



TRABAJO DE GRADO
Opción Seminario-Diplomado.

**ALGORITMO COMPUTACIONAL PARA EL ANÁLISIS Y TOMA DE DECISIONES
EN PROYECCIONES PARA DISMINUIR LAS DEVOLUCIONES EN LA EMPRESA
DESPACHADORA INTERNACIONAL, UTILIZANDO ESTRATEGIAS DE MACHINE
LEARNING.**

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

Estudiantes:
Jose David Gonzalez Padilla
Hernan Sarmiento Arrieta.
Tutor: Juan Carlos Briñez de León
Opción de Trabajo de grado Seminario-Diplomado.
2024.

Dedicatoria

Este trabajo es dedicado a todas las personas que han sido parte de mi proceso académico, las cuales han sido apoyo a lo largo de este camino, agradezco a los profesores por su guía y conocimiento.

a mis padres, por su apoyo y amor incondicional, a mis compañeros de clases por el compañerismo, por compartir desafíos y momentos juntos por hacer gratos mis momentos en la universidad.

por último, queremos dedicarnos este trabajo a nosotros mismos, por nunca rendirnos y seguir adelante incluso cuando todo se tornaba difícil. este logro es el resultado de nuestro esfuerzo. gracias a todos por ser parte de nuestro viaje académico y ayudarnos a llegar hasta aquí.

Agradecimientos

A Dios, a la Universidad, a mis compañeros, al profesor Juan Carlos Briñez, no podemos dejar de agradecer a nuestros padres y familiar quienes siempre han estado con su amor y apoyo incondicional. Por último agradecer a todos los que han participado en este estudio, ya sean participantes, colaboradores o fuentes de información, sin su apoyo este trabajo no hubiese sido posible.

A todo el más sincero agradecimiento por su invaluable ayuda y apoyo.

Contenido

Resumen.....	5
Palabras clave.....	5
Marco conceptual y contextual	6
Pregunta problema	7
Acercamiento a los datos	7
Descripción de variables	8
Aproximaciones con gráficos - analítica.....	9
Objetivos.....	14
Objetivo general.....	14
Objetivos específicos.....	14
Desarrollo e implementación del aprendizaje.....	14
Preparación de los datos.....	15
Modelo machine learning	19
Modelo de toma de decisiones	19
Conclusiones y trabajos futuros.....	21
Referencias.....	22

Resumen

En el mundo de la logística de transporte y entrega de mercancías la empresa DESPACHADORA INTERNACIONAL con 144 años en el mercado, esta posesionada como una de las empresas en Colombia con mayor credibilidad, sin embargo, no se escapa a las problemáticas que surgen en este tipo de actividades como lo es la devolución de productos. Por lo anterior, abordaremos dicho problema a través de herramientas machine learning, algoritmos computacionales que nos mostrarán un diagnóstico del estado actual de esta área empresarial y trazará el camino o estrategia que dará solución en materia de disminuir las devoluciones de mercancías.

Palabras clave

Transporte, Zona, Categorías, Entregas, Mercancía, Machine Learning, Empresarial, Logística de transporte, Devoluciones, Clustering.

Marco conceptual y contextual

La logística de transporte empresarial se ha convertido en una solución práctica para los problemas cotidianos en materia de envío y recepción de artículos de diferente índole, toda empresa relacionada con entregas y transporte busca la manera de potenciar esta área que de una u otra manera pasa a tener un papel fundamental en la relación cliente-proveedor, que independientemente de la Calidad del producto, si se brinda un buen servicio de entrega, genera un ambiente de respaldo y confiabilidad que a su vez se ve reflejado en las calificaciones y recomendaciones del cliente.(Tolosa, 2017)

Existen diferentes formas en que el cliente suele medir la calidad de servicio de una empresa que tenga integrado el transporte de mercancías, estas pueden variar dependiendo del tipo de mercancía y su calidad e incluso el tiempo de entrega del producto y el estado de este. Por lo anterior, se genera una problemática asociada al área de transporte de mercancías y son las devoluciones, estas están supeditadas en gran parte a la subjetividad del cliente a la hora de medir la calidad del servicio, pero también a la objetividad de la misma entrega ya que pueden presentarse retrasos, mercancía en mal estado o una mala relación con los operadores de entrega, en fin, las razones para rechazar un envío pueden ser muchas, por lo que se necesita abordar este problema en las empresas de manera urgente teniendo en cuenta que este tipo de situaciones genera pérdidas económicas y caos logístico (Flamarique, 2018).

