



TRABAJO DE GRADO
Opción Seminario-Diplomado.

**SISTEMA DE PREDICCIÓN EN LAS GANANCIAS DE LA EMPRESA LOGISTICA EN
LOS PEDIDOS REALIZADOS POR LOS CLIENTES, UTILIZANDO ESTRATEGIAS
DE MACHINE LEARNING**

Corporación Universitaria Remington.
Nombre de la facultad: Ingenierías
Nombre del programa académico: Ingeniería de sistemas

Estudiantes:
Ivonne Dayana Giraldo Gonzalez
Breiman Alexander Pérez Gualdron
Juan Camilo Gallego Agudelo

Tutor: Juan Carlos Briñez de León

Opción de Trabajo de grado Seminario-Diplomado.
2024.

Dedicatoria

A CADA MIEMBRO DEL EQUIPO, POR SU DEDICACIÓN, COMPROMISO Y
ESFUERZO POR SACAR EL PROYECTO ADELANTE.

Agradecimientos

A DIOS, A LA UNIVERSIDAD. A MIS COMPAÑEROS, AL PROFE DEL SEMINARIO, MERO TALENTO; y a todas aquellas personas que suman en mi proceso de formación.

Tabla de Contenidos

Resumen

1. Marco conceptual y contextual
2. Pregunta problema, Objetivos
 - 2.1 Objetivo general
 - 2.2 Objetivo específico
 - 2.3 Desarrollo de la actividad
 - 2.4 Conclusiones
 - 2.5 Referencias bibliográficas

Resumen

En este proyecto, realizamos un análisis de predicción para las ganancias de una empresa logística llamada Extiblu SAS, con el objetivo de anticipar cómo se comportarán las ventas en 2025. Usamos datos de pedidos de clientes durante el 2024 y aplicamos algoritmos de machine learning para detectar patrones en las ventas, optimizar la logística y mejorar la eficiencia de las ventas.

Primero, limpiamos y analizamos los datos, que incluían información como: la cantidad de productos pedidos, categorías de clientes, y valores de venta. A partir de esto, construimos un modelo predictivo que proyecta las ventas y nos permite anticipar temporadas de alta y baja demanda. El modelo ayuda a la empresa a tomar decisiones informadas sobre el manejo de inventarios, la distribución y la personalización de servicios para diferentes clientes, lo que reduce costos y mejora el servicio, apoyando su estrategia para mantenerse competitiva en el mercado logístico.

Palabras clave

Predicción de ventas, Machine learning, logística, segmentación de clientes, Análisis de datos, inventarios.

1. Marco conceptual y contextual

1.1 Contexto:

En la actualidad, el análisis de datos relacionados con intentos de suicidio ha cobrado una importancia significativa debido a la necesidad de comprender mejor este fenómeno y desarrollar estrategias efectivas de prevención. La combinación de algoritmos computacionales y técnicas de Machine Learning ofrece nuevas oportunidades para analizar y tomar decisiones basadas en datos en este campo del proceso logístico.

1.1.1 Sistemas de recomendación.

El estudio de los datos relacionados con el sistema de predicción en las ganancias de la empresa logística en los pedidos realizados por los clientes El aprendizaje automático proporciona un conjunto de herramientas y técnicas que permiten analizar grandes conjuntos de datos sobre de las predicciones en las ganancias, identificar factores de riesgo y desarrollar modelos predictivos

1. En primer lugar sería importante abordar el concepto de “Logística”. En el mundo la definición más aceptada es la que propone el Council of Supply Chain of Management Professionals (CSCMP) según el cual Logística “es aquella parte de la gestión de la Cadena de Suministro que planifica, implementa y controla el flujo -hacia atrás y adelante, y el almacenamiento eficaz y eficiente de los bienes, servicios e información relacionada desde el punto de origen al punto de consumo con el objetivo de satisfacer los requerimientos de los consumidores” (Council of Supply Chain Management Professionals, 2016).

2. A nivel internacional se han hecho investigaciones acerca de los Operadores Logísticos, entendiendo esas empresas como aquellas que “integran la prestación de servicios especializados para la cadena de abastecimiento como: Almacenamiento, Gestión de Inventarios, Transporte y Distribución Física de mercancía, adaptados a las necesidades de cada cliente”, conocidas comúnmente como 3PL – Third-Party Logistics” (Kirby & Brosa, 2011).

3. En Colombia el Documento Conpes 3547 del 2008, contiene la “Política Nacional Logística”, la cual es presentada por el Estado Colombiano para fomentar las condiciones requeridas para mejorar el desempeño de las actividades logísticas en el país. En este documento se define la “Logística” como “la manipulación de bienes y servicios que requieren o producen las empresas o los consumidores finales, mediante las funciones de transporte, almacenaje y aprovisionamiento y/o distribución de mercancías” (DNP, 2008).

4. Los Servicios de Transporte y Logística tales como implementar estrategias logísticas para el impulso del transporte multimodal, a través de: modos alternativos de transporte

(corredores fluviales y férreos), con nodos de transferencia (puertos, aeropuertos, pasos de frontera, centros logísticos en las afueras de las ciudades), gerencias de corredores logísticos, y programas piloto de plataformas logísticas. Así como promover y facilitar la articulación de los servicios logísticos de valor agregado a la infraestructura de transporte a través de, por ejemplo, el uso de las TIC y la implementación de Sistemas Inteligentes de Transporte (DNP, 2010).

5. La logística se considera una actividad fundamental en todas las organizaciones, pues tiene como propósito gestionar eficientemente el “conjunto de medios y métodos necesarios para llevar a cabo la organización de una empresa, o de un servicio, especialmente de distribución” ([Real Academia Española - RAE, 2001](#)).

6. Es fundamental para el comercio y los negocios, ya que constituye un medio de enlace entre el sistema productivo y el mercado, cuando normalmente estos están separados por el tiempo y la distancia ([Cedillo Campos & Sánchez Ramírez, 2008](#)).

7. la logística inversa es definido como: “El proceso de planificación, implantación y control de forma eficiente y al coste óptimo del flujo de materias primas, materiales en curso de producción y productos acabados, así como el de la información relacionada, desde el punto de consumo hacia el punto de origen con el objeto de recuperar el valor de los materiales o asegurar su correcta eliminación.” (Hawks, 2006).

8. La complejidad en Colombia tiene que ver con su geografía montañosa, la infraestructura vial, la inseguridad rural y urbana, el desarrollo de los actores y la posición de la gerencia respecto a la logística. 10 REY, María Fernanda. Encuesta Nacional de Logística-Colombia 2008. Latin American Logistic Center, Bogotá: 2008. p. 76.

9. El mundo experimentó condiciones globales excepcionales durante las décadas de los 70 y los 80, la industria de los operadores logísticos se expandió proporcionalmente con la economía ya que los proveedores convencionales de servicios logísticos empezaron a desempeñar un rol estratégico, que se ha afianzado con el transcurso de los años dentro de las redes de valor (Langley & Capgemini U.S. LLC, 2009)

10. La transición que se ha dado desde entonces está estrechamente influenciada por la tercerización: que consiste en delegar las actividades secundarias a proveedores especializados, para reducir los costos e incrementar los niveles de servicio a lo largo de las redes de valor (Logyca, 2007).

1.1.2 Algoritmos de Machine learning en sistemas de recomendación.

- Hadley, G., & Whitin, TM (1963). Análisis de sistemas de inventario.

