**Segmentación de Mercados y Productos agrícolas tradicionales y no tradicionales de Exportación Colombianos mediante K-Means**

**Daniela Cárdenas Bermúdez**

**Jhon Jairo Cárdenas Romero**

**Víctor Manuel Fuentes Botero**

**Eber Alexander Marín**

Trabajo final presentado como requisito para optar por la certificación del:

**Bootcamp de Inteligencia Artificial**

**Talento Tech**

Manizales, Caldas

Colombia

[1. Planteamiento del Problema 2](#_mqczlnjg8st5)

[Beneficiarios del proyecto: 3](#_7alts07tkrzl)

[2. Objetivo General 3](#_eia59l5zt5x1)

[Objetivos Específicos 3](#_8ex0c34mau3r)

[3. Introducción 4](#_w611rjlhasuq)

[4. Justificación 4](#_nttu8vt5apt0)

[5. Metodología 4](#_rvk4porg300)

[5.1. Análisis Exploratorio de Datos (EDA) 5](#_sgog3ojp0beh)

[5.1.1. Descripción de los Datos 5](#_tbyrwvdonf4)

[5.1.3. Visualización: 6](#_4056s6jny19o)

[5.2. Modelo de Agrupación K-Means 9](#_l2vzwkiuf75o)

[5.3. Resultados y Análisis 10](#_t68q3428x75g)

[6. Alcance 10](#_pznd48uq5nn4)

[7. Presupuesto 11](#_lq5mct56dnh0)

[9. Conclusiones 11](#_1am34260ffbj)

[8. Webgrafía: 12](#_2rcvp9xahrzb)

Segmentación de Mercados y Productos de Exportación Colombianos mediante K-Means

## **1. Introducción**

El desarrollo de estrategias efectivas para la diversificación de exportaciones es esencial en el contexto actual de comercio global. Colombia, con su variado portafolio de productos exportables —que abarca desde bienes tradicionales como el café, el carbón y las flores, hasta productos no tradicionales como alimentos procesados y derivados de frutas exóticas— tiene una gran oportunidad para incrementar su competitividad en el mercado internacional. Sin embargo, este potencial enfrenta desafíos importantes, tales como la dependencia de pocos mercados, fluctuaciones en los precios y una falta de segmentación específica que permita maximizar el valor de las exportaciones en función de cada destino.

Para abordar estos desafíos, es fundamental entender los patrones de exportación y segmentar los mercados y productos de manera que se optimicen los recursos de exportación. La segmentación de mercados y productos es una herramienta poderosa que permite agrupar datos en función de características compartidas, facilitando así la toma de decisiones estratégicas. Mediante la aplicación de técnicas avanzadas de inteligencia artificial, como el modelo K-Means, es posible identificar grupos homogéneos de productos y destinos de exportación, proporcionando insights estratégicos para los exportadores colombianos. K-Means es un algoritmo de agrupación no supervisada que permite detectar patrones en grandes volúmenes de datos, ofreciendo una estructura que puede utilizarse para orientar la estrategia de exportación y distribución de productos hacia los mercados con mayor potencial de crecimiento.

## **2. Planteamiento del Problema**

Colombia presenta un amplio portafolio de productos de exportación, que incluye bienes tradicionales, como el café y el carbón, y productos no tradicionales, como los derivados de frutas y flores. Sin embargo, la estrategia actual de exportación enfrenta desafíos importantes: la concentración de exportaciones en pocos mercados, fluctuaciones en los precios de los productos y la falta de conocimiento detallado sobre los segmentos de mercado con mayor potencial de crecimiento. Estos factores limitan la competitividad de Colombia en el mercado global, ya que los recursos se destinan de manera homogénea, sin una segmentación estratégica que permita maximizar el valor y volumen de exportación para cada tipo de producto y mercado específico.

La dificultad de identificar patrones específicos entre miles de registros de exportaciones es un reto que puede abordarse mediante técnicas avanzadas de inteligencia artificial. En particular, el modelo de agrupación K-Means permite segmentar mercados y productos de exportación en función de patrones y características compartidas, facilitando una comprensión detallada de los diferentes segmentos de mercado y sus comportamientos. Sin embargo, actualmente no existen estudios que apliquen esta técnica para optimizar las exportaciones colombianas a través de la segmentación no supervisada de sus datos de exportación.

### **Beneficiarios del proyecto:**

- Entidades gubernamentales: Ministerio de Comercio, Industria y Turismo, ProColombia.