Actualmente encontramos apoyo en herramientas como machine learning, que a través de algunos algoritmos nos puede brindar soluciones específicas a los problemas de

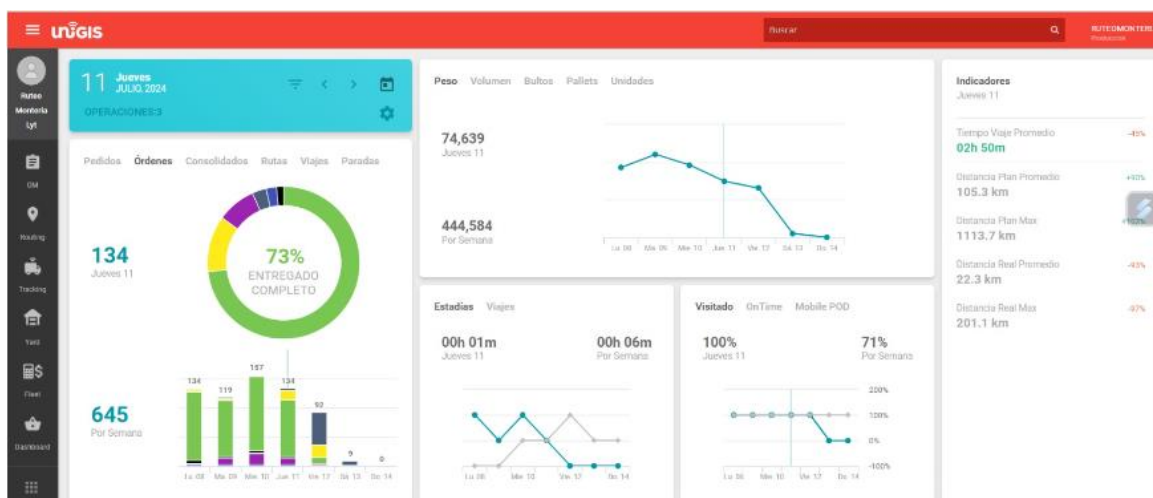
logística de transporte de mercancías, especialmente en la problemática de las devoluciones en este momento histórico en que la tecnología y la inteligencia artificial brindan una gama de herramientas que contribuyen al mejoramiento empresarial (Soler, 2024).

Pregunta problema

¿Cómo desarrollar una estrategia computacional que reduzca las devoluciones de la empresa, a partir de algoritmos de Machine Learning?

Acercamiento a los datos

Los datos para reducir las devoluciones de productos fueron tomados de la herramienta UNIGIS, la cual es utilizada por la empresa Despachadora internacional para el seguimiento y control de las entregas. Consiste en un histórico de datos de seis meses que muestran los estados de entrega, las zonas y el tipo de producto por entregar, además de si la entrega fue o no exitosa.



Gráfica 1. base de datos

Descripción de variables

El dataset escogido para nuestro trabajo cuenta con 9 variables de entrada que corresponden a la operación logística de transporte y una variable de salida que indica el estado de la entrega. A manera de resumen la siguiente imagen muestra la lista de las variables obtenidas en el conjunto de datos.

