

Minería Visual de Datos para la Toma de Decisiones Estratégicas en Importaciones

Visual Data Mining for Strategic Import Decision-Making

Andrés Teodoro Calle Clavijo¹ <https://orcid.org/0009-0009-0104-5458>, Luis Tonon-Ordóñez¹ <https://orcid.org/0000-0003-2360-9911>, Marcos Orellana¹ <https://orcid.org/0000-0002-3671-9362>

¹Laboratorio de Investigación y Desarrollo en Informática (LIDI), Cuenca, Ecuador
acallemsn@es.uazuay.edu.ec, ltonon@uazuay.edu.ec,
marore@uazuay.edu.ec



Esta obra está bajo una licencia internacional
Creative Commons Atribución-NoComercial 4.0

Enviado: 2024/07/12

Aceptado: 2024/10/02

Publicado: 2024/10/15

Resumen

En el comercio internacional, la elección de un proveedor de productos ideal puede ser un desafío. Bajo este antecedente, resulta obligatorio identificar cuáles proveedores y sus países en la importación de cierto bien resultan más adecuados. Por lo tanto, en este trabajo se propone un modelo que permita realizar análisis y visualizaciones de datos de costos de las importaciones al Ecuador dentro del periodo 2008-2018. Para lograrlo, se analizó la información existente sobre las importaciones a través de una revisión de la literatura relacionada. Luego, se aplicó el algoritmo K-means para agrupar los países por cada partida arancelaria, tomando en cuenta el valor FOB y los costos de importación. Finalmente, se diseña una interfaz de visualización para facilitar la toma de decisiones con base en la información obtenida. En consecuencia, el modelo propuesto es útil para la toma de decisiones en las importaciones, debido a que permite realizar análisis de los datos de todos los países en conjunto con los bienes importados, demostrando ser aplicable a una amplia gama de empresas.

Palabras clave: Clusterización, Comercio Internacional, Inteligencia de Negocios, Minería de Datos, Visualización de Datos.

Abstract

In international trade, choosing the ideal supplier can be a challenge. Against this background,

Sumario: Introducción, Trabajos Relacionados, Materiales y Métodos, Resultados, Discusión, Conclusiones.

Como citar: Calle, A., Tonon-Ordóñez, L. & Orellana M. (2024). Minería Visual de Datos para la Toma de Decisiones Estratégicas en Importaciones. *Revista Tecnológica - Espol*, 36(2), 163-176. Recuperado a partir de <https://rte.espol.edu.ec/index.php/tecnologica/article/view/1209>

it is mandatory to identify the countries from which the import of a certain good is most appropriate. Therefore, this paper proposes a model that allows for the analysis and visualization of cost data for imports into Ecuador from 2008-2018. To achieve this, the existing information on imports was analyzed by reviewing the related literature. Then, the K-means algorithm was applied to group countries by tariff heading, considering the FOB value and import costs. Finally, a visualization interface facilitates decision-making based on the obtained information. Consequently, the proposed model is useful for decision-making in imports because it allows data analysis from all countries in conjunction with imported goods, proving to be applicable to a wide range of companies.

Keywords: Business Intelligence, Clustering, Data Mining, Data Visualization, International Trade.

Introducción

El comercio internacional se posiciona como uno de los pilares fundamentales para el desarrollo económico de los países. Al facilitar el intercambio de bienes, servicios y tecnología entre naciones, se dinamiza la economía y se contribuye significativamente al crecimiento del Producto Interno Bruto (PIB) (Rahim et al., 2023). Esta actividad clave genera diversos beneficios como i) aumento del empleo: Los países exportadores experimentan un crecimiento en las oportunidades laborales, impulsando el bienestar social; ii) mayor variedad para los consumidores: Los países importadores se ven beneficiados por una amplia gama de productos y servicios a su disposición, lo cual satisface mejor sus necesidades y preferencias; iii) información para la toma de decisiones: El comercio internacional permite a los consumidores acceder a información sobre diferentes opciones, facilitando la elección de los bienes que mejor se ajustan a sus necesidades y presupuesto (Tâu & Sharfeldin, 2021). Por lo que, el comercio internacional se erige como un motor fundamental para el progreso económico y social de las naciones, impulsando el crecimiento, la generación de empleo y el bienestar de los ciudadanos.

Si bien el comercio internacional es un motor fundamental para el crecimiento económico, la toma de decisiones en este ámbito se ve afectada por altos niveles de incertidumbre (Ponlaem et al., 2021). Este factor eleva el riesgo asociado a las importaciones y exportaciones (Bonaime et al., 2018), dificultando el intercambio fluido de bienes. En este contexto, el análisis de datos de importaciones se convierte en una herramienta crucial para reducir la incertidumbre y facilitar el comercio internacional. Las cifras del Banco Central del Ecuador (BCE) lo confirman: En 2018, las importaciones del país alcanzaron los 22,385 millones de dólares, representando el 20.81% del PIB, y compuestas principalmente por materias primas y bienes de capital (Banco Central del Ecuador, 2023).

El análisis y la comprensión de los costos de importación de las diferentes subpartidas es fundamental para reducir el riesgo en las decisiones de importación y exportación, optimizar los procesos logísticos y de cadena de suministro, identificar nuevas oportunidades comerciales, negociar mejores precios con proveedores internacionales y tomar decisiones estratégicas informadas para la gestión del comercio exterior.