- Sectores privados interesados en la expansión de mercados internacionales.

- Inversionistas

- Instituciones de análisis económico y de mercado.

## **2. Objetivo General**

Identificar patrones y segmentar productos y mercados de exportación mediante el modelo K-Means, facilitando recomendaciones estratégicas para maximizar el valor de las exportaciones colombianas.

## **Objetivos Específicos**

* Explorar los datos de exportación y seleccionar variables relevantes para el análisis de agrupación.
* Aplicar la técnica de K-Means para agrupar productos y mercados con características similares.
* Analizar los patrones de cada clúster e identificar oportunidades de optimización de exportación.
* Desarrollar recomendaciones estratégicas basadas en los perfiles de cada clúster identificado.

## **4. Justificación**

La globalización ha incrementado la competencia internacional, lo que exige una mejor segmentación y especialización en exportaciones. Este proyecto busca aprovechar el análisis de datos y técnicas de machine learning para dotar a los exportadores colombianos de una herramienta que facilite la identificación de mercados objetivo, maximizando así el valor de sus operaciones y recursos. La aplicación de K-Means permitirá descubrir y visualizar patrones en los datos de exportación, proporcionando insights a nivel de mercado y producto que actualmente no están disponibles.

Este proyecto forma parte del Bootcamp de Inteligencia Artificial de Talento Tech y utiliza K-Means para analizar grandes volúmenes de datos de exportación. Al aplicar esta técnica no supervisada, se busca descubrir patrones ocultos que puedan informar una estrategia de exportación optimizada, permitiendo al sector exportador colombiano mejorar su competitividad en los mercados globales.

## 

## **5. Metodología**

### **5.1. Análisis Exploratorio de Datos (EDA)**

#### **5.1.1. Descripción de los Datos**

Al iniciar el Análisis de datos Exploratorio se halló que existían 475788 Filas y 9 Columnas, cuya estructura corresponde a las siguientes variables clave:

**Partida:** se refiere a un código o número que clasifica productos específicos de acuerdo con un sistema de nomenclatura, como el Sistema Armonizado (SA). Este sistema permite identificar y categorizar mercancías para facilitar el comercio internacional.

**Descripción Partida10 Dig:** Descripción detallada de la partida, relacionada con los productos exportados.

**Tradición productos:** Indica la tradición o tipo de productos que se están exportando.

**Departamento:** El departamento o región desde donde se realizan las exportaciones.

**País:** El país al que se exportan los productos.

**Año:** El año en que se registraron las exportaciones.

**Mes:** El mes en que se realizaron las exportaciones.

**Exportaciones en valor (Miles USD FOB):** El valor de las exportaciones expresado en miles de dólares estadounidenses, en condiciones FOB (Free On Board).