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 12891 entries, 0 to 12890
Data columns (total 10 columns):
#   Column                Non-Null Count  Dtype
---  ---                ---
0   FECHA DESPACHO        12891 non-null  int64
1   CATEGORIA              12891 non-null  float64
2   ID VALIDA              12891 non-null  object
3   ZONA                   12891 non-null  int64
4   DESTINATARIO          12891 non-null  object
5   TOTAL PESO            12891 non-null  float64
6   TIPO ENTREGA          12891 non-null  int64
7   CATEGORIA P           12891 non-null  object
8   TOTAL COBRO           12830 non-null  float64
9   ESTADO PARADA         12891 non-null  object
dtypes: float64(3), int64(3), object(4)
memory usage: 1007.2+ KB

```

Gráfica 2. Lista de Variables en el conjunto de datos

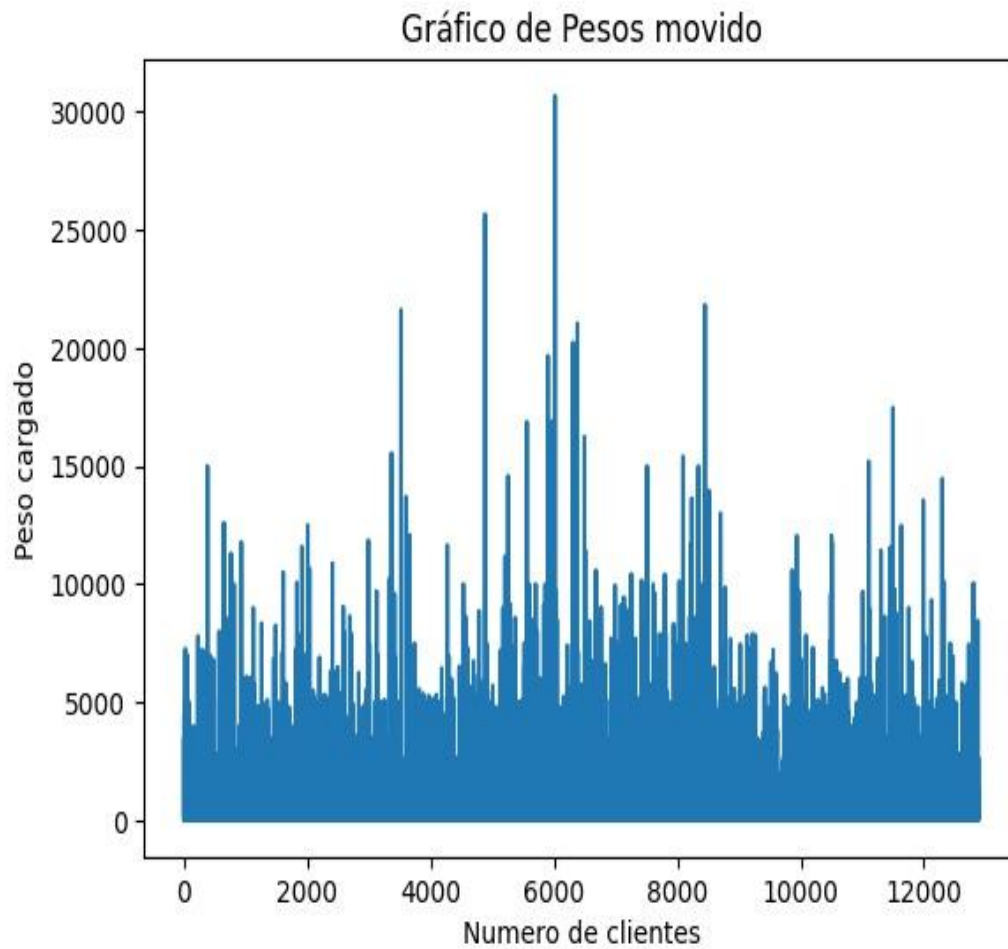
El conjunto de datos posee variables de tipo texto y numérica, donde: 3 son variables enteras, 3 decimales y 4 de texto. Podemos describir cada variable de la siguiente manera:

- fecha de despacho: se refiere al dato de año, mes, día y hora en el que la mercancía sale del almacén.
- Categoría: indica el tipo de vehículo asignado para la entrega, entre ellas tenemos: carri de 600Kg, furgón de 2 ton, planchón de 3,5 y 4,5 ton.
- ID valida: es la referencia de compra con que se valida la compra del producto.

- Zona: área de entregas dispuesta en distancia al almacén de la siguiente manera:
zona 1 < 1 Km, zona 2 de 2 a 3 Km, zona 3 entre 3 y 7 Km y zona 4 todos los destinos fuera de la zona urbana de Montería.