El análisis de datos ha demostrado ser una herramienta invaluable para optimizar la gestión de cadenas de suministro y aumentar la ventaja competitiva en el comercio internacional. Al reducir la incertidumbre en factores como la demanda y las condiciones del mercado, es posible tomar decisiones más informadas y estratégicas (Dubey et al., 2021). Las técnicas de minería de datos son capaces de realizar análisis robustos del comportamiento comercial de los

países, facilitando la toma de decisiones acertadas en materia de importaciones y exportaciones. Sin embargo, la complejidad de la interpretación y comunicación de los resultados obtenidos mediante estas técnicas puede dificultar su comprensión a los usuarios (Chambers, 2018).

Así pues, la visualización de datos es importante, ya que a través de resúmenes visuales y paneles de control o dashboards permite comunicar información de manera clara, concisa y comprensible (J. Kim et al., 2016). Su aplicación en el ámbito del comercio internacional ha demostrado ser particularmente útil para visualizar el comportamiento del comercio entre diferentes regiones, considerando variables como el volumen de compraventa, el costo y su evolución a lo largo del tiempo (S. Kim et al., 2020). La principal ventaja de la visualización de datos radica en su capacidad para ofrecer al usuario una comprensión intuitiva de la información presentada. En el caso de Ecuador, la aplicación de estas técnicas a gran escala permitiría representar gráficamente los costos de importación, facilitando su análisis y la toma de decisiones estratégicas en materia de comercio exterior.

Conforme con los antecedentes expuestos, en esta investigación se propone un modelo de análisis y visualización de datos de importaciones del Ecuador en el período 2008-2018 en base a técnicas de minería de datos. Con la finalidad de reducir la incertidumbre en la toma de decisiones comerciales internacionales al conocer cuáles son los países con los que se puede lograr una mejor relación comercial en base a los costos de importación.

De este modo, este artículo se estructura de la siguiente manera: en la Sección 2 se presentan los trabajos relacionados; la Sección 3 expone la metodología empleada; en la Sección 4 se detallan los resultados obtenidos; la Sección 5 manifiesta las discusiones y finalmente en la última Sección se exhiben las conclusiones y los trabajos futuros.

Trabajos Relacionados

En el ámbito del comercio internacional, el análisis de datos ha emergido como una herramienta fundamental para comprender las dinámicas complejas que lo caracterizan. Diversos estudios han demostrado varias técnicas para identificar patrones, clasificar países y tomar decisiones estratégicas. Siendo así, en un estudio realizado por Simić et al. (2016), examinaron las importaciones de telas de algodón a los Estados Unidos durante el período 1990-2010. Para ello, propusieron un enfoque original que clasificaba a los 200 países proveedores en distintos grupos según su desempeño económico.

Los investigadores obtuvieron datos de valor monetario y peso de las importaciones del repositorio de comercio exterior de Estados Unidos. Posteriormente, aplicaron el algoritmo K-means, una técnica de agrupamiento para identificar grupos de países con características similares. Los resultados revelaron una correlación significativa entre el crecimiento económico de un país y su volumen de exportaciones de algodón a los Estados Unidos. Los países con economías de rápido crecimiento tienden a ser los principales proveedores de algodón al mercado estadounidense.

Un estudio adicional analiza las importaciones de arroz a Indonesia desde diversos países durante el período 2000-2015, incluyendo Vietnam, Tailandia, China, India, Pakistán, Estados Unidos, Taiwán, Singapur y Myanmar. Con el objetivo de identificar los principales proveedores de arroz, se agruparon los países de importación en tres clústeres: alto nivel de importación, introducción media y bajo nivel de importación. Para ello, emplearon el algoritmo K-means y el software RapidMiner, considerando las variables de peso de la mercadería en toneladas, su

valor de costo, seguro y flete o Cost, Insurance, Freight (CIF). Los resultados revelaron que Vietnam y Tailandia, pertenecientes al clúster de alto nivel de importación, eran los principales proveedores de arroz para Indonesia, seguidos por China, India y Pakistán en el clúster de introducción media. Los países restantes se ubicaron en el clúster de bajo nivel de importación (Windarto, 2017).

Dankevych et al. (2018) identificaron socios potenciales para el comercio internacional de productos agrícolas a largo plazo con Ucrania. Se analizaron datos del período 2014-2016 utilizando el algoritmo K-means para agrupar países en tres categorías: "Orientados a la exportación", "Orientados a la importación" y "Socios potenciales". Austria, Bulgaria, Croacia, Chipre, República Checa, Dinamarca, Estonia, Finlandia, Grecia, Irlanda, Letonia, Lituania, Luxemburgo y España fueron identificados como socios potenciales basados en su potencial comercial y perspectivas a largo plazo.

Batarseh et al. (2019) exploraron la viabilidad de utilizar técnicas de minería de datos para agrupar países según su nivel de comercio de productos agrícolas. Se analizaron datos del sistema de agricultura de Estados Unidos entre 1989 y 2018. El algoritmo K-means aplicado con Python agrupa los países en tres clústeres, principales importadores y exportadores entre Estados Unidos y China, países con comercio significativo entre Japón, Alemania, Canadá, Reino Unido, India y Francia, y el resto de los países. Este tipo de análisis puede ser útil para informar la formulación de políticas comerciales.

En otro estudio, se evalúa la eficacia del algoritmo K-means para clasificar países desarrollados y en vías de desarrollo. Se utilizaron datos de 167 países obtenidos del repositorio Kaggle, sin especificar el período de tiempo analizado. Los investigadores analizaron factores relacionados con las importaciones, exportaciones y otras variables. El algoritmo K-means logró clasificar correctamente 32 países como desarrollados y 135 como en vías de desarrollo (Wulandari & Yogantara, 2022).

Los datos del comercio internacional ofrecen información valiosa para comprender las dinámicas económicas globales, identificar oportunidades comerciales y tomar decisiones estratégicas. La visualización de datos juega un papel crucial en la transformación de estos datos complejos en información accesible y útil para diversos actores, como gobiernos, empresas y organismos internacionales.