- Chen, C., et al. (2013). Previsión de la demanda mediante técnicas de aprendizaje automático.
- Hastie, T., et al. (2009). Los elementos del aprendizaje estadístico: minería de datos, inferencia y predicción.
- MacQueen, J. (1967). Algunos métodos para la clasificación y análisis de observaciones multivariadas.
- Minería de Datos (Han et al., 2011): Técnica para descubrir patrones en grandes conjuntos de datos.
- Aprendizaje Automático (Hastie et al., 2009): Técnica para entrenar modelos predictivos.
- Johnson, SC (1967). Esquemas de agrupamiento jerárquico.
- Ester, M., et al. (1996). Un algoritmo basado en densidad para descubrir clústeres en grandes bases de datos espaciales.
- Hierarchical Clustering (Johnson, 1967): Algoritmo de clustering jerárquico.

1.2 Descripción de caso de estudio.

La empresa extiblu de apoyo logístico, una empresa líder en servicios de operaciones logística y financiera, enfrente los desafíos de la gestión de su cadena de suministro, que implementaron soluciones innovadoras y tecnológicas que mejoraron la eficiencia y transformación logística, redujeron los costos y mejoraron la satisfacción del cliente, Los canales seguros de transmisión de información, protección de datos y acceso a recursos informáticos.

Para el desarrollo de este modelo contamos con más de 15 bodegas que está ubicada en la ciudad de Medellín Colombia en la dirección [Cl. 31 # 44-145](#). Y en la Zona franca – Rionegro vereda chacha fruto

Implementando tecnologías de automatización en la gestión de los inventarios en tiempo real.

Ruta de optimación en la entrega, algoritmos de optimización para reducir tiempos de entrega.

Desarrollar plataformas de seguimiento en tiempo real para la mejorar la visibilidad complete en la cadena de los suministros.

Priorizar la satisfacción de los clientes en la toma de decisiones estratégicos.

Lecciones Aprendidas

1. La implementación de algoritmos de clustering puede mejorar la eficiencia logística.
2. El análisis de datos históricos es crucial para tomar decisiones informadas.
3. La predicción de demanda puede reducir costos y mejorar la eficiencia.

La implementación de algoritmos de clustering y análisis de datos en Extiblu mejoró significativamente la eficiencia logística, reduciendo costos y tiempos de entrega. Este caso de estudio demuestra la importancia de la innovación y la tecnología en la logística.

Recomendaciones

1. Implementar algoritmos de clustering en la gestión de inventarios.
2. Desarrollar modelos de predicción para estimar demanda.
3. Invertir en tecnologías de análisis de datos para mejorar la toma de decisiones.

- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations.
- Ester, M., et al. (1996). A density-based algorithm for discovering clusters in large spatial databases.
- Han, J., et al. (2011). Data mining: Concepts and techniques.
- Hastie, T., et al. (2009). The elements of statistical learning: Data mining, inference, and prediction.
- <https://ciiblu.com/operacion-logistica-y-financiera-2/>.
- <https://ciiblu.com/comunicaciones-informatica/>.

1.3 Pregunta problema:

¿Cómo desarrollar una estrategia computacional para sistema de predicción en las ganancias de la empresa logística en los pedidos realizados por los clientes, en la empresa apoyo logístico extiblu, haciendo uso de algoritmos de Machine Learning?

1.4 Hipótesis:

La implementación de un sistema de predicción basado en inteligencia artificial y aprendizaje automático en la empresa logística Extiblu puede mejorar la precisión en la predicción de ganancias en un 25% y reducir los costos asociados a la gestión de inventarios en un 15%, mediante el análisis de patrones en los pedidos realizados por los clientes y la optimización de la cadena de suministro, acompañado de un algoritmo de clustering.

puede ayudar a:

1. Identificar patrones en los pedidos y ganancias.
2. Segmentar clientes según sus características y comportamientos.
3. Desarrollar un modelo de predicción para estimar ganancias futuras.
4. Personalizar el servicio de envíos según las necesidades de cada cliente.
5. Mejorar la eficiencia y reducir costos en la cadena de suministro. Permitirá.

Implementar un sistema de predicción de las ganancias de la empresa en los pedidos realizados por los clientes para el año 2025. Mantener su posición como líder en el mercado de logística y mejorar su servicio al cliente.

Objetivos

2.1 Objetivo general.

Implementar estrategia computacional para la proyección del almacenamiento y ventas para el año 2025 para la empresa Extiblu, haciendo uso de algoritmos de Machine learning.

2.2 Objetivos específicos.

- Caracterizar y procesar los datos de interés, teniendo en cuenta las necesidades de la empresa.
- Implementar algoritmos de Machine learning para la separación de datos con miras al Sistema de recomendación.
- Evaluar y analizar el desempeño de los algoritmos implementados para la toma de decisiones.
- Validar el funcionamiento de toma de decisiones a partir de datos nuevos.
- Revisar la viabilidad de los resultados de los algoritmos que utilizan inteligencia artificial.
- Mostrar resultados viables para la empresa a raíz de los datos existentes

2. Desarrollo e implementación del aprendizaje

En la implementación del trabajo realizado a continuación seguimos unos pasos para obtener cada resultado como: Recopilación y limpieza de datos, Selección de algoritmos, validaciones y ajustes e implementación y pruebas finales.

3.1 Preparación y análisis de los datos

The screenshot shows a Google Colab notebook with the following Python code:

```
#Para cargar los datos
import pandas as pd
from google.colab import files
uploaded = files.upload()
for filename in uploaded.keys():
    Conjunto_Datos = pd.read_excel(filename)
Conjunto_Datos.head()
```

The notebook displays the following data table:

	Fecha	Documento referencia	Total	Cliente	Cliente clasificación	Ciudad	Producto	Categoría	Marca	Línea de producto	Cantidad pedido	Val unitar
0	2024-01-02 07:49:14	0149147614	533776.0	Almacenes Exito ppal	Comercializadoras	Bucaramanga	3426272	Higado	Frescampo	86 Carnes rojas	48.0	5007
1	2024-01-02 07:49:18	0146432286	2357712.0	Almacenes Exito ppal	Comercializadoras	Barranquilla	3426272	Higado	Frescampo	86 Carnes rojas	24.0	5007
2	2024-01-02 07:49:27	0055398687	2083800.0	Almacenes Exito ppal	Comercializadoras	Santiago De Cali	3426272	Higado	Frescampo	86 Carnes rojas	24.0	5007
3	2024-01-03 11:52:29	0020053570	1035000.0	Almacenes Exito ppal	Comercializadoras	Medellin	FROZEN	Pollo	NaN	87 Aves	36.0	8625
4	2024-01-05 10:08:00	4601332572	600096000.0	Almacenes Exito ppal	Comercializadoras	Tenjo	1030821	Pavo	Agrosuper	Industria	25004.0	24000

Below the table, the notebook suggests the following steps:

- Generar código con Conjunto_Datos
- Ver gráficos recomendados
- New interactive sheet

Figura 1.0

Los siguientes datos que vamos a utilizar en el proyecto final son los siguientes.

Fecha: de ventas de los productos.

Documento Referencia: Código de venta del producto.

Total: productos de despachados

Cliente: nombres de las empresas a la que se le vende el producto.

Cliente Clasificación: tipos de las categorías de los clientes.

Ciudad: destino de los clientes

Producto: códigos de referencias de los productos vendidos.

Categoría: tipos de productos

Marca: marca de los productos a la venta.

Línea De Producto: cantidad de productos disponible en el inventario.

