**Logotipo

Descripción generada automáticamente**

**Análisis, optimización y predicción en rutas de transporte de última milla para una empresa de consumo masivo**

Carlos Alberto Jusquini Tinoco

Erika Tatiana Rodríguez Chávez

Monografía presentada para optar al título de Especialista en Analítica y Ciencia de Datos

Asesor  
David Manuel Villanueva Valdés, M. Sc.

Universidad de Antioquia  
Facultad de Ingeniería

Especialización en Analítica y Ciencia de Datos

Medellín, Antioquia, Colombia

2023

|  |  |
| --- | --- |
| **Cita** | (Jusquini Tinoco & Rodríguez Chávez, 2023) |
| **Referencia**  **Estilo APA 7 (2020)** | [Jusquini Tinoco, C. A. & Rodríguez Chávez, E. T. (2023). *Análisis, optimización y predicción en rutas de transporte de última milla para una empresa de consumo masivo.* Trabajo de grado especialización]. Universidad de Antioquia, Medellín, Colombia. |

**** 

Especialización en Analítica y Ciencia de Datos, CohorteVI.

Centro de Investigación Ambientales y de Ingeniería (CIA).

|  |  |
| --- | --- |
|  | Diagrama  Descripción generada automáticamente con confianza media |

Centro de Documentación Ingeniería (CENDOI)

**Repositorio Institucional:** http://bibliotecadigital.udea.edu.co

Universidad de Antioquia - www.udea.edu.co

Rector: John Jairo Arboleda Céspedes.

Decano: Julio Cesar Saldarriaga Molina

Jefe departamento: Diego José Luis Botia Valderrama

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Antioquia ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

**Tabla de contenido**

[1. Descripción del problema 6](#_heading=h.35nkun2)

[1.1. Problema de negocio 6](#_heading=h.1ksv4uv)

[1.2. Aproximación desde la analítica de datos 7](#_heading=h.44sinio)

[1.3. Origen de los datos 7](#_heading=h.z337ya)

[1.4. Métricas de desempeño 8](#_heading=h.3j2qqm3)

[2. Objetivos 9](#_heading=h.4i7ojhp)

[2.1. Objetivo general 9](#_heading=h.2xcytpi)

[2.2. Objetivos específicos 9](#_heading=h.3whwml4)

[3. Datos 10](#_heading=h.2bn6wsx)

[3.1. Datos originales 10](#_heading=h.qsh70q)

[3.2. Datasets 12](#_heading=h.3as4poj)

[3.3. Analítica descriptiva 13](#_heading=h.1pxezwc)

[**Referencias 15**](#_heading=h.vx1227)

**Lista de figuras**

**Figura 1***.* Muestra del dataset procesado13

[**Figura 2** Tipos de datos](#_heading=h.1664s55) 13

**Figura 3** Descripción estadística del dataset 14

**Figura 4** Pedidos por tipo de día14

**Siglas, acrónimos y abreviaturas**

**ERP** Enterprise Resourcing Planning - Planificación de recursos empresariales

**MSc** Magister Scientiae

# Descripción del problema

Se estima que cerca del 30% de los gastos asociados a la distribución de un producto está concentrado en la logística de última milla (España. Universidad de Valladolid, 2018), esa última fase de la logística en que el producto sale del almacén o centro de distribución para llegar al consumidor final. El reto para las compañías de consumo masivo es encontrar modelos más eficientes que les permita garantizar el cumplimiento de la promesa de entrega con sus clientes y optimizar el gasto derivado de este proceso.

La distribución tienda a tienda tiene adicional a los retos de infraestructura (estados de las vías), movilidad y condiciones meteorológicas, el de considerar las condiciones de seguridad en las zonas de entrega ya que los procesos de recaudo de efectivo las hacen más susceptibles a hurtos. La mayoría de las soluciones de enrutamiento disponibles consideran tan solo número de paradas, un tiempo promedio de atención por cliente y la distancia a recorrer para el cálculo de tiempo de servicio del vehículo, lo que limita el balanceo de las rutas y conlleva a sobrecostos por no finalización de estas y pérdida de la credibilidad por parte del cliente por la afectación en el servicio.