**Exportaciones en volumen (Toneladas):** La cantidad de productos exportados medida en toneladas.

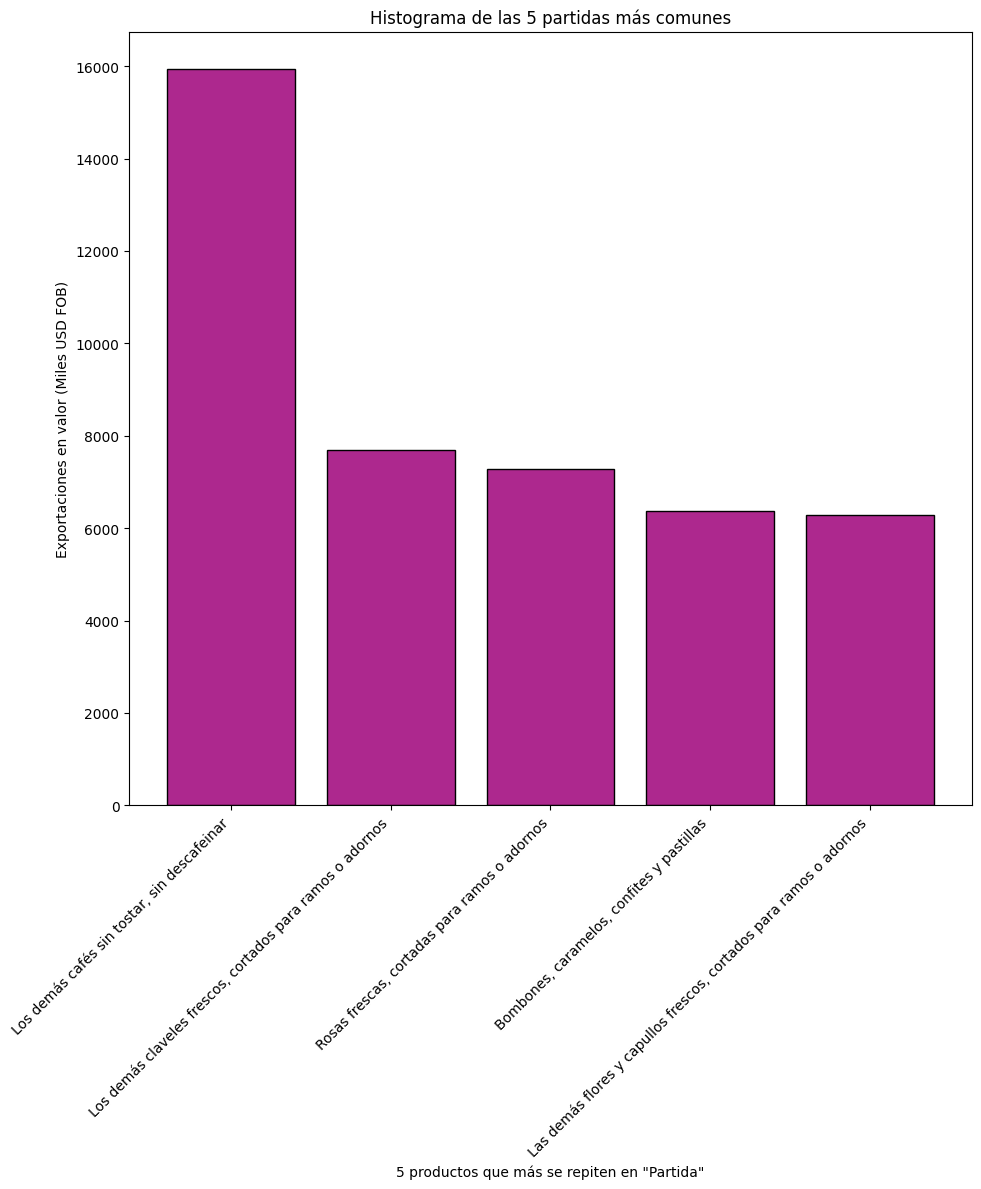
**5.1.2. Limpieza y Preprocesamiento**La visualización de la composición de la base de datos se realizó utilizando un DataFrame de pandas, en Python para manejar y analizar datos tabulares.  
Evidenciando los siguientes resultados:

**Datos No Nulos:** Cada columna tiene un conteo de valores no nulos igual a 475,785, lo