- Destinatario: Número o código de referencia del cliente.
- Total peso: indica el peso en Kg del domicilio.
- Tipo de entrega: hace referencia al origen de donde se toma la mercancía así:
tipo 1 mercancía tomada de la tienda, tipo 2 provenientes del proveedor.
- Categoría P: indica el tipo de mercancía que será asignado, existe cuatro categorías: pisos, HC (electrodomésticos), madrea y patio.
- Total cobro: valor total cobrado al cliente.
- Estado de parada: muestra el estado actual de la mercancía, existen dos tipos:
entregado completo y devuelto

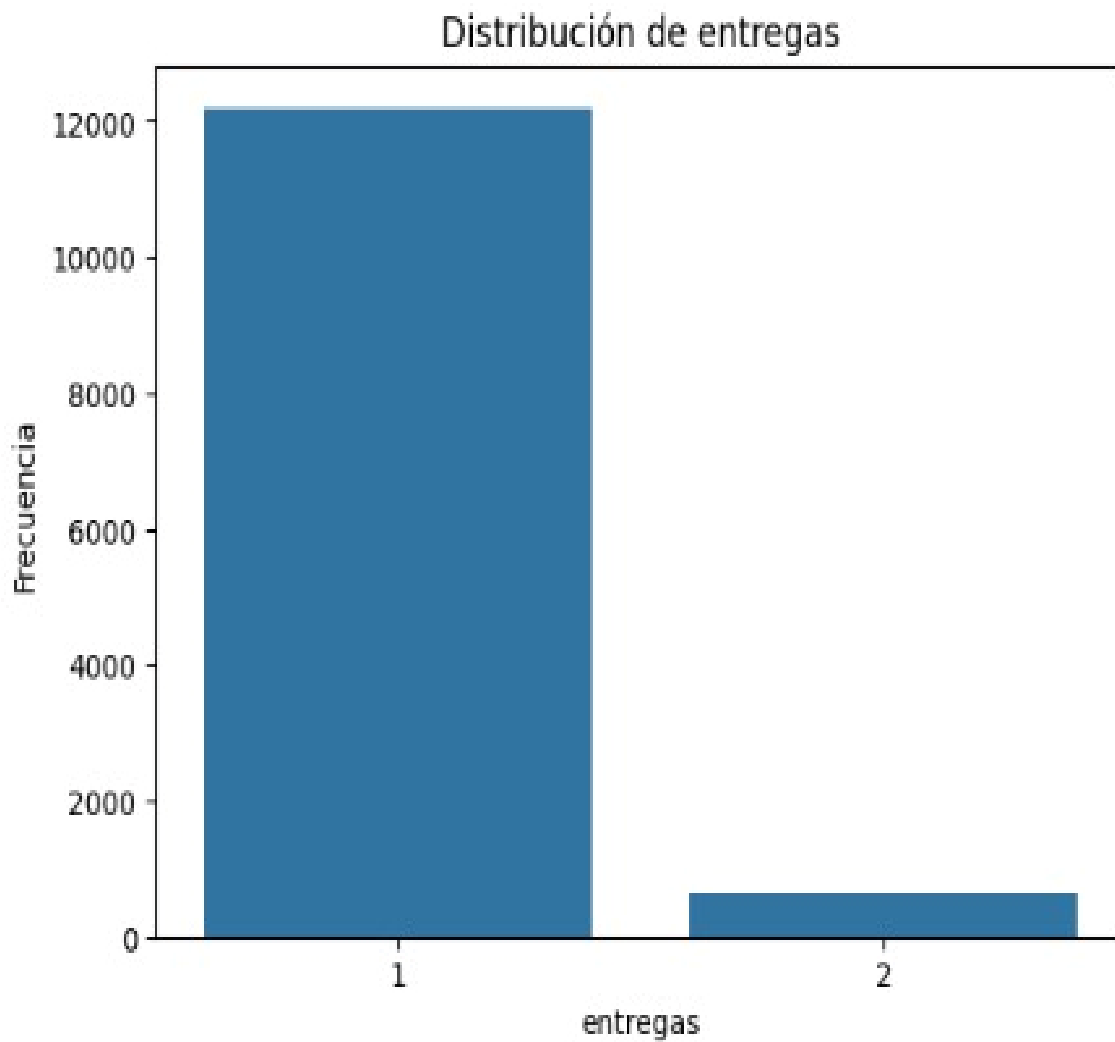
Aproximaciones con gráficos - analítica.

Con respecto a la variable peso movido, podemos observar que en promedio se mueven 5000 Kg, teniendo algunos valores atípicos que pueden llegar hasta los 30 mil Kg.



Gráfica 3.peso movido

Con respecto a la variable estado de parada podemos notar que el éxito de entrega es mayor por mucho a las devoluciones.

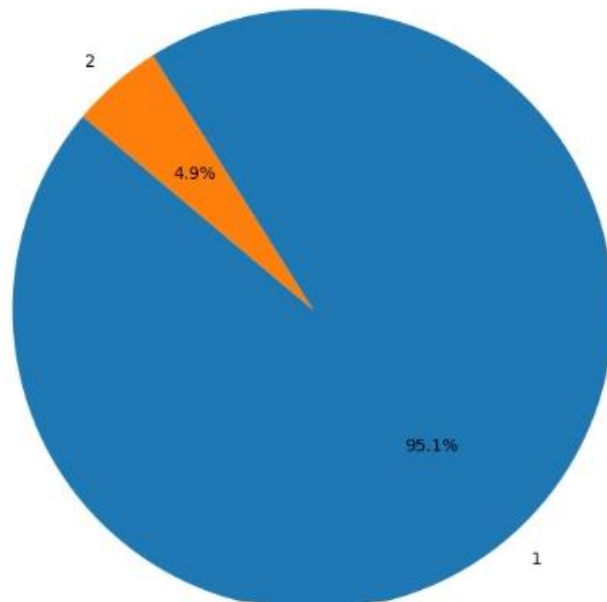


Gráfica 4. variable distribución de entregas

En esta grafica podemos apreciar el porcentaje de exitos con respecto a la cantidad de devoluciones, haciendo énfasis en ese 4,9% de desaciertos que son la razón de ser de este estudio.

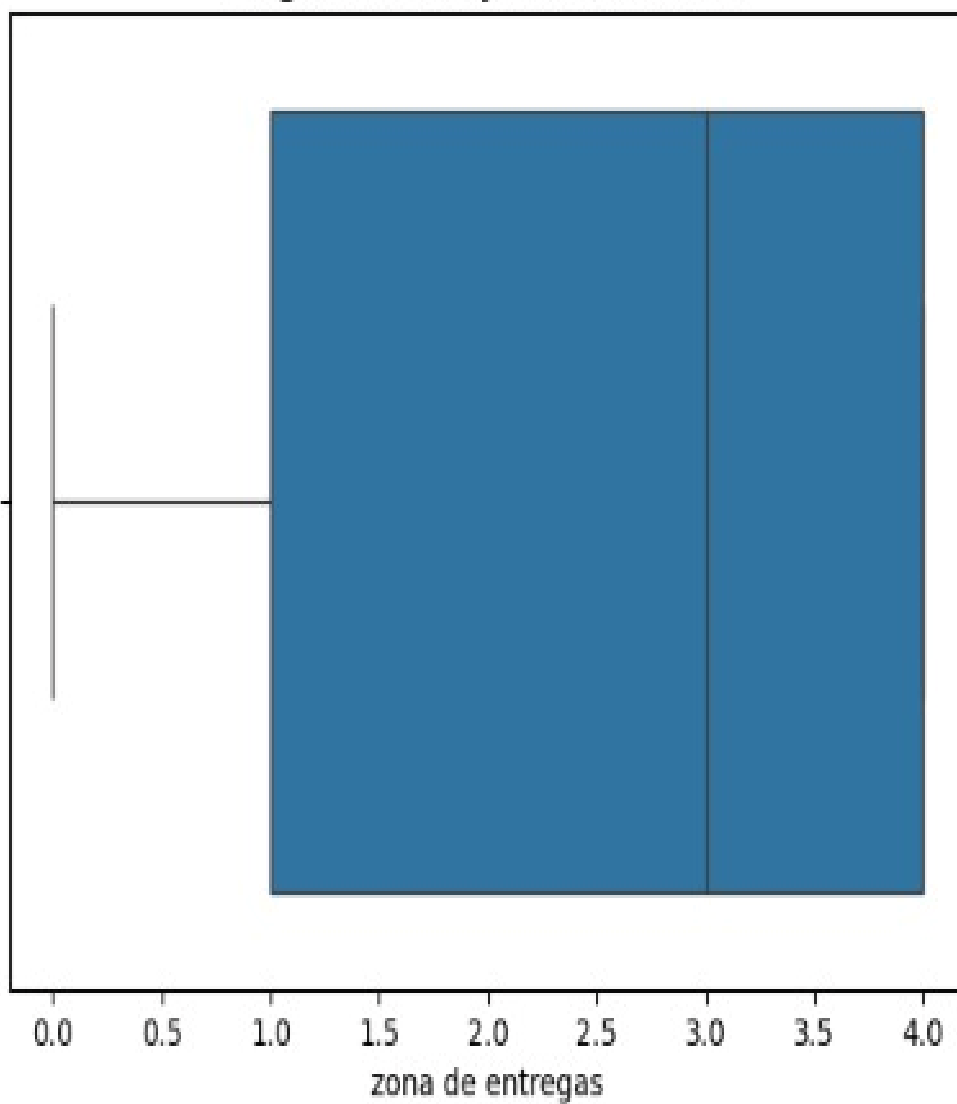


Diagrama de Torta de Distribución de Entregas

*Grafica 5. porcentaje de distribución de entregas*

Para la variable zona de entrega podemos apreciar que el área de mayor demanda se concentró en la zona 3 que abarca una distancia entre 3 y 7 Km de distancia a partir del almacén.

Diagrama de Caja de la Columna