## Problema de negocio

El caso particular que abordaremos en este documento parte de un volumen de venta diaria que oscila entre los 480 y 560 clientes, los cuales se atienden con una promesa de entrega a 48 horas, es decir, la venta de lunes se distribuye el día miércoles y así sucesivamente, todos los días esta venta es enrutada en un número fijo de vehículos a través de un aplicativo de enrutamiento llamado *Roadnet*, el balanceo de estas rutas se hace de manera lineal; buscando el mismo número de pedidos por vehículo, el mismo volumen (metros cúbicos) y peso (kilogramos) y dinero a recaudar, sin considerar factores propios de la zona de entrega como seguridad, estados de las vías, y/o restricciones de movilidad vehicular, tampoco se hace una predicción de los próximos días de distribución lo que no permite anticiparse a crecimientos estacionarios de venta; intuitivamente se identifica la concentración de la venta en las dos últimas semanas del mes así como una posible desconcentración en los días de la semana dependiendo del tipo de producto que se tenga en oferta, pero a hoy nada de esto se tiene analíticamente en cuenta haciendo rígida la estructura de recursos para la operación y el gasto logístico el cual se ve impactado entre un 15% y 20% mensualmente por concepto de sobrecostos para el reenvío de pedidos no entregados, afectando a su vez el nivel de servicio y la credibilidad del cliente con la promesa de entrega.

## Aproximación desde la analítica de datos

El modelo a desarrollar tiene como objetivo predecir el comportamiento de la venta e integrar con las condiciones propias de la zona de entrega (limitaciones horarias por condiciones de seguridad, estado de las vías o restricciones de movilidad, tiempo en tráfico y condiciones meteorológicas) para plantear un balanceo de las rutas a partir de la estimación del tiempo de servicio.

Utilizando como entrenamiento el histórico de los últimos seis (6) meses será posible predecir el volumen de venta por día de la semana, identificando variaciones a nivel de volumen (metros cúbicos), peso (kilogramos) y número de pedidos de las rutas, con lo cual lo primero que permitirá identificar es la densidad de la distribución por zonas de entrega.

A través de la integración con la API de Google Maps y la geolocalización de los clientes el modelo podrá definir tiempo en tráfico, velocidad promedio de tránsito, posibles desvíos requeridos y/o la ruta lógica más eficiente. Adicional, será posible construir un mapa de calor de seguridad que permita generar restricciones de volumen, peso y/o número de pedidos en aquellas zonas en donde la seguridad sea un factor crítico a partir del cual por ejemplo limite la franja horaria, ajustando el enrutamiento para que sea posible finalizar recorrido bajo estas limitaciones.

## Origen de los datos

Los datos a utilizar provienen del histórico de ventas de los últimos 6 meses disponible en el *ERP* (Enterprise Resourcing Planning - Planificación de recursos empresariales) usado por la compañía, herramienta que guarda todos los registros de los pedidos realizados por los clientes incluyendo: fecha de facturación, artículos, cantidad por artículo, número de factura, número de planilla (vehículo en el cual fue distribuido) y vendedor.

Por otra parte, usaremos la maestra de clientes proporcionada por el área de ventas en la cual se relaciona para cada cliente su ubicación geográfica en términos de longitud y latitud.

## Métricas de desempeño

* Se utilizarán las métricas: R2, F1-Score, ROC, RMSE y MAE para la evaluación del modelo
* A nivel de negocio la métrica a evaluar para entender balanceadas las rutas será el tiempo total en ruta por vehículo, un tiempo total en servicio inferior a 10 horas, es considerada viable y con un riesgo bajo de no terminación de recorrido.

# Objetivos

## Objetivo general

Desarrollar una herramienta que integre la estimación de los volúmenes de venta y el comportamiento de factores externos asociados a las zonas de entrega para la predicción del tiempo en ruta de cada vehículo buscando un tiempo total de servicio inferior a las 10 horas.

## Objetivos específicos

Realizar el análisis descriptivo de los datos históricos de venta de los últimos seis meses.

Construir un modelo predictivo de la venta para estimar los recursos necesarios para su distribución.

Integrar la predicción de la venta y el análisis de las condiciones en ruta para la construcción del modelo de enrutamiento y el cálculo del tiempo en servicio por vehículo.

# Datos

## Datos originales

Se descarga el reporte de pedidos en formato CSV de los meses abril a septiembre de 2023, directamente desde el *ERP* este reporte contiene los siguientes campos:

El ***consecutivo de pedido***, es asignado al momento que el vendedor toma el pedido en la visita al cliente con una longitud de ocho (8) caracteres siendo la primera una letra y los otros siete (7) dígitos.

El ***código de cliente,*** un ID de cliente asignado por la compañía al momento de su creación, con una longitud de ocho (8) caracteres, los dos (2) primeras letras y los otros seis (6) dígitos.

El ***código de vendedor***, un ID asignado al vendedor que atiende el al cliente, con una longitud de ocho (8) caracteres, todos dígitos.