En el estudio de SIKOS T & Meirmanova (2020), examinaron las interconexiones en el comercio mundial de trigo utilizando representaciones geográficas y grafos. Se identifica una red de escala libre, donde algunos países tienen un papel central en el comercio, mientras que otros participan de manera marginal. Esta información puede ser útil para comprender la vulnerabilidad del mercado ante eventos disruptivos y para diseñar políticas comerciales efectivas.

Otro estudio evalúa el comercio internacional de Corea del Sur durante el período 2007-2017, creando visualizaciones que muestran las importaciones y exportaciones por país y tipo de bien (Dar et al., 2020). Se identificó a China, Estados Unidos, Vietnam, Hong Kong y Japón como los principales socios comerciales de Corea del Sur. Esta información puede ser utilizada por empresas y autoridades gubernamentales para identificar oportunidades de mercado y optimizar las estrategias comerciales.

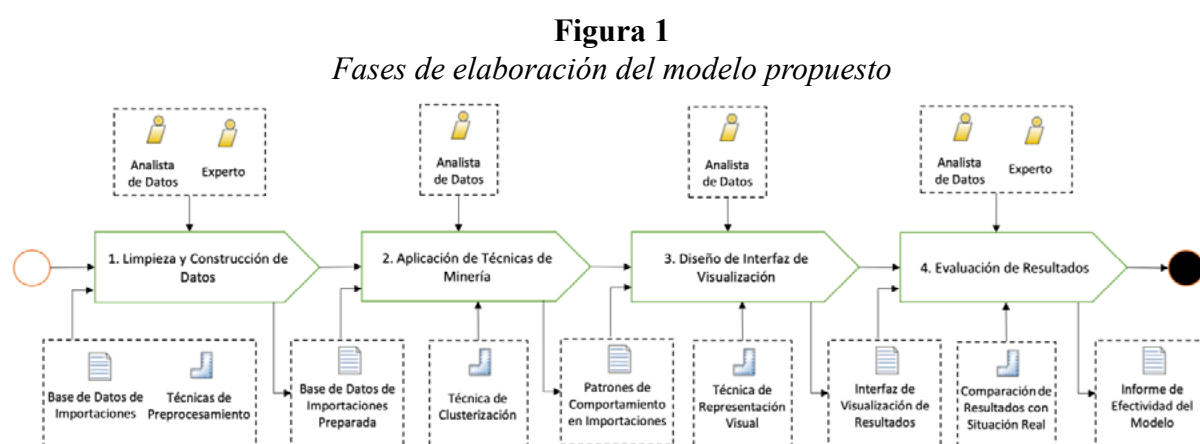
Pacini et al. (2021) investigaron sobre el comercio internacional de residuos de plástico en 2018 utilizando visualizaciones en forma de grafos. En este estudio, el país de Alemania fue determinado como el mayor comerciante, seguido de Estados Unidos, Bélgica, Italia y otros países. Esta información puede ser útil para comprender los flujos de residuos y para implementar políticas de gestión ambiental efectivas.

Una investigación explora las exportaciones de frutas de Brasil entre 1986 y 2017, utilizando visualizaciones en forma de grafos para mostrar los destinos y el nivel de ingresos de los países importadores, Bernal et al. (2021). En este estudio se señala a Europa como el principal mercado, destacando la importancia de la visualización de datos para comprender las tendencias del comercio internacional de productos agrícolas.

En esta investigación se propone una interfaz interactiva para visualizar los precios de franco a bordo o Free on Board (FOB) y costos de importación del Ecuador. La interfaz incluye gráficos de barras, grafos y filtros para facilitar la toma de decisiones comerciales internacionales.

Materiales y Métodos

Para la elaboración del modelo se contemplan las siguientes fases: i) limpieza y construcción de los datos, ii) aplicación de técnicas de minería, iii) diseño de interfaz de visualización y iv) evaluación de resultados. La Figura 1 presenta las fases de la metodología utilizando el Meta-Modelo 2.0 de Ingeniería de Procesos de Software y Sistemas (SPEM 2.0).



Limpieza y construcción de datos

Para esta fase se utilizó como entrada la base de datos de importaciones del Ecuador obtenida del sitio web del Banco Central del Ecuador (Banco Central del Ecuador, 2023). El análisis fue ejecutado para el periodo 2008-2018. Este conjunto de datos fue seleccionado debido a la disponibilidad de una base de datos consolidados de todos los países de origen y partidas arancelarias importadas entre esos años (Medina López et al., 2023). En esta instancia, el objetivo fue transformar la base disponible en un conjunto de datos adecuado y aplicar las técnicas de minería de datos necesarias. Para ello, se evaluó la necesidad y utilidad de cada variable y se utilizó el lenguaje de programación Python.

El conjunto de datos consta de 1,471,125 registros, una cantidad grande para ser manejada en un solo archivo. Por este motivo, fue dividida en 2 hojas del documento, las mismas que fueron cargadas al procesador de forma individual y luego fueron unificadas. Luego de la carga, se evidenció que la base de datos constaba de 42 variables Categóricas, Nominales, Numéricas, Categóricas Ordinales y Dicotómicas.