*Gráfica 6. variable zona de entregas*

Objetivos

Objetivo general.

Implementar un algoritmo computacional para el análisis y toma de decisiones a partir de datos en línea para la reducción de devoluciones en almacenes CORONA, utilizando estrategias de machine learning.

Objetivos específicos.

- Caracterizar y procesar los datos de interés, con miras a la toma de decisiones informadas.
- Implementar un algoritmo de Machine learning para la toma de decisiones a partir de los datos de interés.
- 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.

Desarrollo e implementación del aprendizaje

Para la realización de este trabajo requerimos como primera medida, cargar una base de datos a Python desde una página oficial de la empresa DESPACHADORA INTERNACIONAL. Luego realizamos un resumen detallado de las variables que contenía la dataset, seguidamente procedimos a limpiar los datos utilizando la librería Python. El siguiente paso fue eliminar las variables que no tuvieran incidencia directa con la toma de decisiones, paso a seguir, se convirtieron los datos de texto a numérico para que el sistema tomara lectura correcta de ellos. Después hicimos un análisis estadístico rápido para conocer máximos y mínimos de las variables, por ultimo se implementó un

modelo de Machine Learning en clasificación, evaluando su desempeño y puesto a prueba con nuevos datos.

Preparación de los datos

- Cargar datos: de la página oficial de la empresa cargamos los datos hacia Python como se muestra en la siguiente imagen.

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

Dataset.head(7)
```