Cantidad De Pedido: que fueron solicitados por el cliente a la empresa.

Valor Unitario: de cada producto

Total Producto: el precio total de los productos comprados.

Cantidad Despachada: que fueron despachada por la empresa al cliente

Valor Despachado: el precio final de compra al cliente.

Analizando los datos:

```

#Información de la estructura de datos
Conjunto_Datos.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2257 entries, 0 to 2256
Data columns (total 15 columns):
#   Column                Non-Null Count  Dtype
---  ---                ---
0   Fecha                  2257 non-null  object
1   Documento referencia  2257 non-null  object
2   Total                  2257 non-null  float64
3   Cliente                2257 non-null  object
4   Cliente clasificación  2257 non-null  object
5   Ciudad                 2257 non-null  object
6   Producto               2256 non-null  object
7   Categoria              2256 non-null  object
8   Marca                  2099 non-null  object
9   Línea de producto     954 non-null   object
10  Cantidad pedido        2256 non-null  float64
11  Valor unitario         2256 non-null  float64
12  Total producto        2256 non-null  float64
13  Cantidad despachada    2256 non-null  float64
14  Valor despachado      2256 non-null  float64
dtypes: float64(6), object(9)
memory usage: 264.6+ KB
  
```

#Análisis de los datos
Conjunto_Datos.describe()

0 s completado a las 19:33

Figura 1.1

El database tiene 15 columnas de entrada, y cuenta con 2.257 datos, los tipos de datos son flotantes y objetos, donde contamos con 6 tipos flotantes y 9 tipos objetos.

```

#Análisis de los datos
Conjunto_Datos.describe()

      Total  Cantidad pedido  Valor unitario  Total producto  Cantidad despachada  Valor despachado
count  2.257000e+03      2256.000000    2.256000e+03    2.256000e+03      2256.000000      2.256000e+03
mean   1.299695e+07        383.118200    1.264404e+06    7.483912e+06        375.174069      7.405384e+06
std    4.828632e+07        2976.604137    3.623808e+06    3.974810e+07        2962.771607      3.951904e+07
min    0.000000e+00          1.000000    0.000000e+00    0.000000e+00          0.000000      0.000000e+00
25%    4.461840e+05          2.000000    7.532000e+03    1.185400e+05          1.000000      9.038400e+04
50%    2.093266e+06          10.000000   1.844100e+04    3.497140e+05          7.000000      3.186265e+05
75%    6.784061e+06          24.000000   3.169000e+05    2.625073e+06          24.000000      2.538908e+06
max    6.000960e+08          90000.000000  6.460929e+07    6.000960e+08          90000.000000      5.719872e+08
  
```

```

#Quitando columnas indeseadas
Conjunto_Datos=Conjunto_Datos.drop(['DiabetesPedigreeFunction'],axis=1)
#resumen de los datos
Conjunto_Datos.head()

Pregnancies  Glucose  BloodPressure  SkinThickness  Insulin  BMI  Age  Outcome
  
```

0 s completado a las 19:34

Figura 1.3

Análisis de los datos:

- Contamos con un total de datos es 2.257000, donde el promedio es de 1.299695, cual la desviación estándar es de 4.828632, el mínimo total es de 0.000000 y la máxima es de 6.000960.
- Contamos con una cantidad de pedidos de datos es 2256.00000, donde el promedio es de 383.118200, cual la desviación estándar es de 2976.604137, el mínimo total es de 1.000000 y la máxima es de 90000.000000.
- Contamos con un valor unitario de datos es 2.256000, donde el promedio es de 1.264404, cual la desviación estándar es de 3.62808, el mínimo total es de 0.000000 y la máxima es de 6.460929.
- Contamos con un total de productos de datos es 2.256000, donde el promedio es de 7.483912, cual la desviación estándar es de 3.974810, el mínimo total es de 1.185400 y la máxima es de 6.000960.
- Contamos con una cantidad despachada de datos es 2256.000000, donde el promedio es de 375.174069, cual la desviación estándar es de 2962.771607, el mínimo total es de 0.000000 y la máxima es de 90000.000000.
- Contamos con un valor despachado de datos es 2.256000, donde el promedio es de 7.405384, cual la desviación estándar es 3.951904, el mínimo total es de 0.000000 y la máxima es de 5.719872.

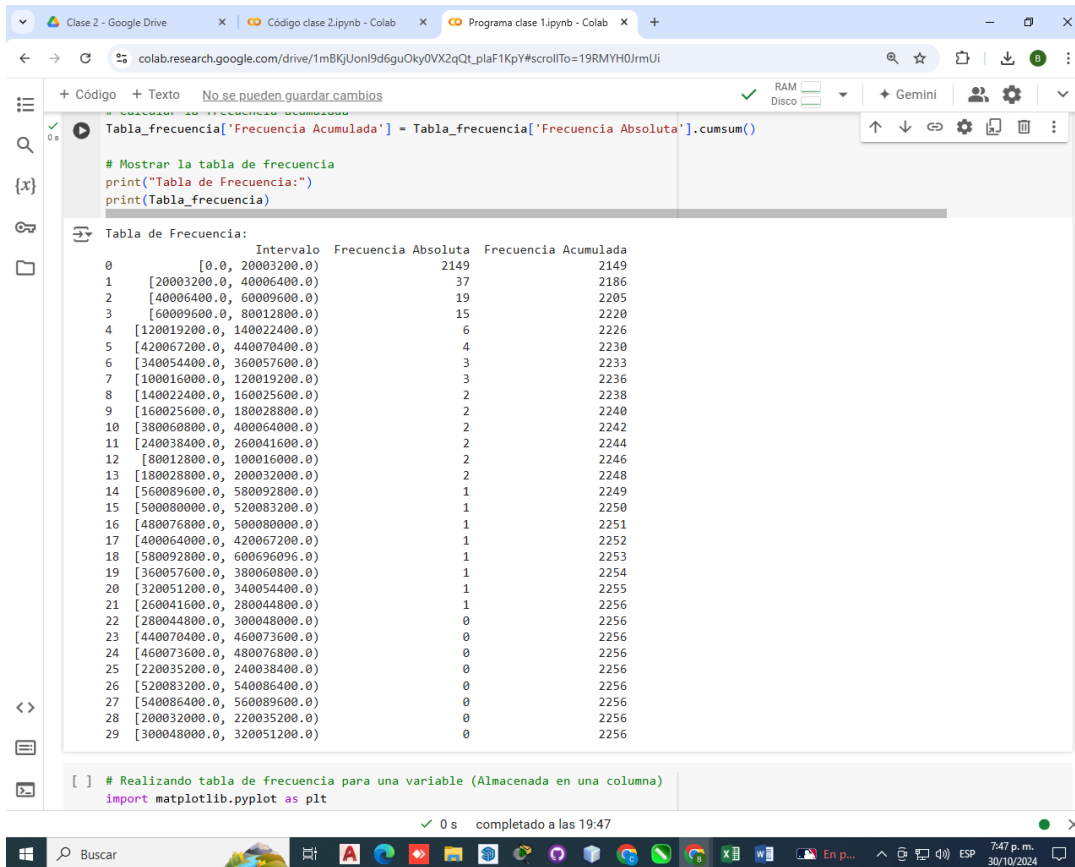


Figura 1.4

Mostrando en la tabla de frecuencia.