que significa que no hay datos faltantes en ninguna de las columnas. Esto es importante porque asegura que todos los registros son completos y se pueden utilizar para análisis.

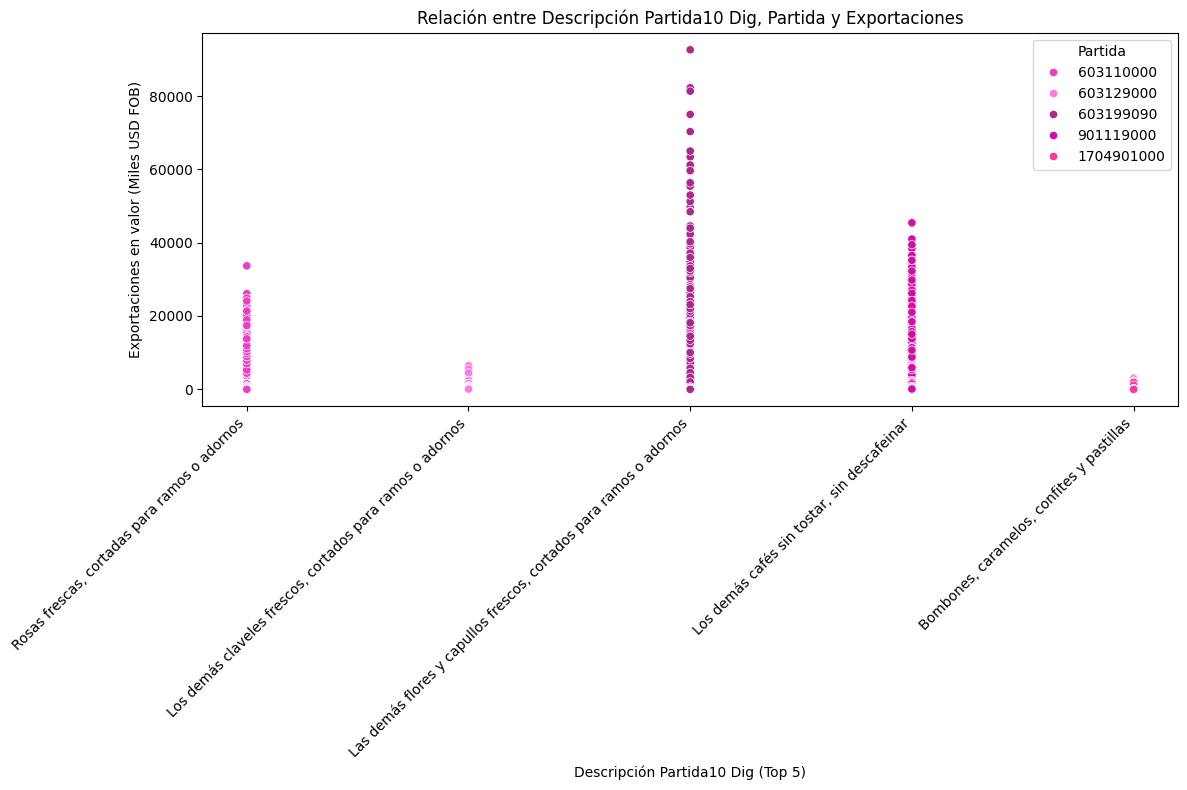
**Tipos de Datos:** Los tipos de datos incluyen enteros (int64) para las columnas numéricas y objetos (object) para las columnas que contienen texto. Esto es típico en bases de datos donde se manejan tanto datos cuantitativos como cualitativos.