Elegir archivos Sin archivos seleccionados Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.

Saving Base dato Domicilio Despachadora.xlsx to Base dato Domicilio Despachadora.xlsx

	FECHA DESPACHO	PLACA	CATEGORIA	ID VALIDA	CUBRIMIENTO	ZONA	DESTINATARIO	DEPARTAMENTO	CIUDAD	TOTAL PESO	TIPO ENTREGA	CATEGORIA P	TOTAL COBRO	Id Viaje	ESTADO PARADA
0	1	YHL302	4.5	56-71279	NACIONALES	4	56-629438	ANTIOQUIA	TURBO	3400.500	1	PISOS	954332.0000	1308426	Entregado Completo
1	1	NLV972	2.0	56-71277	URBANOS	1	56-627090	CORDOBA	MONTERIA	130.000	2	HC	19269.6111	1308617	Entregado Completo
2	1	TJY356	2.0	56-71177	NACIONALES	4	99-18474236	SUCRE	COROZAL	2.030	1	HC	36763.0000	1302551	Entregado Completo
3	1	UPK274	4.5	56-71114	URBANOS	3	56-629235	CORDOBA	MONTERIA	1761.125	1	MIXTA	84869.5000	1321079	Devolución
4	1	YHK350	4.5	56-71121	URBANOS	1	56-626799	CORDOBA	MONTERIA	13.250	2	HC	22713.0000	1308497	Entregado Completo
5	1	UPK274	4.5	56-70815	NACIONALES	4	41-1796684	ANTIOQUIA	ARBOLETES	7.200	1	MADERAS/PUERTAS	40380.0000	1303266	Entregado Completo
6	1	YHK350	4.5	56-71321	NACIONALES	4	41-1796685	CORDOBA	AYAPEL	1377.610	1	PISOS	667924.0000	1303268	Devolución