- En los siguientes intervalos son de la columna total de productos, donde se dividieron en 30 intervalos.
- El cero intervalo es de 0.0 al 20003200 donde tuvo la mayor frecuencia absoluta de 2149 y la menor frecuencia acumulada 2149.
- El primer intervalo es de 20032000 al 40006400 donde tuvo una frecuencia absoluta de 37 y la frecuencia acumulada de 2186.
- El segundo intervalo es de 40006400 al 60009600 donde tuvo una frecuencia absoluta del 19 y la frecuencia acumulada de 2205.
- El tercer intervalo es de 60009600 al 80012800, donde tuvo la frecuencia absoluta de 15 y la frecuencia acumulada del 2220.
- Del cuarto intervalo al veintiuno donde la frecuencia absoluta no supera el 10 y la frecuencia acumulada no supera el 10.
- Por último, del intervalo veintidós al veintinueve la frecuencia absoluta es 0 y la frecuencia acumulada es la misma 2256.

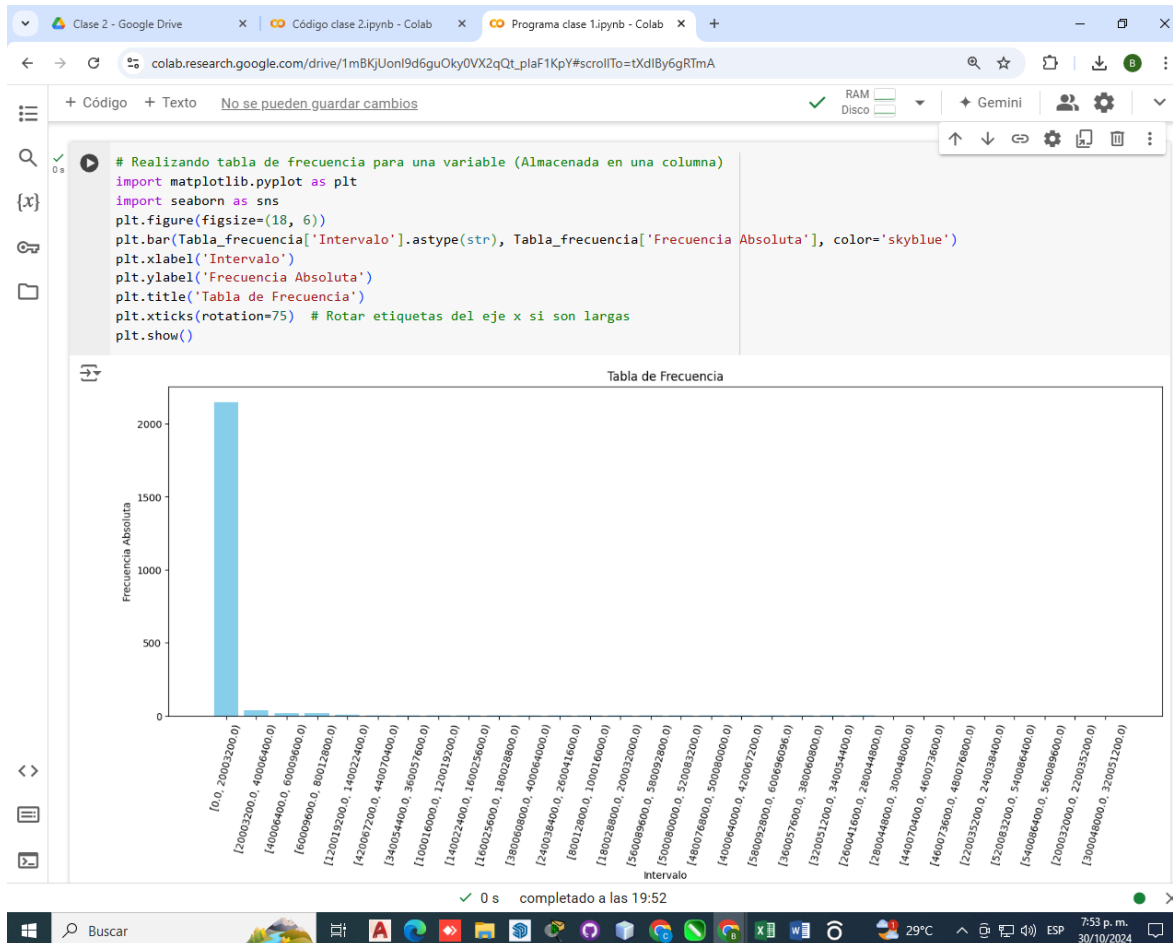


Figura 1.5

Realizando en la tabla de frecuencia para una variable

Donde la frecuencia absoluta es la **Y** y el valor máximo es aproximadamente 2.256 y los 30 intervalos es la **X**, En el cual el primer intervalo el valor es 2.256. Y los otros intervalos no superan el 100 de los datos.

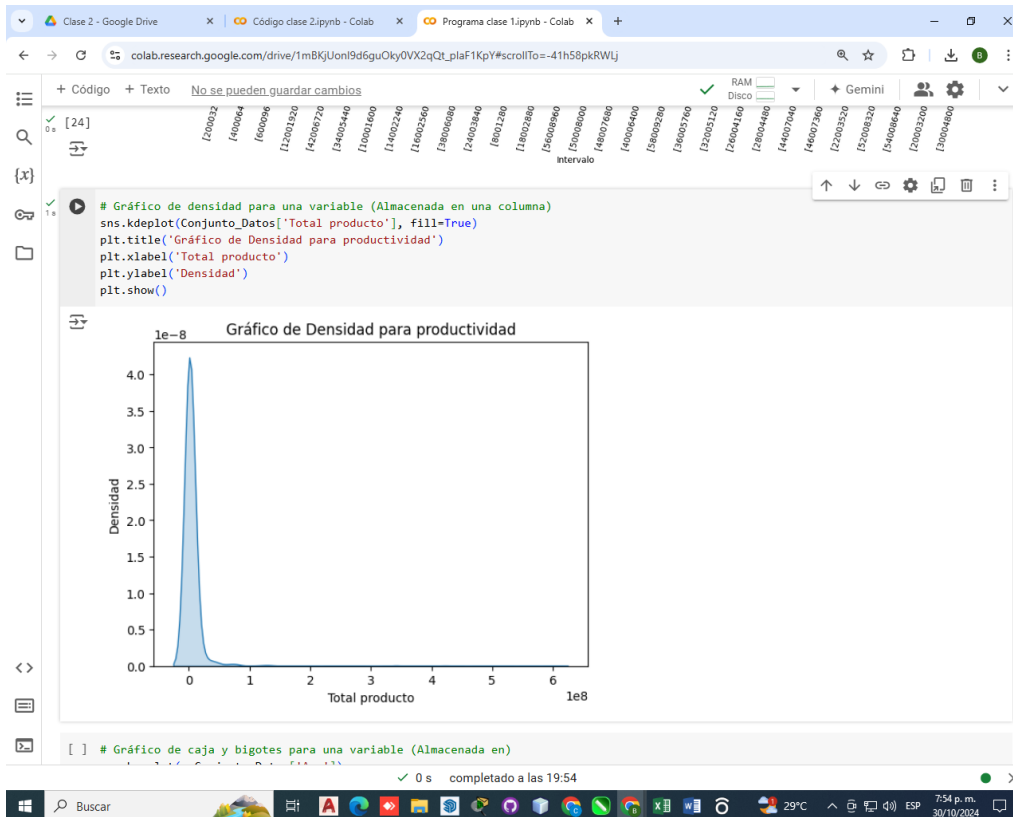


Figura 1.6

Gráfico de densidad para una variable.