El análisis fue ejecutado en base a los datos de los precios del franco a bordo o Free On Board (FOB) y costos de importación. Para esto, fueron consideradas las variables “Valor FOB de la importación” y “Costo de importación”. Sin embargo, los valores de estas variables se presentaban en base al total de importaciones anuales de cada partida arancelaria, por lo que estaban en función de la cantidad de toneladas importadas. A fin de evitar sesgos en la interpretación, se analizaron los valores por tonelada. Para esto, se generaron las variables “FOB por tonelada” y “Costo de importación por tonelada”. Esto se logró dividiendo respectivamente los valores de las variables “Valor FOB de la Importación” y “Costo de importación” entre los datos de la variable “Toneladas importadas” para cada año, país de importación y partida arancelaria.

Por otra parte, para el proceso de clusterización, es preferible trabajar con aquellas variables que brinden información útil durante el análisis. Por este motivo, una metodología conocida para prescindir de variables que no aportan valor es eliminar aquellas que brindan la misma información o tienen un alto nivel de correlación (Hsu et al., 2010). Así, no todas las variables disponibles fueron necesarias. El objetivo de este estudio es conocer los países importadores más adecuados para cada partida, por lo que se seleccionaron las variables como el nombre del país de importación, el año de la importación, código de la partida “NANDINA” actual y el costo de importación por tonelada y FOB por tonelada. Los criterios fueron seleccionados en base a cuáles variables pueden brindar información sobre si es conveniente o no importar un bien seleccionado de un país específico desde el punto de vista económico.

Luego, se evidenció la existencia de valores en blanco en ciertas variables. Estos valores pueden repercutir en complicaciones al aplicar los algoritmos. Sin embargo, dichos valores pueden ser reemplazados o eliminados (Sehgal & Bhargava, 2018). En los análisis comerciales, la ausencia de registros implica que no existió intercambio alguno, por lo tanto, a los datos en blanco se los reemplazó con ceros.

Para ejecutar el análisis, primero se obtuvieron los valores promedio de las variables “FOB por tonelada” y “Costo de importación por tonelada” en función de las variables “Año de la importación”, “Nombre del país de importación en inglés” y “Código de partida NANDINA actual”. De esta manera, se consiguió un total de 124,150 registros correspondientes a 224 países y 7,131 partidas arancelarias.

Aplicación de Técnicas de Minería

Una vez culminado el preprocesamiento del conjunto de datos, fue aplicada la técnica de clusterización. Como se mencionó anteriormente, el algoritmo K-means ha sido ampliamente utilizado para clusterización en varios estudios relacionados a agrupación de países para análisis comerciales (Dankevych et al., 2018; Simić et al., 2016; Windarto, 2017). Sin embargo, este método requiere que previamente se defina el número de clústeres para trabajar.

Aplicación de clusterización

Aunque la aplicación del algoritmo K-means es el método más utilizado para clusterizar (Sinaga & Yang, 2020), la cantidad de clústeres (k) en los que se desea agrupar los datos es requerida como entrada; lo cual puede resultar complejo de determinar. Por otra parte, dicho número puede ser predefinido en base a las necesidades del negocio (Kodinariya et al., 2013). Así, trabajos previos que han agrupado países en base a variables económicas han dividido los datos en 3 niveles –alto, medio y bajo– (Dankevych et al., 2018).

En esta investigación se persigue un objetivo análogo y se ha considerado definir tres grupos de países por cada partida arancelaria ($k = 3$). Por otra parte, existen partidas arancelarias que fueron importadas desde una cantidad de países menor o igual a la cantidad de grupos predefinida. En estos casos, se utilizó un número de clústeres igual a la cantidad de naciones desde las que se importó el bien, resultando en un país dentro de cada clúster.

Una vez definido el número de clústeres, se utilizó el algoritmo K-means con los datos de importaciones utilizando la librería scikit-learn de Python. El script desarrollado realiza una iteración por cada partida arancelaria. En cada repetición se agrupa a los países desde donde se importó la partida en base a las variables “FOB por tonelada” y “Costo de importación por tonelada” de acuerdo con el número de clústeres definido para dicha clasificación arancelaria. En esta instancia, los clústeres obtenidos fueron representados por los números 0, 1 y 2 y no siempre están conformados por los mismos grupos de países. Luego, se añadió un campo al dataset o conjunto de datos para especificar el número de clústeres obtenidos del proceso, de modo que pueda ser incluido en la fase de visualización.

Diseño de Interfaz de Visualización

Para facilitar la interpretación de los resultados obtenidos y la toma de decisiones, se diseñó un entorno para la visualización de los grupos obtenidos y su relación con los datos iniciales mediante interacción del usuario final. Este proceso fue realizado mediante el software Microsoft Power BI, debido a que permite representar los datos de forma efectiva y el tipo de visualización depende de factores como los tipos de datos, las preguntas que se busca responder y al público que va dirigido el informe. Lo que se busca con la visualización de los datos es la simplicidad, claridad, coherencia, interactividad, y accesibilidad de los datos.

Interfaz e Interacción del usuario

Primero, se vinculó el código de la partida de la base de datos clusterizados con la información de las partidas arancelarias de la base inicial, de modo que en la interfaz fueron colocadas las variables “Nivel 1 de la Partida”, “Nivel 2 de la Partida”, “Nivel 3 de la Partida” y “Nivel 4 de la Partida” para facilitar la búsqueda de información con respecto a cada tipo de bien importado. Por otra parte, fue incluida la variable “Clúster” para identificar a qué grupo pertenece cada país desde la interfaz.

Como siguiente paso, en la interfaz se colocaron los valores promedio de las variables “Free On Board (FOB)” y “Costo de importación” de cada grupo de países clusterizados por su importancia en el análisis de importaciones. Además, fue incluido un elemento de identificación de las unidades en las que se mide cada partida arancelaria. Finalmente, se incluyeron dos representaciones de las variables “Free On Board (FOB)” y “Costo de importación” mediante grafos con el país Ecuador en el nodo central. En esta instancia, los tamaños de los nodos de los grafos fueron parametrizados de tal manera que resulten proporcionales a los valores que toman las variables, mientras que su color fue utilizado para indicar el clúster al que pertenecía el país representado.