Imagen 1. primer paso cargar los datos

- Resumen: se hace exhibir todas las variables contenidas en la dataset como se muestra en la siguiente imagen.

```
[ ] #Resumen de los datos
Dataset.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 12891 entries, 0 to 12890
Data columns (total 15 columns):
#   Column                Non-Null Count  Dtype
---  -
0   FECHA DESPACHO        12891 non-null  int64
1   PLACA                 12891 non-null  object
2   CATEGORIA            12891 non-null  float64
3   ID VALIDA            12891 non-null  object
4   CUBRIMIENTO          12891 non-null  object
5   ZONA                 12891 non-null  int64
6   DESTINATARIO         12891 non-null  object
7   DEPARTAMENTO         12891 non-null  object
8   CIUDAD              12891 non-null  object
9   TOTAL PESO           12891 non-null  float64
10  TIPO ENTREGA         12891 non-null  int64
11  CATEGORIA P          12891 non-null  object
12  TOTAL COBRO          12830 non-null  float64
13  Id Viaje             12891 non-null  int64
14  ESTADO PARADA        12891 non-null  object
dtypes: float64(3), int64(4), object(8)
memory usage: 1.5+ MB
```

Imagen 2. resume de los datos

- Limpieza: en esta parte se busca encontrar errores en los datos como celdas vacías, datos duplicados etc. Como se muestra en la siguiente imagen.

Limpieza de datos

```
[ ] # Identificar valores faltantes
print(Dataset.isnull().sum())

# Eliminar filas con valores faltantes
Dataset = Dataset.dropna()
Dataset.head(7)
```

```
FECHA DESPACHO    0
PLACA             0
CATEGORIA        0
ID VALIDA        0
CUBRIMIENTO      0
ZONA             0
DESTINATARIO     0
DEPARTAMENTO      0
CIUDAD           0
TOTAL PESO       0
TIPO ENTREGA     0
CATEGORIA P      0
TOTAL COBRO      0
Id Viaje         0
ESTADO PARADA    0
dtype: int64
```

Imagen 3.limpieza de datos

- Eliminación de variables: del cumulo de variables se eliminan aquellas que no tienen incidencia directa en la toma de decisiones, como lo muestra la siguiente imagen.

```
[ ] #Quitando columnas que no aportan al proceso
variable_a_eliminar = input("Ingrese el nombre de la variable a eliminar: ")

# Verificar si la variable existe en el DataFrame
if variable_a_eliminar in Dataset.columns:
    # Eliminar la variable
    Dataset = Dataset.drop(columns=[variable_a_eliminar])
    print(f"La variable '{variable_a_eliminar}' ha sido eliminada del DataFrame.")
else:
    print(f"La variable '{variable_a_eliminar}' no existe en el DataFrame.")

#Muestra un resumen del dataset
Dataset.head(10)
```

Imagen 4.eliminación de variables

- Convertir datos de texto a numérico: para una lectura correcta de los datos por parte del sistema, se transforman los datos de la variable *estado de parada*, en formato numérico como se muestra en la siguiente imagen.

```
[ ] #Convirtiendo a numérico
Conversion = {'entregado completo':1,'devolución':2}
Dataset['ESTADO PARADA']=Dataset['ESTADO PARADA'].replace(Conversion)
Dataset.head(7)
```

	FECHA DESPACHO	PLACA	CATEGORIA	ID VALIDA	CUBRIMIENTO	ZONA	DESTINATARIO	DEPARTAMENTO	CIUDAD	TOTAL PESO	TIPO ENTREGA	CATEGORIA P	TOTAL COBRO	ESTADO PARADA
0	1	YHL302	4.5	56-71279	NACIONALES	4	56-629438	ANTIOQUIA	TURBO	3400.500	1	PISOS	954332.0000	1
1	1	NLV972	2.0	56-71277	URBANOS	1	56-627090	CORDOBA	MONTERIA	130.000	2	HC	19269.6111	1
2	1	TJY356	2.0	56-71177	NACIONALES	4	99-18474236	SUCRE	COROZAL	2.030	1	HC	36763.0000	1
3	1	UPK274	4.5	56-71114	URBANOS	3	56-629235	CORDOBA	MONTERIA	1761.125	1	MIXTA	84869.5000	2
4	1	YHK350	4.5	56-71121	URBANOS	1	56-626799	CORDOBA	MONTERIA	13.250	2	HC	22713.0000	1
5	1	UPK274	4.5	56-70815	NACIONALES	4	41-1796684	ANTIOQUIA	ARBOLETES	7.200	1	MADERAS/PUERTAS	40380.0000	1
6	1	YHK350	4.5	56-71321	NACIONALES	4	41-1796685	CORDOBA	AYAPEL	1377.610	1	PISOS	667924.0000	2

