

# Análisis del valor las órdenes en los dataset

1

# Análisis de serie de tiempo - Predicción ARIMA

2

## # Análisis del valor las órdenes en los dataset

En la tabla de ordenes y pagos, (olist\_order\_payments\_dataset) se puede ver tanta cantidad valores de pago como payment\_sequential. Es decir, que una orden de pago está dividida en varias secuencias de pago. Esto quiere decir, que cada secuencia de pago tiene un valor de pago asociado y la suma de cada uno de ellos será el total monetario de la orden.

Asimismo, en la tabla de ordenes y productos, olist\_order\_items\_dataset, se puede entender que una orden tiene asociada varios productos . El valor de cada producto mas su valor de envio, da el mismo total que la orden.

Al poder ver esto, se puede entender que:

- existen varios registros con el mismo order\_id, no solo por la cantidad de diferentes productos asociados a la orden, si no tambien por las distintas "payment\_sequential" con la que cuenta la orden. Cada una de ellas tiene un valor distinto en el pago (payment\_installments) pero si se suman, va a dar el total de lo que se pago por producto y la suma de productos, dara el total del pago de la orden.

Por lo tanto se decide:

Realizar una agrupamiento por orden y dejar el valor total unificado en una nueva variable para poder utilizarla en la predicción de ventas.

Las tablas a utilizar para el modelo de ventas serán:

Tabla (5): olist\_orders\_dataset

- Variables a utilizar:

- \* order\_id

- \* order\_status: se tomarán aquellas con valor delivered y se excluirán aquellas canceled.

- \*order purchase timestamp (para saber la fecha de la orden)

Tablas y variables a utilizar en el modelado de machine learning

Tabla (6): olist\_order\_payments\_dataset.

- Variables a utilizar:

- \* order\_id

- \* payment sequential (para agrupar la orden y calcular el monto total)
- \* payment value (será la base del monto total de la orden)

Existe dos tablas que aportan información de las ventas, pero a priori se descartan del análisis. Una de ellas contiene información de cuantos productos existen por orden y cuales productos son los comprados. La otra tiene información de las categorías de los productos.

Tabla (3): olist\_products\_dataset

- Variables:

- \* product\_id
- \* product\_category\_name

Tabla (7): olist\_order\_items\_dataset

- Variables:

- \* order\_id
- \* order\_item\_id
- \* product\_id

## # Análisis de serie de tiempo - Predicción ARIMA

Se ha decidido implementar un modelo ARIMA para predecir en base al tiempo, los valores de órdenes que se tendrán de aquí a 3 años.

Los pasos que se han tomado son los siguientes:

Resumen:

1. Preparación de los datos:

- Carga de datos: Cargar los conjuntos de datos en Python utilizando bibliotecas como Pandas para manipular los datos.
- Procesamiento: Se crearán los dataframes necesarios para llegar al dataframe final que se procesará con ARIMA. Se aclara que se ha creado una variables llamado pago\_orden que contiene el valor de ventas por orden y valor promedio por mes y año. Esta variable es la variable a predecir..

2. Aplicación de modelo de series de tiempos:

- Arima:

Preparación de los Datos: Los datos se agrupan por mes (con el objetivo de obtener una serie temporal más manejable y para ayudar a revelar patrones estacionales o tendencias)

Revisión de la Estacionariedad: Se realizan pruebas de estacionariedad, como la prueba de Dickey-Fuller aumentada, para determinar si es necesario diferenciar los datos para

hacerlos estacionarios. Los datos no eran estacionarios, por lo que se aplicó la técnica para convertirlos en estacionarios y poder aplicarles ARIMA.

Identificación de los Parámetros ARIMA: Se utilizaron las funciones de autocorrelación (ACF) y autocorrelación parcial (PACF) para identificar los parámetros  $p$ ,  $d$ , y  $q$  del modelo ARIMA.

Ajuste del Modelo ARIMA: Se ajustó el modelo ARIMA con los parámetros identificados a los datos. Y adicionalmente se optimizaron los parámetros con Optuna.

Predicciones: Se realizaron las predicciones para los próximos años.