#### 

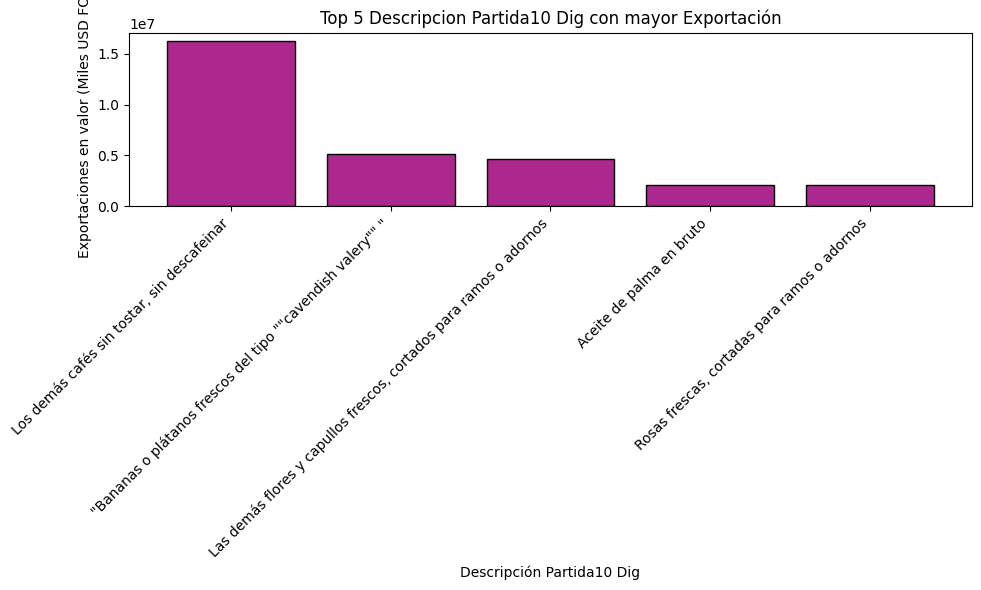
#### **5.1.3. Visualización:**

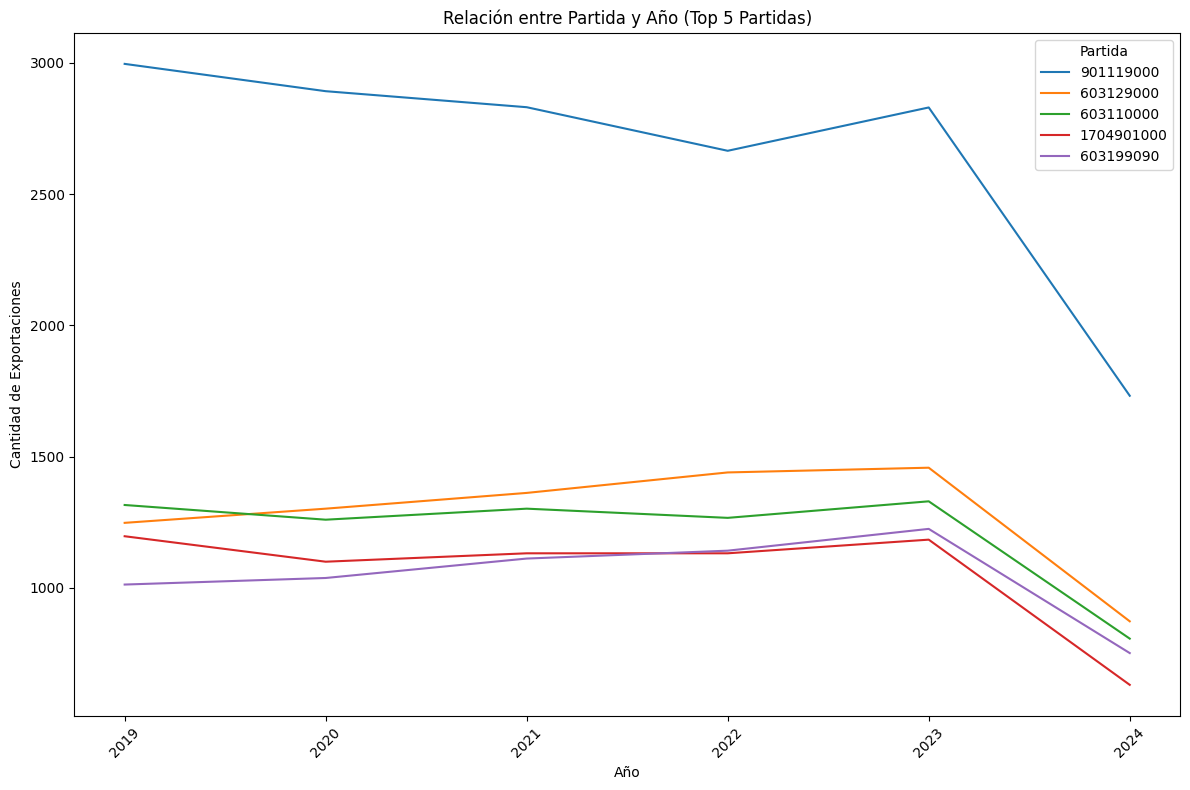
1. Histograma (Imagen 1) que muestra las cinco partidas más comunes en los datos de exportaciones. Cada barra representa una partida y su altura indica cuántas veces aparece en los registros. Esto permite identificar rápidamente cuáles son los productos más exportados.

2. Gráfico de dispersión que muestra la relación entre las cinco descripciones de partidas más comunes y las cinco partidas más comunes en términos de sus exportaciones. Cada punto en el gráfico representa una combinación de "Descripcion Partida10 Dig" y "Partida", y su posición indica el valor total de exportaciones (en Miles USD FOB) para esa combinación.



3. Gráfico de barras que muestra las 5 descripciones de partidas que tienen el mayor valor de exportación en términos de Valor en USD FPB y Volúmen". Cada barra representa una descripción y su altura indica el total de exportaciones para esa descripción.



4. Permite observar tendencias a lo largo de los años para los productos más exportados. Por ejemplo, puedes ver si la cantidad de exportaciones de una partida específica ha aumentado o disminuido con el tiempo, lo que puede ayudar a entender el comportamiento del mercado y las preferencias de exportación en esos años.

5. Histograma que permite observar la distribución de las exportaciones de las 10 descripciones más comunes en función de la región.