El objetivo del trabajo es facilitar la toma de decisiones para el usuario final a través de representaciones visuales de los datos y que permita comprender de forma sencilla las complejas relaciones de los datos. Para ello, se utilizó la visualización de los datos como una herramienta que permita identificar a los países con los que sería más conveniente ejecutar procesos de importación en base al costo de importación y valor FOB. Así, para facilitar la interacción del usuario, se colocaron opciones de filtrado a las variables “Nivel 1 de la Partida”, “Nivel 2 de la Partida”, “Nivel 3 de la Partida”, “Nivel 4 de la Partida” de modo que se pueda acceder de

manera sencilla e intuitiva a la información requerida. Además, se incluyó el campo “Clúster”, que también puede ser filtrado, para poder identificar los países pertenecientes a cada clúster de las partidas.

Evaluación de resultados

El modelo fue sometido a una evaluación por parte de expertos en el campo del comercio internacional y la minería de datos, quienes confirmaron su validez y aplicabilidad en diversos escenarios. Los expertos destacaron la solidez de la metodología, la precisión de los resultados y la utilidad potencial del modelo para la toma de decisiones estratégicas. Por lo tanto, los expertos evaluaron el modelo en base a la solidez de la metodología, precisión de los resultados, aplicabilidad del modelo y el análisis para identificar los aspectos positivos y negativos del modelo, así como las áreas de mejora potenciales.

Resultados

Como parte del proceso de investigación, fue elaborada una base de datos completa que contiene promedios de precios FOB por tonelada y costos de importación por tonelada para cada país de origen de las importaciones realizadas en Ecuador durante el período 2008-2018. La información se encuentra organizada por partida arancelaria “NANDINA”, lo que facilita el análisis y la comparación de datos entre diferentes productos. Para ilustrar la aplicación de la base de datos, fue seleccionada la partida arancelaria 85171200, que corresponde a "teléfonos móviles (celulares) y los de otras redes inalámbricas". Esta partida fue elegida por su simplicidad y por ser un producto de uso común en la actualidad. La descripción de la partida arancelaria con el código 85171200 es:

- Sección VIII: Máquinas, aparatos y material eléctrico; sus partes; aparatos de grabación o reproducción de sonido, aparatos de grabación o reproducción de imagen y sonido en televisión, y las partes y accesorios de estos aparatos.
- Capítulo 85: Máquinas, aparatos y material eléctrico, y sus partes.
- Partida 8517: Teléfonos, incluidos los teléfonos móviles (celulares) y los de otras redes inalámbricas, los demás aparatos para emisión, transmisión o recepción de voz, imagen u otros datos, incluidos los de comunicación en red con o sin cable.
- Subpartida 851712: No hay descripción disponible.
- Ítem 85171200: Teléfonos móviles (celulares) y los de otras redes inalámbricas.

El análisis del conjunto de datos para la partida arancelaria 85171200 permitió identificar los países de origen con los precios FOB y costos de importación más favorables para la importación de teléfonos móviles hacia Ecuador. Esta información puede ser de gran utilidad para empresas importadoras, entidades gubernamentales y otros actores interesados en el comercio internacional de este tipo de productos. En la Tabla 1 se muestra la agrupación obtenida de la partida arancelaria de ejemplo.

Tabla 1
Conjunto de datos resultante (Fragmento)

COSTO DE IMPORTACIÓN POR TONELADA	FOB POR TONELADA	NOMBRE DEL PAÍS DE IMPORTACIÓN	GRUPO
28,808.81	1,295,847.76	Canada	0
16,425.13	783,938.49	Colombia	0

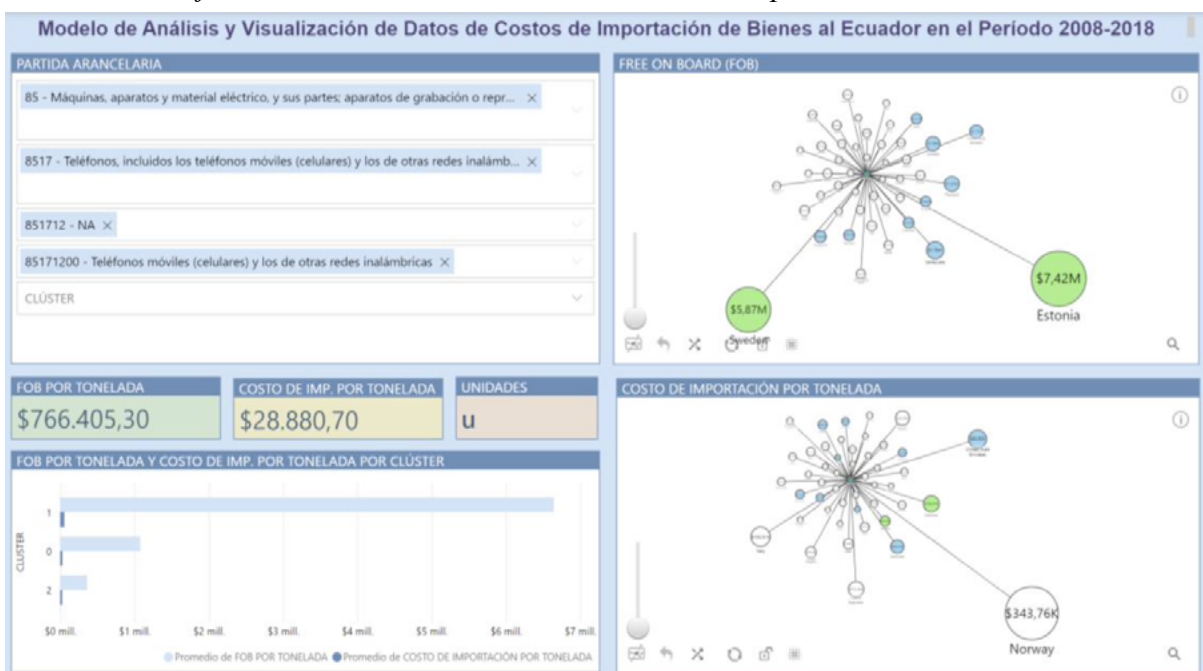
COSTO DE IMPORTACIÓN POR TONELADA	FOB POR TONELADA	NOMBRE DEL PAÍS DE IMPORTACIÓN	GRUPO
58,716.67	799,096.67	Denmark	0
5,293.19	369,255.89	United States of America	2
6,652.95	532,318,35	Vietnam	2