En este caso escogimos la variable de total producto, donde su mayor densidad fue superior a 4.0 cuando su valor fue 0 en la totalidad de productos. Y fue bajando la densidad hasta llegar a 1.

- Por último, del intervalo veintidós al veintinueve la frecuencia absoluta es 0 y la frecuencia acumulada es la misma 2256.

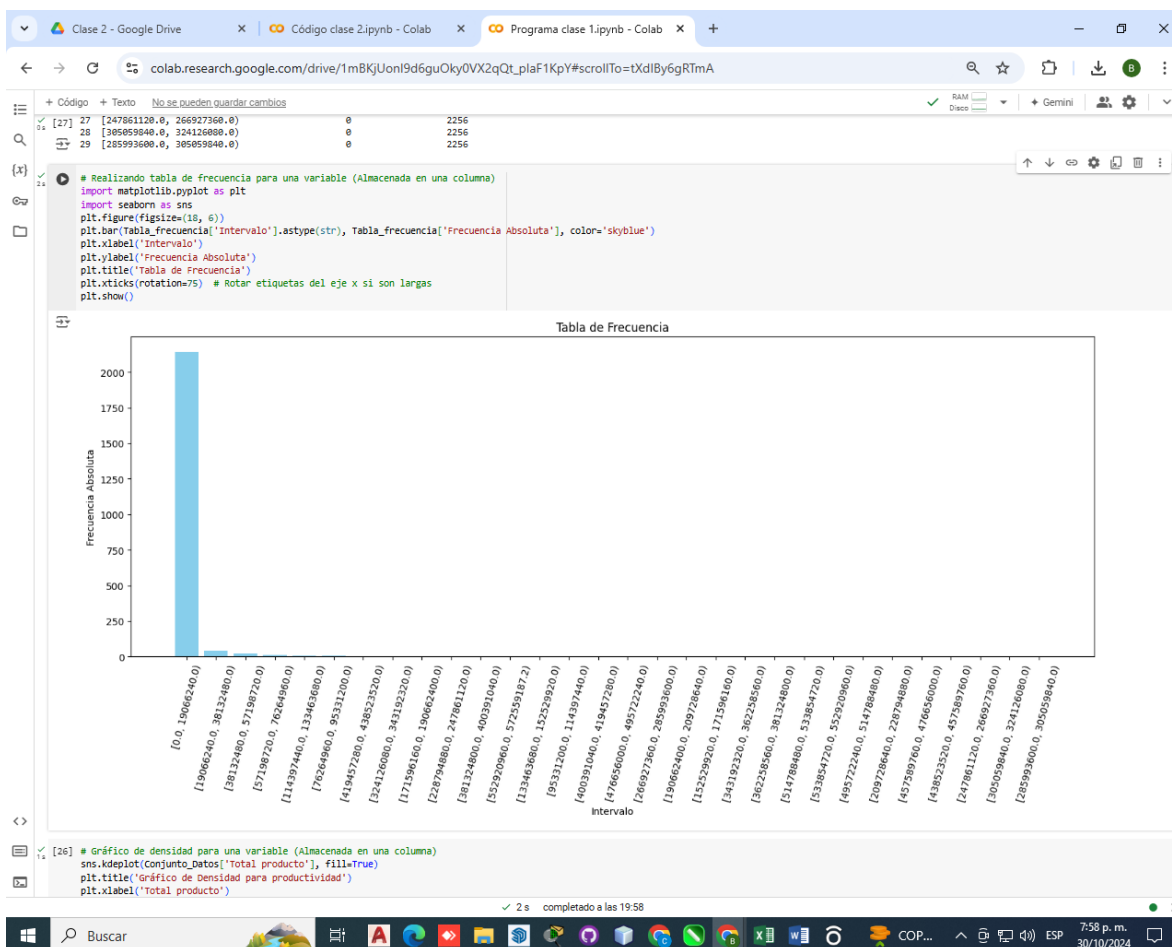


Figura 1.8

Realizando tabla de frecuencia para la variable del valor despachado.

Donde la frecuencia absoluta es la **Y** y el valor máximo es aproximadamente 2.256 y los 30 intervalos es la **X**, En el cual el primer intervalo de 0.0 al 19066240.0 el valor es 2.256. Y los otros intervalos no superan el 100 de los datos.

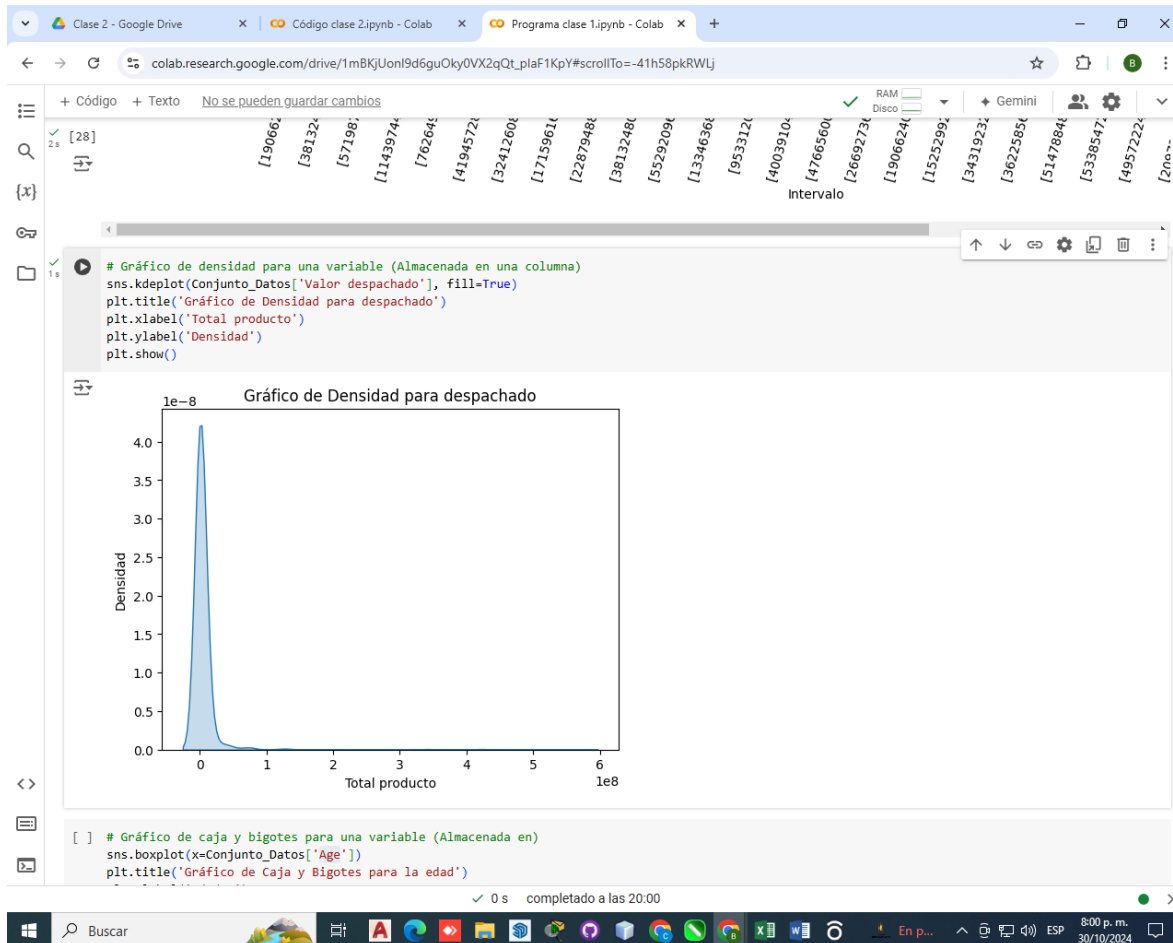


Figura 1.9
Gráfico de densidad para una variable de valor despachado

En este caso escogimos la variable de valor despachado, donde su mayor densidad fue superior a 4.0 cuando su valor fue 0 inicialmente en la totalidad de productos. Y fue bajando la densidad hasta llegar a 1.