#### 

# 

### 

### 

### 

### 

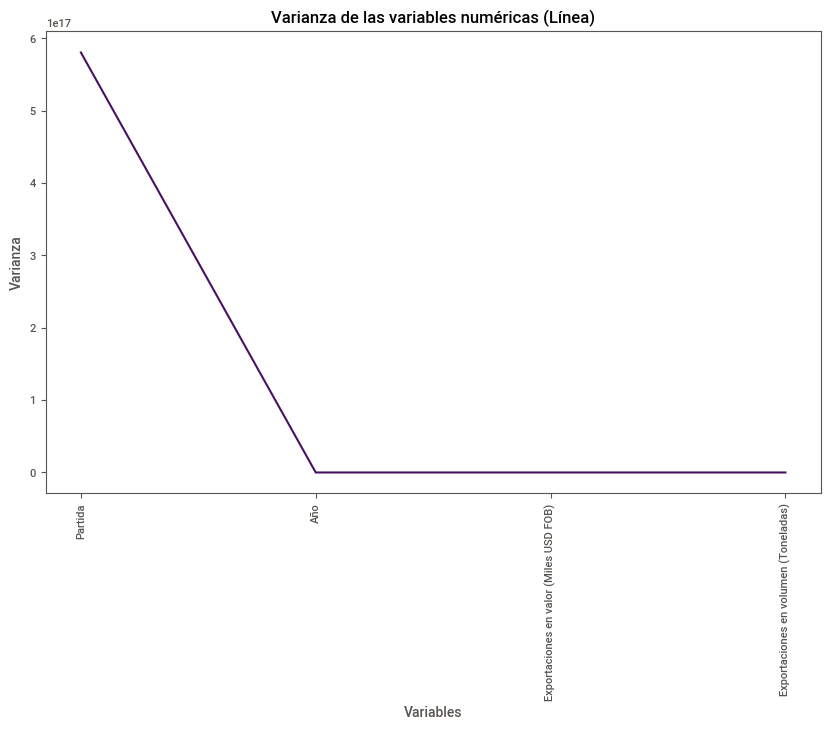
### 

### 

### **5.2. Modelo de Agrupación K-Means**

* **Selección de Variables:**

**1. Selección de Variables Numéricas de Mayor Varianza:**

Se decide seleccionar las columnas numéricas para luego proceder a visualizar de cómo están relacionadas las variables numéricas dentro de la base de datos, ayudando a identificar patrones y relaciones entre ellas. La varianza indica cuánto varían los datos. Una varianza alta significa que los valores están muy separados entre sí, mientras que una varianza baja indica que los valores son más similares.

La alta varianza de la variable "Partida" refleja una gran diversidad en la cantidad de productos, equivalente a 1180 valores únicos en esta columna.

La serie de números observados en la variable “Partida” se refieren a la segmentación los productos según el Sistema Armonizado (HS Code) en donde se tiene que:

Los Primeros 2 dígitos identifican la categoría más general, Capítulo.

Los Primeros 4 dígitos detallan subcategorías dentro del capítulo.

**Subpartida:** Primeros 6 dígitos, que especifican aún más el producto dentro de la partida.

**Partida Completa:** Código de 10 dígitos, proporcionando el nivel máximo de detalle.

La Segmentación por Capítulos se encuentra establecida de la siguiente manera:

**Sección 1:** animales y productos animales. Esta sección incluye los

capítulos del 1 al 5. En cada capítulo encontramos partidas y subpartidas, según la homogeneidad de cada grupo de productos.

**Sección 2:** productos vegetales, Capítulos del 6 al 14.

**Sección 3:** grasas y aceites animales o vegetales. Capítulo 15.

**Sección 4:** preparados alimenticios. Capítulos 16 al 24.

**Sección 5:** productos minerales. Capítulos del 25 al 27.

**Sección 6:** productos químicos y derivados. Capítulos 28 al 38.

**Sección 7:** plásticos y gomas. Capítulos 39 y 40.

**Sección 8:** cueros y pieles brutas y curtidas. Capítulos 41 al 43.

**Sección 9:** madera y productos de madera. Capítulos 44 al 46.

**Sección 10:** pasta de madera y materiales fibrosos. Capítulos del 47 al 49.

**Sección 11:** productos textiles. Capítulos del 50 al 63.

**Sección 12:** calzado y protección para la cabeza. Capítulos 64 a 67.

**Sección 13:** piedra y cristal. Capítulos 68 a 70.

**Sección 14:** perlas naturales y cultivadas. Capítulo 71.

**Sección 15**: metales comunes. Capítulos del 72 al 83.

**Sección 16:** maquinaria y productos eléctricos. Capítulos 84 y 85.

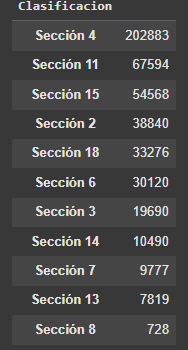
**Sección 17:** productos de transporte. Capítulos del 86 al 89.