En la Tabla 1 se observa que cada país presente en el campo “Nombre del País de Importación” pertenece a un clúster asignado en el campo “Grupo” identificado por los valores 0, 1 y 2 (3 clústeres) en base a las variables planteadas inicialmente. Para el ejemplo presentado, la interpretación puede ser la siguiente:

- Clúster 0: Se encuentran los países desde los cuales el costo promedio de importación es 30,383.00 dólares por tonelada y el valor FOB promedio de la importación es 1,078,415.29 dólares por tonelada.
- Clúster 1: Se encuentran los países desde los cuales el costo promedio de importación es 56.411,34 dólares por tonelada y el valor FOB promedio de la importación es 6,644,241.87 dólares por tonelada.
- Clúster 2: Se encuentran los países desde los cuales el costo promedio de importación es 26,975.34 dólares por tonelada y el valor FOB promedio de la importación es 361,856.33 dólares por tonelada.

De acuerdo con las necesidades del negocio, es posible identificar cuáles son los países desde los que resulta más conveniente importar una partida “NANDINA” específica en base a los costos y valor FOB en los que se incurre para adquirir el bien descrito desde el Ecuador. A su vez, estos resultados fueron representados en un informe visual (dashboard), cuya interfaz se diseñó para facilitar el acceso a la información disponible de todas las partidas y países con interacciones sencillas e intuitivas por parte del usuario, como se representa en la Figura 2.

Figura 2
Interfaz de visualización de datos de costos de importaciones al Ecuador



En el informe visual presentado, cada cuadro permite el acceso a la información que requiera el usuario de acuerdo con el tipo de bien que desea importar. Para esto, los filtros y los gráficos dinámicos facilitan la comparación entre los costos de importación y FOB de cada país. Los campos de la sección superior izquierda de la Figura 2 muestran todos los posibles filtros del informe visual, desde lo más general (Nivel de Partida 1) a lo más específico (Clúster), como se puede observar en la Figura 3.

Figura 3
Filtros utilizados del informe

Los valores de la sección media izquierda muestran el promedio de las variables FOB y “Costo de importación”. Se colocaron como elementos separados por su relevancia en el análisis, como se observa en la Figura 4.

Figura 4
Valores visualizados en el informe y unidades de la partida arancelaria

FOB POR TONELADA	COSTO DE IMP. POR TONELADA	UNIDADES
\$766.405,30	\$28.880,70	u

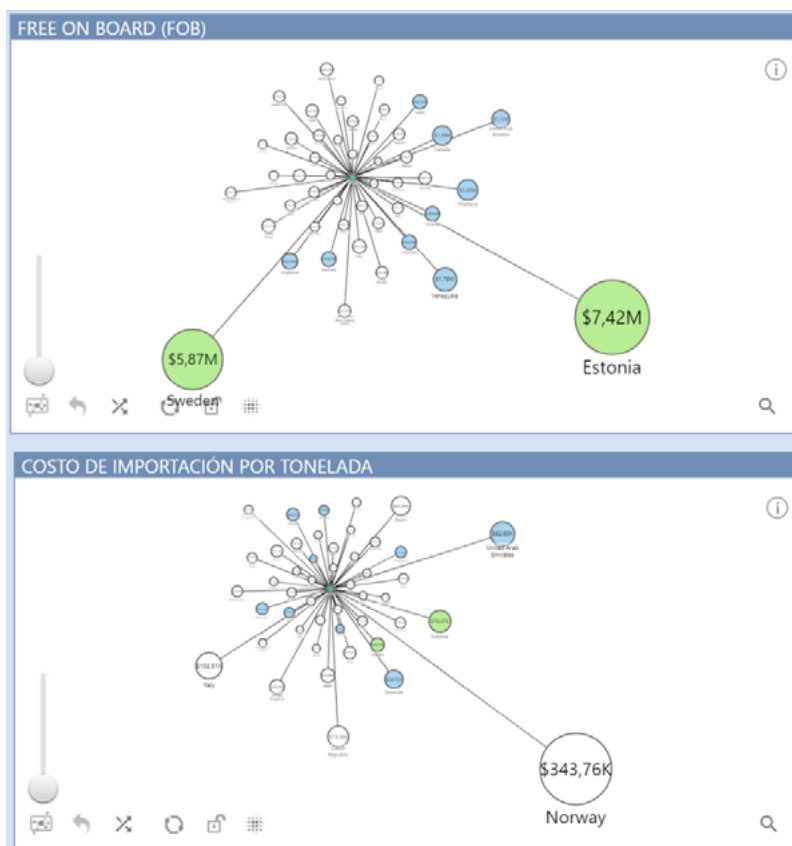
En la sección inferior izquierda, se muestra la gráfica de barras que puede ser utilizada para comparar la diferencia entre los valores promedio que tomaron las variables FOB y “Costo de importación” para cada clúster, como se muestra en la Figura 5.