El **e*stado de transmisión del pedido***, indica si el pedido fue transmitido o no, el pedido pasa por las siguientes fases: toma (digitación al momento de la visita en el cliente), transmisión (al momento en que el dispositivo móvil usado por el vendedor hace interfaz con el ERP de la compañía), liberación (el sistema de acuerdo al tren de atención, día de atención, configurado para cada cliente libera automáticamente el pedido para su facturación y despacho), planillado (el momento de asignación del pedido a un vehículo en el proceso de enrutamiento) y facturación (cuando se genera la factura de cada pedido para su despacho y entrega). El estado de transmisión puede ser: ESTA OK, que significa que el pedido fue transmitido correctamente y dentro del tren de atención asignado, y SIN TX, que significa que el pedido fue transmitido fuera de la frecuencia configurada y por ende no es posible que sea tendido, el horario de transmisión es de 7:00am a 6:00pm si un pedido de un cliente que está configurado para visitarse el lunes se transmite el martes quedará en estado SIN TX, así mismo, si fue transmitido a las 7:00pm aunque sea el mismo lunes.

La ***ciudad a la cual pertenece***, en este caso estamos tomando los registros de la ciudad de CARTAGENA (BOLÍVAR) por lo que todos los registros indicarán esta descripción.

El ***estado de facturación***, indica si el pedido fue anulado o facturado y despachado, los valores en esta columna pueden ser; ANUA, Anulado automáticamente (cuando el pedido está fuera de frecuencia, aplica lo mismo que se indicó en el estado de transmisión), ANUT, Anulado por corte transmisión (cuando el pedido está fuera de ruta, eso significa que es posible que haya sido transmitido el día que se configuró para su atención, la geolocalización indica que está distante de la zona en donde se concentra la venta de ese día, por lo que atenderlo implicaría un desvío no considerado), ANUU, Anulado por el usuario (Anulación manual por error en digitación que se identifique o por solicitud del área de ventas), FACA, Factura aprobada por proveedor tecnológico, FACE, Factura enviada a proveedor tecnológico, FACP, Facturado y planillado, FACX, Facturado no aprobado por proveedor tecnológico, (estos últimos cuatro están asociados al flujo de facturación, una vez embarcado el pedido la factura pasa por diferentes estados para cumplir la normatividad de facturación electrónica, el estado exitoso final de un pedido será FACP y el estado FACX será un pedido que no cumplió correctamente su ciclo de facturación por lo cual no será despachado)

El ***valor del pedido***, indica el valor en pesos producto a producto de la cantidad solicitada por el cliente.

El ***valor embarcado***, indica el valor efectivamente facturado producto a producto, este puede ser inferior al valor pedido cuando en el almacén no hay suficiente inventario disponible generando agotados.

La ***factura***, consecutivo de facturación de cada pedido.

El ***código del artículo***, ID asignado por el *ERP* a cada artículo de la compañía, con una longitud de ocho (8) caracteres.

La ***descripción del artículo***, breve descripción del artículo solicitado.

La ***unidad de medida***, puede ser UN (unidades), CJ (caja) y DP (display) y está asociada a la presentación del producto en la cual se le hará entrega al cliente.

La ***cantidad pedida***, cantidad solicitada por el cliente de cada artículo en la unidad de medida correspondiente.

La ***cantidad embarcada***, cantidad exitosamente embarcada y facturada que será entregada y cobrada al cliente.

***Fechas del flujo del pedido*** (toma, transmisión, liberación, facturación y planillado), trazabilidad del pedido en las diferentes etapas del proceso. Las fechas se presentan en formato mes, día, año.

Este reporte tiene un peso de 40 MB y ha sido exportado en formato CSV para los fines del presente trabajo, con un total de 272,643 registros.

Se cuenta con otro CSV; la maestra de clientes, en la cual solo se incluyen las columnas de código, (ID del cliente), código del vendedor (ID de vendedor que tiende al cliente), los datos de geolocalización: latitud y longitud. Con un total de 4,495 registros que corresponden a los clientes activos y el archivo pesa 188 KB.