**Sección 18:** instrumentos de precisión. Capítulos 90 a 92.

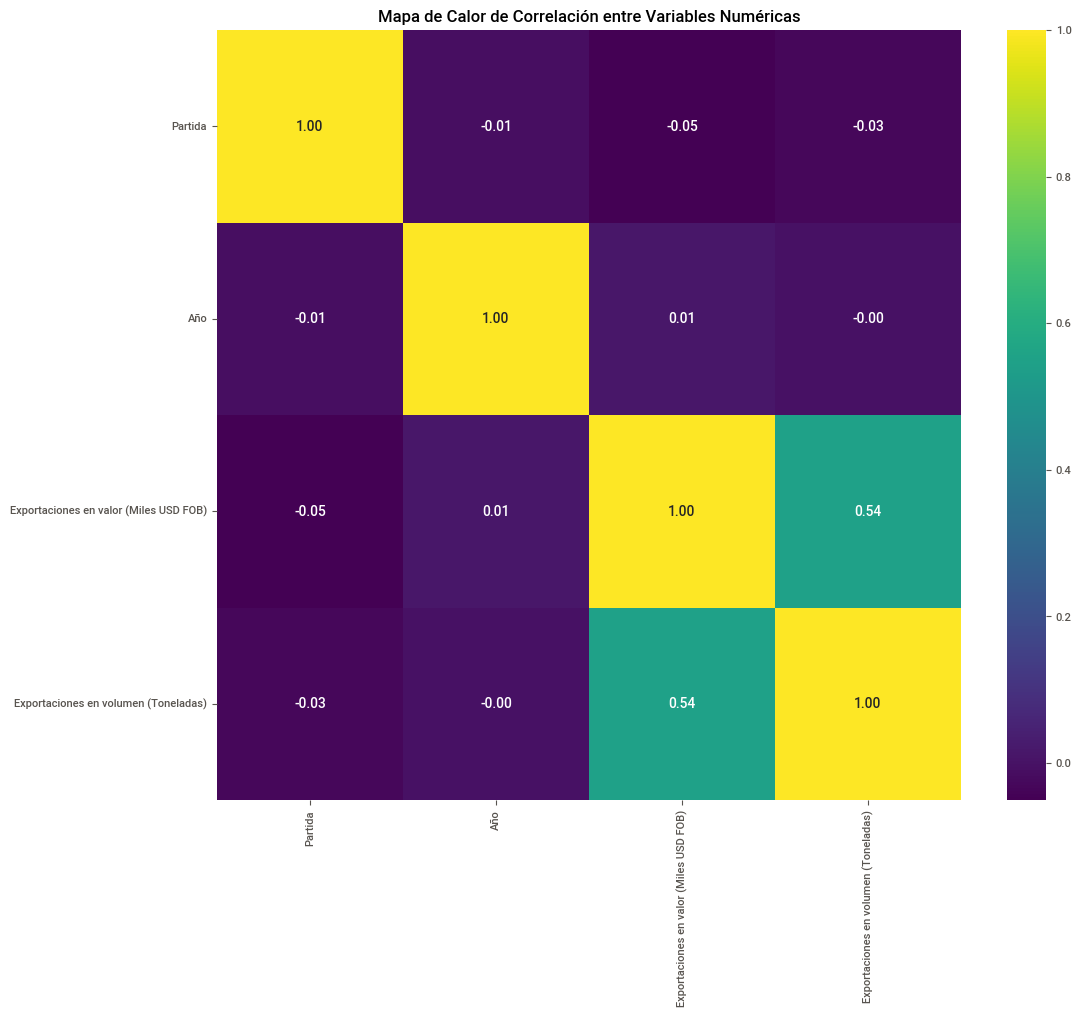
**Sección 19**: armas y munición. Capítulo 93.

**Sección 20:** artículos manufacturados variados. Capítulos del 94 al 96.

**Sección 21:** obras de arte. Capítulo 97.

Con la finalidad de simplificar la cantidad de datos que se tienen se procede a realizar una clasificación de las variables en la columna “Partida”:

1. Extrayendo los dos primeros dígitos de la serie de partida.
2. Se crea un diccionario llamado secciones, que contiene diferentes secciones y los rangos de capítulos que corresponden a cada sección. Por ejemplo, podría haber una sección para capítulos del 1 al 5, otra para capítulos del 6 al 10, etc
3. **Clasificación: Basándose** en el capítulo extraído, la función clasifica la partida en una de las secciones definidas.

* Para identificar las variables numéricas de la base de datos que tienen más relación entre ellas a partir de una matriz de correlación, se deben observar los valores más cercanos a 1 o -1 en el mapa de calor. 

