889.0
Varianza	27482.822237
Desviación Estándar	165.779439
Rango	5778
Cuartiles	Q1: 0.0, Q3: 0.0
Rango Intercuartil	0.0

Atributo	COD_LIBERACION
Tipo de Dato	Nominal no categórico
Tipo de Dato	object
Valores Nulos	36325
Valores Atípicos	No aplica
Media	No aplica
Mediana	No aplica
Moda	0424-TRANSFERENCIAS E IMPORTACIONES CON TARIFA CERO "IVA", APLICA TEXTO LEGAL #7 ART. 55 LRTI - Refo
Rango Medio	No aplica
Varianza	No aplica
Desviación Estándar	No aplica
Rango	No aplica
Cuartiles	No aplica
Rango Intercuartil	No aplica

Atributo	ADV_PAG_PARTIDA
Tipo de Dato	Numérico
Tipo de Dato	float64
Valores Nulos	0
Valores Atípicos	7286
Media	86.85122
Mediana	0.0
Moda	0.0
Rango Medio	31137.39
Varianza	633858.718178
Desviación Estándar	796.152447
Rango	62274.78
Cuartiles	Q1: 0.0, Q3: 4.3
Rango Intercuartil	4.3

Atributo	ADV_LIQ_PARTIDA
Tipo de Dato	Numérico
Tipo de Dato	int64
Valores Nulos	0
Valores Atípicos	0
Media	0.0
Mediana	0.0
Moda	0
Rango Medio	0.0
Varianza	0.0
Desviación Estándar	0.0
Rango	0
Cuartiles	Q1: 0.0, Q3: 0.0
Rango Intercuartil	0.0

Atributo	CARACTERISTICA
Tipo de Dato	Nominal no categórico
Tipo de Dato	object
Valores Nulos	13208
Valores Atípicos	No aplica
Media	No aplica
Mediana	No aplica
Moda	REPUESTOS
Rango Medio	No aplica
Varianza	No aplica
Desviación Estándar	No aplica
Rango	No aplica
Cuartiles	No aplica
Rango Intercuartil	No aplica

Atributo	PRODUCTO
Tipo de Dato	Nominal no categórico
Tipo de Dato	object
Valores Nulos	13176
Valores Atípicos	No aplica
Media	No aplica
Mediana	No aplica
Moda	PRODUCTO TERMINADO
Rango Medio	No aplica
Varianza	No aplica
Desviación Estándar	No aplica
Rango	No aplica
Cuartiles	No aplica
Rango Intercuartil	No aplica

Atributo	MARCA_COMERCIAL
Tipo de Dato	Nominal no categórico
Tipo de Dato	object
Valores Nulos	13288
Valores Atípicos	No aplica
Media	No aplica
Mediana	No aplica
Moda	SIN MARCA
Rango Medio	No aplica
Varianza	No aplica
Desviación Estándar	No aplica
Rango	No aplica
Cuartiles	No aplica
Rango Intercuartil	No aplica

Atributo	ANIO_PRODUCIDO
Tipo de Dato	Numérico
Tipo de Dato	float64
Valores Nulos	19945
Valores Atípicos	2336
Media	2021.998917
Mediana	2022.0
Moda	2022.0
Rango Medio	2194.5
Varianza	28.770655
Desviación Estándar	5.363828
Rango	391.0
Cuartiles	Q1: 2022.0, Q3: 2022.0
Rango Intercuartil	0.0

Atributo	MODELO_MERCADERIA
Tipo de Dato	Nominal no categórico
Tipo de Dato	object
Valores Nulos	13234
Valores Atípicos	No aplica
Media	No aplica
Mediana	No aplica
Moda	SIN MODELO
Rango Medio	No aplica
Varianza	No aplica
Desviación Estándar	No aplica
Rango	No aplica
Cuartiles	No aplica
Rango Intercuartil	No aplica

Atributo	NUM_FOB_UNITARIO
Tipo de Dato	Numérico
Tipo de Dato	float64
Valores Nulos	0
Valores Atípicos	5494
Media	1365.980483
Mediana	2.071719
Moda	0.0
Rango Medio	19500000.0

Varianza	40615455410.849304
Desviación Estándar	201532.765105
Rango	39000000.0
Cuartiles	Q1: 0.0, Q3: 15.333159
Rango Intercuartil	15.333159

Atributo	VIA
Tipo de Dato	Nominal no categórico
Tipo de Dato	object
Valores Nulos	1568
Valores Atípicos	No aplica
Media	No aplica
Mediana	No aplica
Moda	AEREO (IMPORTACION)
Rango Medio	No aplica
Varianza	No aplica
Desviación Estándar	No aplica
Rango	No aplica
Cuartiles	No aplica
Rango Intercuartil	No aplica