Imagen 5.convertir datos a formato numérico

- Análisis estadístico rápido: es la forma de obtener valores máximos y mínimos, medidas de tendencia central, gráficos etc.

```
#Análisis estadístico rápido
Dataset.describe()
```

	FECHA DESPACHO	CATEGORIA	ZONA	TOTAL PESO	TIPO ENTREGA	TOTAL COBRO	ESTADO PARADA
count	12830.000000	12830.000000	12830.000000	12830.000000	12830.000000	1.283000e+04	12830.000000
mean	3.405924	2.540896	2.685347	837.113364	1.146687	7.153078e+04	1.049337
std	1.748332	1.610822	1.327287	1693.092652	0.353808	1.042011e+05	0.216580
min	1.000000	0.600000	0.000000	0.000000	1.000000	0.000000e+00	1.000000
25%	2.000000	0.600000	1.000000	15.680000	1.000000	2.324000e+04	1.000000
50%	3.000000	2.000000	3.000000	111.255000	1.000000	4.038000e+04	1.000000
75%	5.000000	4.500000	4.000000	945.115000	1.000000	5.603613e+04	1.000000
max	6.000000	4.500000	4.000000	30682.802000	2.000000	2.345678e+06	2.000000

Imagen 6. Análisis estadístico rápido

Modelo machine learning

Hacer uso de un modelo de Machine Learning para la toma de decisiones de la empresa Despachadora internacional tiene varias ventajas significativas, como la predicción de problemas potenciales, ya que con la implementación de estos modelos podemos analizar patrones históricos de devoluciones y así deducir cuales envíos tienen mayor probabilidad de ser devueltos, lo cual permite tomar medidas preventivas antes de que las devoluciones se den, también podemos optimizar las rutas de entregas para poder minimizar los tiempos de tránsito y reducir la probabilidad de tener retrasos o entregas fallidas, también se puede mejorar la precisión en la entrega, esto gracias a la utilización de datos históricos y en tiempo real lo cual asegura a que los envíos lleguen a tiempo, al destinatario correcto y en el tiempo correcto. con el modelo también podemos identificar la principal razón de las devoluciones permitiendo a la empresa abordar los problemas de una manera más específica.

Modelo de toma de decisiones

El uso del modelo de clustering es muy interesante ya que como empresa nos puede ayudar a agrupar los clientes en segmentos basados en comportamientos de compras, preferencias y patrones de devolución. Lo cual permite a la empresa tener una personalización de sus estrategias logísticas para cada tipo de clientes. Identificar de patrones de devoluciones en común, como tipos de productos, regiones o periodos específicos del año, lo cual permite anticipar y mitigar las causas principales de las devoluciones, nos ayuda a agrupar envíos y destinos en clusters para optimizar las rutas

de entregar y de esta manera reducir los tiempos y costos operativos. Proporciona una visión clara de las diferentes categorías de datos y de los comportamientos lo cual es ideal para la toma de decisiones informadas y basadas en datos específicos. En fin, con el clustering podemos analizar grandes volúmenes de datos de manera efectiva, podemos identificar patrones y segmentos claves, de esta manera ser más precisos en la toma de decisiones.

Conclusiones y trabajos futuros.

En este trabajo se materializó la construcción de un algoritmo para la toma de decisiones en la proyección de las devoluciones de mercancías en la logística empresarial, utilizando algoritmo machine learning.

En cuanto a los modelos utilizados, el algoritmo con mayor desempeño fue Clustering el de desempeño intermedio fue el de segmentación

Para el caso de la toma de decisiones, este trabajo opta por los dos mejores modelos.

Para trabajos futuros, se proyecta la implementación de modelos directamente en el sector de logística y transporte de mercancías, donde se pueda tener un comparativo en tiempo real entre la eficacia de entregas utilizando los métodos tradicionales y los ofrecidos mediante el modelo machine learning.

Referencias

Flamarique, S. (2018). *Flujos de mercancías en el almacén*. MARGE BOOKS.

Soler, D. (2024). *Digitalizar la gestión del transporte*. MARGE BOOKS.

Tolosa, L. (2017). *Técnicas de mejora continua en el transporte*. MARGE BOOKS.