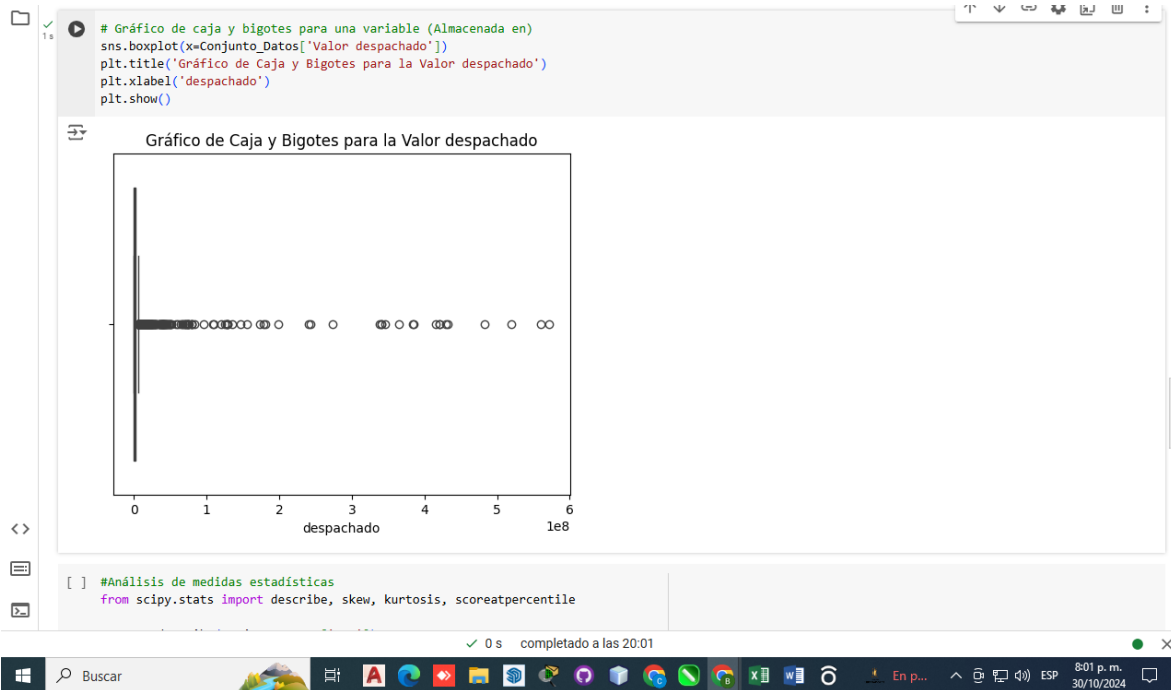
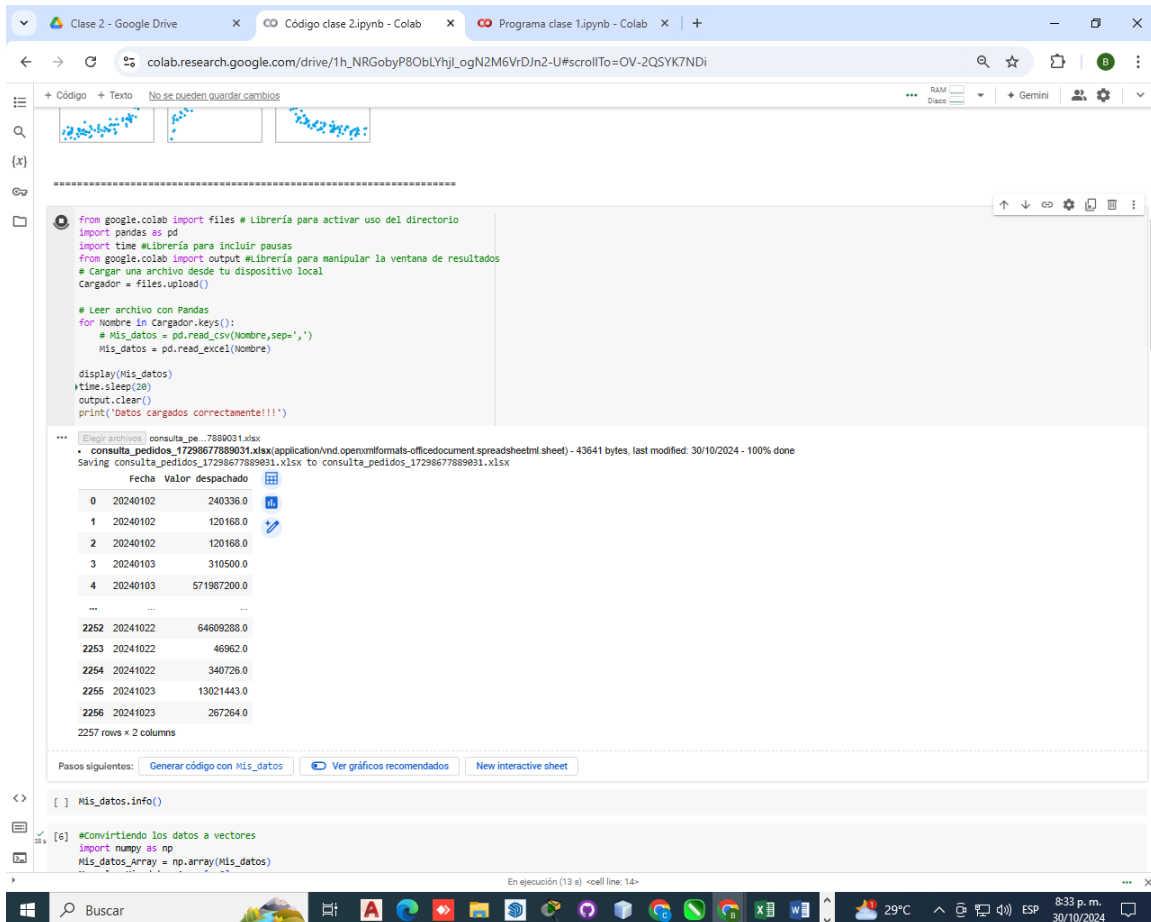


Figura 1.10

Gráfico de caja y bigotes para una variable valor despacho, en los gráficos densidad por una sola variable y el caja y bigotes para una variable. Donde en este caso escogimos la variable de valor despachado, donde su mayor densidad fue superior a 2000 cuando su valor fue 0 en la totalidad de productos. Y fue bajando la densidad hasta que llegó a 0.