Figura 5
Comparación de valores entre los clústeres



Finalmente, en la Figura 6 se aprecia la representación de las variables FOB y “Costo de importación” en dos visualizaciones de grafos para facilitar la comparación de los valores que adquieren en base al país de origen, así como el clúster al que pertenece cada nación.

Figura 6
Valores visualizados en el informe y unidades de la partida arancelaria



El modelo obtenido fue validado, de manera cualitativa, por un equipo de expertos conformado por profesores de las escuelas de Estudios Internacionales y Administración. Asimismo, para la validación se compararon los resultados obtenidos con la información provenientes de otras fuentes, como las bases de datos oficiales, estudios previos e informes de mercado. El resultado de esta fase indica que, si bien hay aspectos en los que se podría profundizar para determinar cuál es el mejor socio comercial, el modelo es un buen punto de partida para identificar cuáles son los mejores países para importar cada partida arancelaria.

Discusión

Este trabajo revisa la literatura existente sobre la visualización y el procesamiento de datos en estudios sobre importaciones internacionales. Si bien estudios previos han analizado datos de importaciones de países, este trabajo se distingue por su enfoque práctico, dirigido a facilitar la toma de decisiones por parte de empresarios que buscan identificar los países óptimos para importar determinados bienes.

El estudio de Simić et al. (2016) ha analizado datos de importaciones para agrupar países mediante la técnica de clusterización. Sin embargo, su enfoque se limita a un solo tipo de producto, lo que restringe su aplicación a empresas de un sector específico. En contraste, este trabajo presenta un modelo de análisis generalizable a todos los tipos de productos (Partida NANDINA) importados a Ecuador. Esta amplitud lo convierte en una herramienta valiosa para una gran variedad de empresas, ya que les permite identificar los países óptimos para importar cualquier tipo de bien según sus necesidades específicas. Windarto (2017), quien agrupó países importadores de un solo producto, observó una tendencia a clasificarlos en 3 clústeres, enfoque

también utilizado por Dankevych et al. (2018) y Batarseh et al. (2019). Considerando la validez de este enfoque, se replicó para agrupar a los países proveedores de Ecuador.

El análisis de los resultados evidenció la robustez del algoritmo K-means para la agrupación de países, corroborando su efectividad, como lo mencionan (Wulandari & Yogantara, 2022). En cuanto a la visualización, se optó por métodos tradicionales como los propuestos por SIKOS T & Meirmanova (2020) y Dar et al. (2020). Si bien estos trabajos se enfocan en variables como el nivel de importaciones y exportaciones, el presente trabajo mantiene el foco en el usuario final, permitiendo identificar diferencias en las variables de interés, en este caso, el FOB y “Costo de importación”. Bernal et al. (2021) y Pacini et al. (2021), quienes exploraron con éxito la representación de datos de comercio internacional mediante grafos. Este estudio adoptó este enfoque. Cada país se representa como un nodo, cuyo tamaño se relaciona directamente con el costo total (bien e importación) y cuyo color indica el clúster al que pertenece, facilitando una identificación rápida y sencilla.

En consonancia con el trabajo de Liu et al. (2021) sobre la utilidad de los análisis de datos de comercio para la gestión de rutas comerciales enfocadas en el transporte de bienes, este estudio se presenta como un punto de partida para identificar potenciales nuevas rutas comerciales o potenciar las existentes entre Ecuador y los países con los que resulta más conveniente comerciar desde una perspectiva económica.

Conclusiones

Elegir el país adecuado para importar productos puede ser un desafío sin las herramientas de análisis correctas. Para abordar esto, se desarrolló un modelo para analizar y visualizar los costos de importación de bienes a Ecuador entre 2008 y 2018. Este modelo facilita la identificación de los mejores países para importar cada producto, lo que reduce la incertidumbre en la toma de decisiones comerciales internacionales.

Tras revisar la literatura sobre visualización y procesamiento de datos de comercio internacional, se encontró que el algoritmo K-means es el modelo de clusterización más utilizado debido a su facilidad de uso y efectividad. Además, la representación de datos en forma de grafos se ha vuelto popular para mostrar relaciones entre entidades, ya que permite transmitir una gran cantidad de información a través de recursos visuales como el color y el tamaño de los nodos. La clusterización, utilizando el algoritmo K-means, demostró que para encontrar patrones de comportamiento y obtener información relevante de una base de datos, era crucial comprender las necesidades del usuario y seleccionar las variables correctas antes de realizar cualquier procesamiento tecnológico. La creación del modelo de visualización basado en representaciones gráficas reveló una amplia gama de posibilidades para transmitir información de forma visual. Esto lleva a la conclusión que enfocarse en la funcionalidad del modelo y en la creatividad para simplificarlo sin sacrificar la calidad de la información transmitida son aspectos esenciales en la visualización de datos.

La evaluación de expertos confirmó la validez y aplicabilidad del modelo en diversos escenarios. Un ejemplo es su uso por parte de empresarios que buscan el mejor país para importar productos específicos de manera más económica, lo que se traduce en una mayor rentabilidad para sus empresas. Este análisis también es útil para que los funcionarios públicos formulen y establezcan políticas comerciales que mejoren la competitividad de Ecuador. Además, el personal que establece rutas comerciales hacia Ecuador puede utilizar el modelo para identificar nuevas conexiones potenciales en función de los costos de importación.