## Datasets

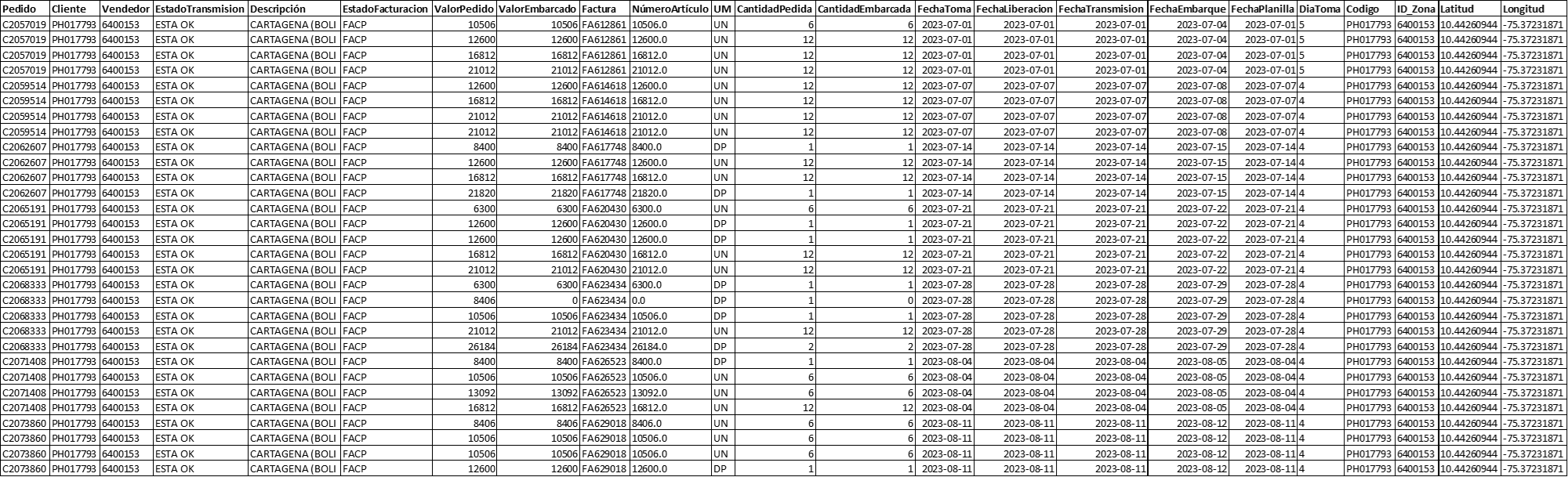
Para la construcción del dataset a trabajar, inicialmente leemos las dos bases de datos mencionadas anteriormente, se realiza el proceso de normalización de las columnas y de limpieza de registros sin fecha o con fecha en formato diferente a día, mes, año, con el fin de depurar pedidos que no han sido facturados y despachados trabajando solo con registros que se hayan distribuido.

Se agrega en el proceso una columna con el día de la semana a la cual corresponde la fecha del pedido, esto con el fin de hacer un análisis según la frecuencia de venta, las zonas de venta se repiten semana a semana, en efecto, la venta de un lunes solo es comparable con la de otro lunes.

Por último, se integran las dos bases relacionando en el DataFrame final la geolocalización del cliente, que se requerirá para identificar la zona de entrega y realizar posteriores análisis geoespaciales.

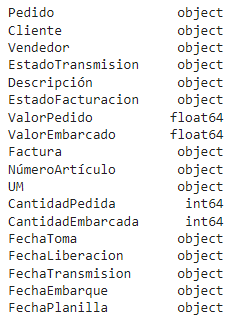
## Analítica descriptiva

Una muestra del dataset procesado se observa a continuación en la *Figura 1.*



*Figura 1. Muestra del dataset procesado.*

En la *Figura 2* se muestran los tipos de datos del dataset.



*Figura 2. Tipos de datos*

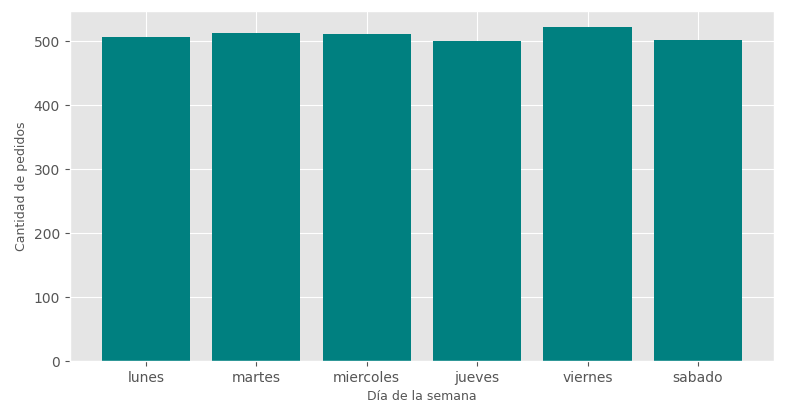
La descripción estadística de las variables numéricas se observa a continuación en la *Figura 3.*



*Figura 3. Descripción estadística del dataset*

En la *Figura 4* se presentan los pedidos por tipo de día, con el fin de entender en promedio en el rango de fecha de análisis cuantos pedidos son programados para entrega

# 



*Figura 4. Pedidos por tipo de día*

La cantidad de artículos distintos que se programaron para entrega en el tiempo de análisis fue: 354 artículos.

# 

# Referencias

Torre Lloreda, F. (2018). *Aplicaciones del Big Data en la Logística*. (pp. 58). Universidad de Valladolid.