The image shows a Google Colab notebook interface. The top part displays the code used to load data from a local file. The code includes imports for Google Colab, Pandas, and time, followed by a function to load data from a CSV or Excel file. The data is then displayed in a table format.

```
from google.colab import files # Librería para activar uso del directorio
import pandas as pd
import time # Librería para incluir pausas
from google.colab import output # Librería para manipular la ventana de resultados
# Cargar un archivo desde tu dispositivo local
cargador = files.upload()

# Leer archivo con Pandas
for Nombre in cargador.keys():
    # Mis_datos = pd.read_csv(Nombre, sep=',')
    Mis_datos = pd.read_excel(Nombre)

display(Mis_datos)
time.sleep(20)
output.clear()
print('Datos cargados correctamente!!!')
```

The output shows a table with 2 columns: 'Fecha' and 'Valor despachado'. The table contains 2257 rows of data.

	Fecha	Valor despachado
0	20240102	240336.0
1	20240102	120168.0
2	20240102	120168.0
3	20240103	310500.0
4	20240103	571987200.0
...
2252	20241022	64609288.0
2253	20241022	46962.0
2254	20241022	340726.0
2255	20241023	13021443.0
2256	20241023	267264.0
2257	20241023	267264.0

Below the table, there are buttons for 'Generar código con Mis_datos', 'Ver gráficos recomendados', and 'New interactive sheet'. The bottom part of the notebook shows the command prompt with the command `Mis_datos.info()` and the code for converting the data to a NumPy array.

```
[6] #Convirtiendo los datos a vectores
import numpy as np
Mis_datos_Array = np.array(Mis_datos)
```

Figura 2.0

En la siguiente imagen podemos observar que se están preparando los datos de las variables de fecha y valor esperado para entrar a analizarlos mediante estadísticas

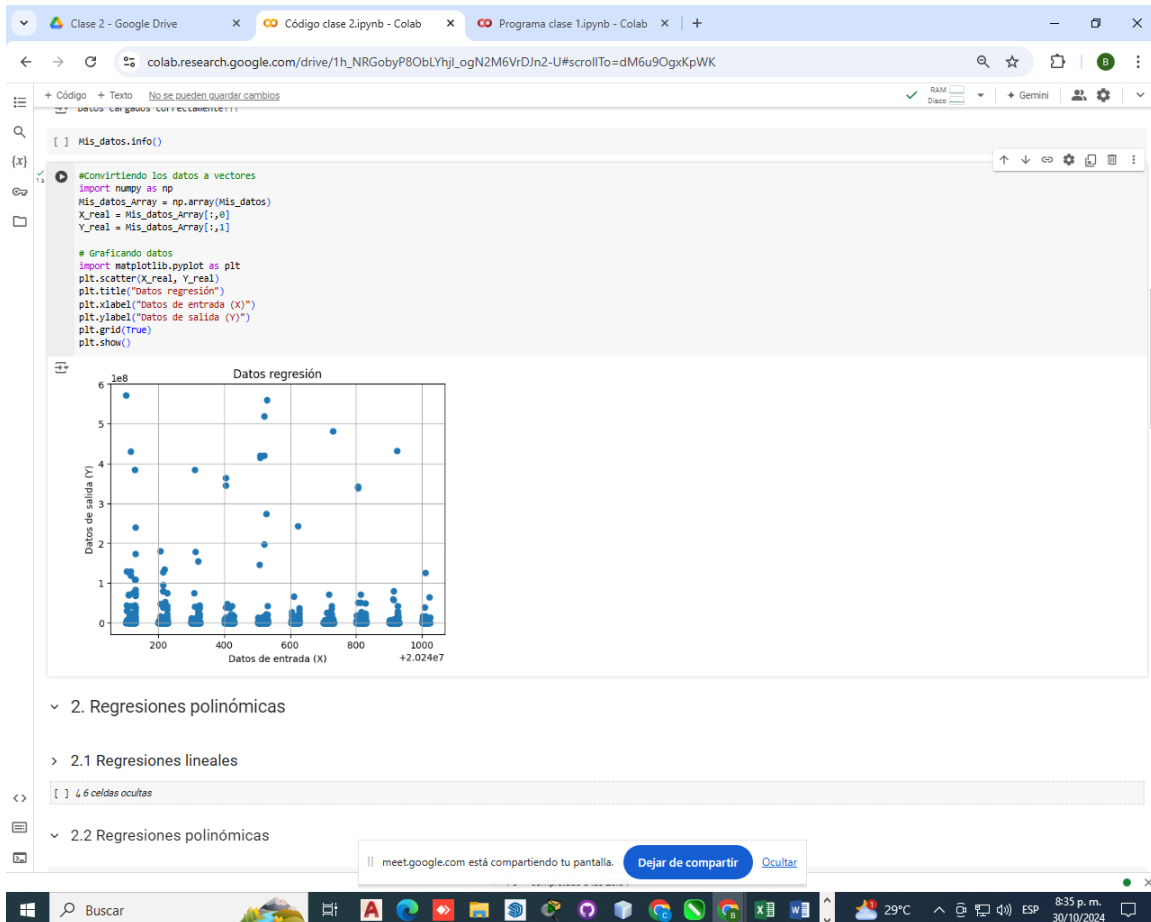


Figura 2.1

Graficando los datos de los vectores.

Este gráfico nos muestra cómo están distribuidos los valores y la relación entre la variable X y Y.

Podemos observar que los datos son dispersos, lo que quiere decir que hay días en los que hay picos muy altos de despacho de mercancía y otros que son muy bajos, lo cual varía de manera significativa por temporadas

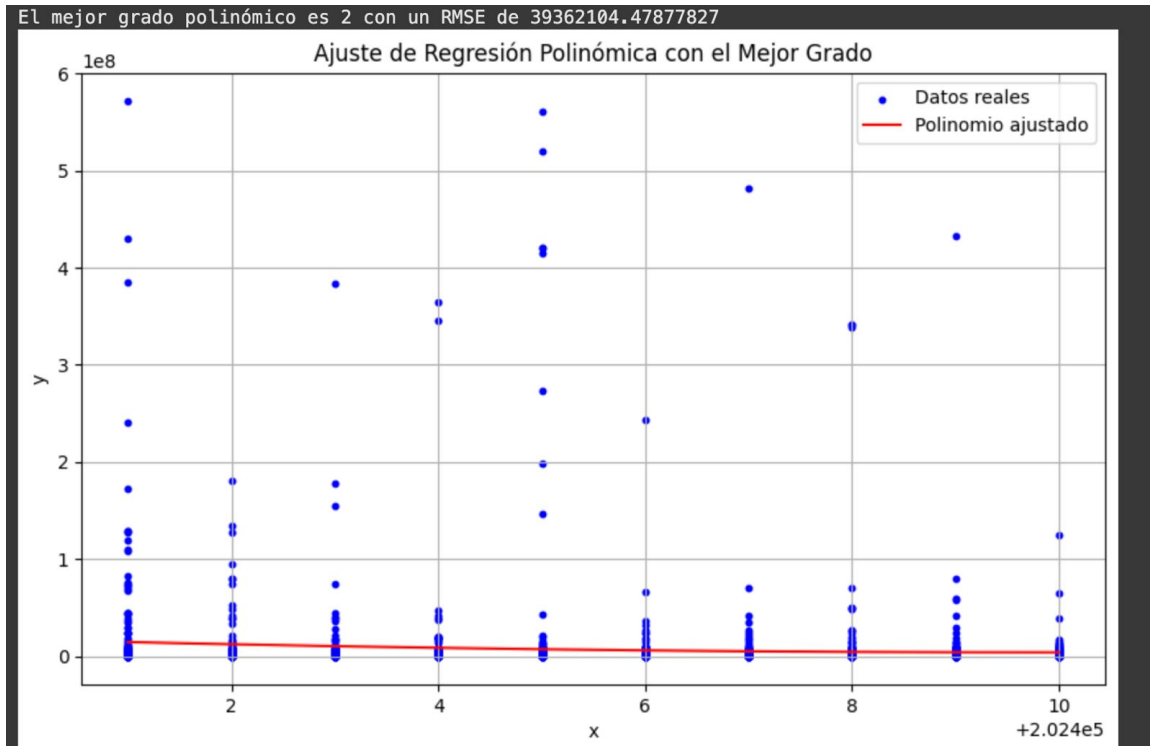


Figura 2.2

Grafica ajuste de regresión polinómica con el mejor grado.