Atributo	REGIMEN_TIPO
Tipo de Dato	Nominal no categórico
Tipo de Dato	object
Valores Nulos	0
Valores Atípicos	No aplica
Media	No aplica
Mediana	No aplica
Moda	IMP.GRAL.
Rango Medio	No aplica
Varianza	No aplica
Desviación Estándar	No aplica
Rango	No aplica
Cuartiles	No aplica
Rango Intercuartil	No aplica

Atributo	INCOTERM
Tipo de Dato	Nominal no categórico
Tipo de Dato	object
Valores Nulos	13164
Valores Atípicos	No aplica
Media	No aplica
Mediana	No aplica
Moda	FOB
Rango Medio	No aplica
Varianza	No aplica
Desviación Estándar	No aplica
Rango	No aplica
Cuartiles	No aplica
Rango Intercuartil	No aplica

Atributo	CONSOLIDADORA
Tipo de Dato	Numérico
Tipo de Dato	float64
Valores Nulos	37483
Valores Atípicos	0
Media	nan
Mediana	nan
Moda	Sin moda
Rango Medio	nan
Varianza	nan
Desviación Estándar	nan
Rango	nan
Cuartiles	Q1: nan, Q3: nan
Rango Intercuartil	nan

Atributo	COD_PROVINCIA
Tipo de Dato	Nominal no categórico
Tipo de Dato	object
Valores Nulos	0
Valores Atípicos	No aplica
Media	No aplica
Mediana	No aplica
Moda	17
Rango Medio	No aplica
Varianza	No aplica
Desviación Estándar	No aplica
Rango	No aplica
Cuartiles	No aplica
Rango Intercuartil	No aplica

Atributo	PROVINCIA
Tipo de Dato	Nominal no categórico
Tipo de Dato	object
Valores Nulos	0
Valores Atípicos	No aplica
Media	No aplica
Mediana	No aplica
Moda	Pichincha
Rango Medio	No aplica
Varianza	No aplica
Desviación Estándar	No aplica
Rango	No aplica
Cuartiles	No aplica
Rango Intercuartil	No aplica

Atributo	FORMULARIO
Tipo de Dato	Numérico
Tipo de Dato	float64
Valores Nulos	37483
Valores Atípicos	0
Media	nan
Mediana	nan
Moda	Sin moda
Rango Medio	nan
Varianza	nan
Desviación Estándar	nan
Rango	nan
Cuartiles	Q1: nan, Q3: nan
Rango Intercuartil	nan

Atributo	FORM_VIA_ENVIO
Tipo de Dato	Numérico
Tipo de Dato	float64
Valores Nulos	37483
Valores Atípicos	0
Media	nan
Mediana	nan
Moda	Sin moda
Rango Medio	nan
Varianza	nan
Desviación Estándar	nan
Rango	nan
Cuartiles	Q1: nan, Q3: nan
Rango Intercuartil	nan

Atributo	ESTADO_MERCANCIA
Tipo de Dato	Nominal no categórico
Tipo de Dato	object
Valores Nulos	2168
Valores Atípicos	No aplica
Media	No aplica
Mediana	No aplica
Moda	1-NUEVA
Rango Medio	No aplica
Varianza	No aplica
Desviación Estándar	No aplica
Rango	No aplica
Cuartiles	No aplica
Rango Intercuartil	No aplica

Atributo	NUM_FLETE_PAGADO
Tipo de Dato	Numérico
Tipo de Dato	float64

Valores Nulos	0
Valores Atípicos	6300
Media	180.057817
Mediana	4.5
Moda	0.0
Rango Medio	237440.0
Varianza	10699551.093125
Desviación Estándar	3271.016829
Rango	474880.0
Cuartiles	Q1: 0.54, Q3: 19.328
Rango Intercuartil	18.788

Atributo	NUM_CIF2
Tipo de Dato	Numérico
Tipo de Dato	float64
Valores Nulos	0
Valores Atípicos	6253
Media	7644.563807
Mediana	120.22
Moda	10.0
Rango Medio	24835881.6405
Varianza	171181248335.50952
Desviación Estándar	413740.556793
Rango	49671763.279
Cuartiles	Q1: 37.768, Q3: 526.5015
Rango Intercuartil	488.7335

Atributo	CFR
Tipo de Dato	Numérico
Tipo de Dato	float64
Valores Nulos	0
Valores Atípicos	6254
Media	7603.003669
Mediana	119.88
Moda	10.0
Rango Medio	24589981.8205
Varianza	167847227494.70276
Desviación Estándar	409691.624877
Rango	49179963.639
Cuartiles	Q1: 37.555, Q3: 524.375
Rango Intercuartil	486.82

Atributo	ESTADO_DECLARACION
Tipo de Dato	Nominal no categórico
Tipo de Dato	object
Valores Nulos	0
Valores Atípicos	No aplica
Media	No aplica
Mediana	No aplica
Moda	SALIDA AUTORIZADA
Rango Medio	No aplica
Varianza	No aplica
Desviación Estándar	No aplica
Rango	No aplica
Cuartiles	No aplica
Rango Intercuartil	No aplica