Entre las recomendaciones para futuros trabajos está la incorporación de otras herramientas visuales que permitan la observación de los datos de importaciones por un periodo mayor al 2018, con la finalidad de conglomerar más datos y que se presente de forma simple y estructurada. Asimismo, con respecto al análisis temporal, sería factible comparar la variación de los costos entre un período y otro.

Reconocimientos

Los autores desean agradecer al Vicerrectorado de Investigaciones de la Universidad del Azuay por el apoyo financiero y académico, así como a todo el personal de la escuela de Ingeniería de Ciencias de la Computación y el Laboratorio de Investigación y Desarrollo en Informática (LIDI).

Referencias

- Banco Central del Ecuador. (2023). *Comercio Exterior*.
- Batarseh, F., Gopinath, M., Nalluru, G., & Beckman, J. (2019). Application of machine learning in forecasting international trade trends. *ArXiv Preprint ArXiv:1910.03112*.
- Bonaime, A., Gulen, H., & Ion, M. (2018). Does policy uncertainty affect mergers and acquisitions? *Journal of Financial Economics*, 129(3), 531–558.
- Bornal, D. R., Silvestrini, M. M., Pio, L. A. S., Costa, A. C., Peche, P. M., & Ramos, M. C. P. (2021). Brazilian position in the international fresh fruit trade network. *Revista Brasileira de Fruticultura*, 43, e-021.
- Chambers, J. M. (2018). *Graphical methods for data analysis*. Chapman and Hall/CRC.
- Dankevych, V., Dankevych, Y., & Pyvovar, P. (2018). Clustering of the international agricultural trade between Ukraine and the EU. *Management Theory and Studies for Rural Business and Infrastructure Development*, 40(3), 307–319.
- Dar, Q., Dar, G. F., Ma, J.-H., & Ahn, Y.-H. (2020). Visualization, economic complexity index, and forecasting of South Korea international trade profile: a time series approach. *Journal of Korea Trade*, 24(1), 131–145.
- Dubey, R., Gunasekaran, A., Childe, S. J., Fosso Wamba, S., Roubaud, D., & Foropon, C. (2021). Empirical investigation of data analytics capability and organizational flexibility as complements to supply chain resilience. *International Journal of Production Research*, 59(1), 110–128.
- Hsu, H.-H., Hsieh, C.-W., & others. (2010). Feature Selection via Correlation Coefficient Clustering. *J. Softw.*, 5(12), 1371–1377.
- Kim, J., Jo, I.-H., & Park, Y. (2016). Effects of learning analytics dashboard: analyzing the relations among dashboard utilization, satisfaction, and learning achievement. *Asia Pacific Education Review*, 17, 13–24.
- Kim, S., Ku, S., Chang, W., & Song, J. W. (2020). Predicting the direction of US stock prices using effective transfer entropy and machine learning techniques. *IEEE Access*, 8, 111660–111682.
- Kodinariya, T. M., Makwana, P. R., & others. (2013). Review on determining number of Cluster in K-Means Clustering. *International Journal*, 1(6), 90–95.
- Liu, H., Chen, X., Wang, Y., Zhang, B., Chen, Y., Zhao, Y., & Zhou, F. (2021). Visualization and visual analysis of vessel trajectory data: A survey. *Visual Informatics*, 5(4), 1–10.
- Medina Lopez, D. S., Orellana, M., Tonon Ordóñez, L. B., & Zambrano-Martinez, J. L. (2023). Modelo Visual del Comercio Externo en Exportaciones Ecuatorianas. *Revista Tecnológica - ESPOL*, 35(2), 143–156. <https://doi.org/10.37815/rte.v35n2.1051>

- Pacini, H., Shi, G., Sanches-Pereira, A., & da Silva Filho, A. C. (2021). Network analysis of international trade in plastic scrap. *Sustainable Production and Consumption*, 27, 203–216.
- Ponlaem, K., Kaewsompong, N., Maneejuk, P., & Sirisrisakulchai, J. (2021). Impact of Economic Policy Uncertainty on Thailand Macroeconomic Variables. *Behavioral Predictive Modeling in Economics*, 437–451.
- Rahim, T., Ibrahim, M., Shah, T. A., & Mehmood, W. (2023). Impact of Foreign Trade on Economic Growth: Empirical Evidence from Pakistan. *Journal of Policy Research*, 9(1), 188–195.
- Sehgal, M., & Bhargava, D. (2018). Knowledge mining: an approach using comparison of data cleansing tools. *Journal of Information and Optimization Sciences*, 39(1), 337–343.
- SIKOS T, T., & Meirmanova, A. (2020). *Geo-based visual network analysis of export and import patterns in international wheat trade*.
- Simić, D., Svirčević, V., Sremac, S., Ilin, V., & Simić, S. (2016). An efficiency k-means data clustering in cotton textile imports. *Proceedings of the 9th International Conference on Computer Recognition Systems CORES 2015*, 255–264.
- Sinaga, K. P., & Yang, M.-S. (2020). Unsupervised K-means clustering algorithm. *IEEE Access*, 8, 80716–80727.
- Tâu, N., & Sharfeldin, I.-M.-M. (2021). Abordarea teoretică a mediului internațional de afaceri. *Administrarea Publică*, 109(1), 113–126.
- Windarto, A. P. (2017). Implementation of data mining on rice imports by major country of origin using algorithm using k-means clustering method. *International Journal of Artificial Intelligence Research*, 1(2), 26–33.
- Wulandari, L., & Yogantara, B. O. (2022). Algorithm analysis of K-means and fuzzy C-means for clustering countries based on economy and health. *Faktor Exacta*, 15(2).