Este nos está mostrando datos de los valores despachados por fechas, se está analizando el comportamiento de los despachos a lo largo del tiempo, tanto de altos y bajos valores, permitiéndonos hacer predicciones futuras para meses y años próximos y así estar preparados para temporadas bajas y aprovechar las altas.

Donde ingresamos un valor x , es este caso fue 202501, el cual nos arroja la siguiente salida

```
# Haciendo predicciones con el mejor modelo
X_new = float(input('Ingrese un valor X: '))
Y_Pnew = modelo(X_new)
print(f'Según el valor de entrada {X_new}, la salida será: {Y_Pnew}')

Ingrese un valor X: 202501
Según el valor de entrada 202501.0, la salida será: 1113257792.0
```

Aquí nos arroja el resultado de lo que se está tratando de predecir en cuanto a las ventas del 2025.

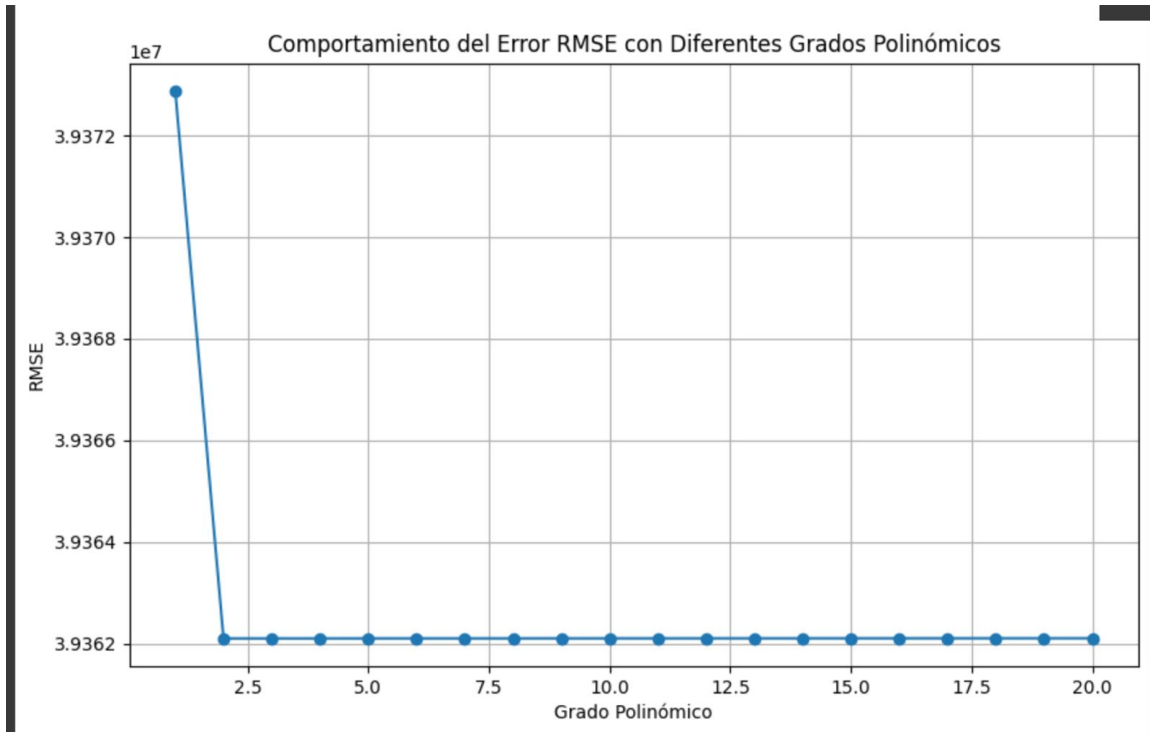


Figura 2.3

Podemos observar como este comienza con un valor alto de aproximadamente de 3.9372 y luego disminuye a medida que el grado polinómico aumenta. A partir de un grado de 5.0 este se mantiene estable con pequeñas variaciones.

3. Conclusiones y trabajos futuros

La implementación de algoritmos de machine learning nos permitió identificar patrones claros en los pedidos y en la demanda de los clientes. La empresa ahora cuenta con una visión más precisa de cómo varía las temporadas a lo largo del año, lo cual facilita la logística y el inventario.

La aplicación de clustering para segmentar clientes nos ayudó a categorizar los distintos tipos de clientes, permitiendo crear estrategias personalizadas para cada grupo. Esto podría mejorar la fidelización, optimizar los recursos y tener una logística más eficiente. El modelo predictivo con datos de 2024 mostró buenos resultados al prever la demanda futura. Esto no solo mejora la eficiencia operativa, sino que también puede contribuir a la reducción de costos en almacenamiento y distribución, alineándose con la estrategia de optimización de recursos de la empresa. Al anticipar picos de demanda, la empresa puede planificar mejor su cadena de suministro, asegurándose de que los productos estén disponibles cuando se necesiten, evitando roturas de stock o acumulación de inventario innecesario.

En trabajos futuros, se podría mejorar la automatización de predicciones en tiempo real, optimizar inventarios según las predicciones, analizar la satisfacción del cliente, y adaptar el modelo para otras áreas, maximizando su valor para la toma de decisiones estratégicas de la empresa.

Referencias bibliográficas

- María Fernanda. Encuesta Nacional de Logística-Colombia 2008. Latin American
- Logistic Center, Bogotá: 2008. p. 76.
- (Langley & Capgemini U.S. LLC, 2009)
- (Logyca, 2007).
- Hadley, G., & Whitin, TM (1963). Análisis de sistemas de inventario.
- Chen, C., et al. (2013). Previsión de la demanda mediante técnicas de aprendizaje automático.
- Hastie, T., et al. (2009). Los elementos del aprendizaje estadístico: minería de datos, inferencia y predicción.
- MacQueen, J. (1967). Algunos métodos para la clasificación y análisis de observaciones multivariadas.

- Minería de Datos (Han et al., 2011): Técnica para descubrir patrones en grandes conjuntos de datos.
- Aprendizaje Automático (Hastie et al., 2009): Técnica para entrenar modelos predictivos.
- Johnson, SC (1967). Esquemas de agrupamiento jerárquico.
- Ester, M., et al. (1996). Un algoritmo basado en densidad para descubrir clústeres en grandes bases de datos espaciales.
- Hierarchical Clustering (Johnson, 1967): Algoritmo de clustering jerárquico.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations.
- Ester, M., et al. (1996). A density-based algorithm for discovering clusters in large spatial databases.
- Han, J., et al. (2011). Data mining: Concepts and techniques.
- Hastie, T., et al. (2009). The elements of statistical learning: Data mining, inference, and prediction.
- <https://ciiblu.com/operacion-logistica-y-financiera-2/>
- <https://ciiblu.com/comunicaciones-informatica/>
- [Real Academia Española - RAE, 2001](#)
- [Cedillo Campos & Sánchez Ramírez, 2008](#)