En el caso de la Matriz realizada con las variables “Partida, Año, Exportaciones en valor (Miles USD FOB) y Exportaciones en volumen(Toneladas)” se evidencia una Correlación positiva (cerca de 1) entre “Exportaciones en valor (Miles USD FOB) y Exportaciones en volumen(Toneladas)” es decir, a medida que una variable aumenta, la otra también tiende a aumentar, lo cual sugiere que a mayor volúmen, mayor valor.

Codificación

La codificación de variables categóricas es un paso fundamental en el preprocesamiento de datos, especialmente en el contexto del aprendizaje automático. Este informe detalla el proceso de codificación de la columna 'Región' en el DataFrame Exportaciones\_df utilizando la técnica de One-Hot Encoding. Esta técnica permite convertir variables categóricas en variables numéricas, facilitando así su uso en algoritmos de machine learning.

* Razonamiento detrás de la elección de variables relevantes para la agrupación (ej., valor FOB, destino, tipo de producto).
* **Normalización de Datos:** Explicación de cómo se normalizan los datos para evitar que variables con diferentes escalas afecten el modelo.
* **Determinación del Número de Clústeres:** Métodos utilizados, como el codo y la silueta, para determinar el número óptimo de clústeres.
* **Entrenamiento del Modelo:** Procedimiento de entrenamiento y resultados iniciales de la agrupación.
* **Evaluación del Modelo:** Métricas de evaluación utilizadas (inercia, índice de silueta) para asegurar la calidad y coherencia de los clústeres.

### **5.3. Resultados y Análisis**

* **Clústeres Identificados**: Descripción de los clústeres generados por el modelo K-Means, incluyendo características comunes dentro de cada grupo.
* **Interpretación de los Resultados**: Análisis de patrones en cada clúster para identificar oportunidades estratégicas de exportación.
* **Visualización de Resultados**: Diagramas de dispersión y gráficos de radar para mostrar cómo se distribuyen los productos y mercados en cada clúster.

## **6. Alcance**

Este proyecto está diseñado para analizar y segmentar los datos de exportación de Colombia mediante el modelo de agrupación K-Means, aplicando esta técnica a nivel exploratorio y estratégico. El alcance del proyecto incluye:

Análisis Exploratorio de Datos (EDA): Se realizará una limpieza y preprocesamiento exhaustivo del conjunto de datos, incluyendo la selección de variables y la normalización de los datos, lo cual es esencial para el desempeño del modelo K-Means.

Aplicación de K-Means: Se entrenará el modelo de K-Means para identificar clústeres significativos de productos y mercados. La agrupación se hará en función de variables clave como el valor y volumen de exportación, país de destino, tipo de producto y año, proporcionando una segmentación útil para identificar oportunidades en cada clúster.

Análisis de Resultados y Recomendaciones Estratégicas: Una vez creados los clústeres, se interpretarán los resultados para entender los patrones característicos de cada grupo, como productos con mayor demanda en ciertos mercados y regiones, variabilidad de precios y similitudes en las tendencias de exportación. Con base en estos perfiles, se desarrollarán recomendaciones para optimizar la estrategia de exportación en función de los segmentos con mayor potencial.

El estudio se enfoca en utilizar datos históricos de exportación de Colombia para comprender mejor los patrones en los distintos segmentos, pero no se incluyen análisis específicos sobre el impacto de factores macroeconómicos o coyunturales fuera de los ya presentes en los datos. Tampoco se contempla la evaluación de datos en tiempo real; este proyecto se limita a un análisis retrospectivo con fines exploratorios y estratégicos. Sin embargo, los resultados y recomendaciones generados serán aplicables para el desarrollo de políticas comerciales y para orientar a los exportadores colombianos en la toma de decisiones informadas.

## **7. Presupuesto**

## 

## **9. Conclusiones**

**Resumen de Hallazgos:** Resumen de los principales hallazgos y cómo el modelo K-Means permite identificar patrones significativos en el comercio exterior colombiano.

**Impacto del Proyecto:** Descripción de cómo los resultados pueden contribuir a una mayor competitividad en las exportaciones colombianas.

**Limitaciones y Futuros Estudios:** Limitaciones en términos de datos y análisis, así como posibles áreas para futuras investigaciones, como la integración de datos macroeconómicos y el análisis en tiempo real.

## 8. Webgrafía:

https://analdex.org/2022/11/08/retos-y-oportunidades-lo-que-necesita-colombia-para-ser-una-potencia-exportadora/

<https://www.altosempresarios.com/noticias/exportacion-de-productos-colombianos-oportunidades-y-desafios>

Anexos.  
Imagen1



Imagen2

