

DEFINICION DEL PROBLEMA DE INVESTIGACION, PREGUNTA DE INVESTIGACIÓN Y BASES DE DATOS SELECCIONADAS

Johan Sebastián Rodríguez Erick Stiven Vanegas Sergio Giovanny Cepeda

Facultad de Ingeniería, Universidad Minuto de Dios

NRC: 2480

Luisa Arciniegas Zamora

Marzo del 2023



Definición Del Problema

En una empresa de ventas Online almacena la información sobre sus ventas para determinar la causa del porque no se efectúan todas las ventas, determinar los productos más vendidos y así obtener características económicas y buscar la manera de aumentar sus ventas por medio de un portal Online, y reducir costos a la iniciativa de marketing incrementar el alta suma de consumidores.

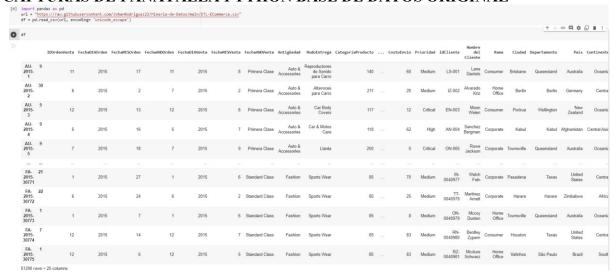
A partir del planteamiento del problema surgen las siguientes preguntas de investigación:

- ¿Qué Producto es el más vendido y menos vendido?
- ¿Cuál es el género que más Compra?
- > ¿Cuál es el mes en el que más compra se efectúan y menos efectúan?
- > ¿Por qué no se efectúan las ventas?
- ¿Qué producto es el más efectivo para promocionar en publicidad?
- ¿Cómo aumentar las ventas?
- En qué ciudad o Continente se efectúan más ventas?

Rango de datos

Se obtiene un rango de datos de enero del 2015 a enero de 2016, obteniendo información de datos en un rango de 12 meses

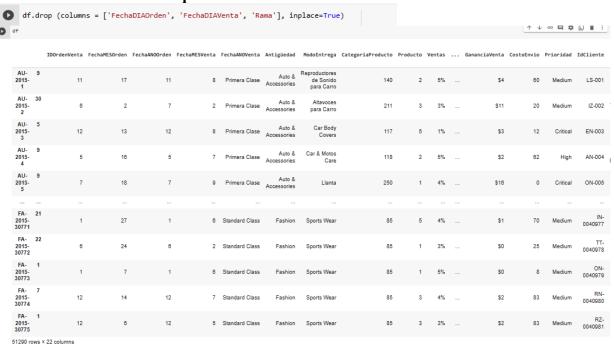
CAPTURAS DE PANATALLA PYTHON BASE DE DATOS ORIGINAL



Utilizamos el comando path para guardar los cambios

```
vos [35] path = "C:\ETL-ECommerce"
vos [36] df.to_csv(path)
```

Eliminación de columnas que no vamos a utilizar



Definición de las variables de investigación definidas previamente.

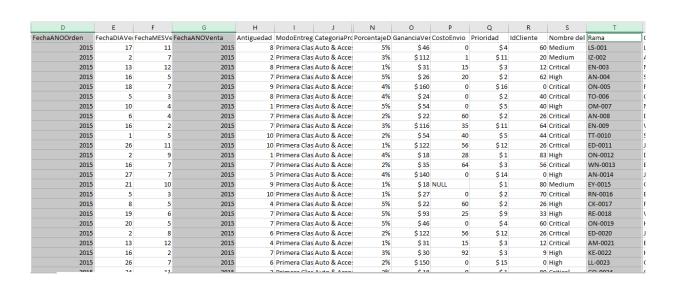
- IDOrdenVenta
- FechaDIAOrden
- FechaMESOrden
- FechaDIAVenta
- FechaMESVenta
- Antigüedad
- ModoEntrega
- CategoriaProducto
- CostoEnvio
- Prioridad

- IdCliente
- Nombre del Cliente
- Ciudad
- Departamento
- Pais
- Continente
- VentaEfectiva

Variables Eliminadas

Las variables no usadas y que consideramos no relevantes son:

- FechaAÑOOrden
- FechaAÑOVenta
- Rama



Eliminación de Variables Colab (Por un incidente con el aplicativo Colab, los índices de cada columna se visualizan uno a la derecha)

| | | IDOrdenVenta | FechaDIAOrden | FechaMESOrden | FechaDIAVenta | FechaMESVenta | Antigüedad | ModoEntrega | CategoriaProducto | Producto | Ventas | GananciaVe |
|-----------------------|----|--------------|---------------|---------------|---------------|----------------|-----------------------|--|-------------------|----------|--------|----------------|
| AU- 2015- 1 | 9 | 11 | 17 | 11 | 8 | Primera Clase | Auto & Accessories | Reproductores de Sonido para Carro | 140 | 2 | 5% | |
| AU- 2015- 2 | 30 | 6 | 2 | 7 | 2 | Primera Clase | Auto & Accessories | Altavoces para Carro | 211 | 3 | 3% | |
| AU- 2015- 3 | 5 | 12 | 13 | 12 | 8 | Primera Clase | Auto & Accessories | Car Body Covers | 117 | 5 | 1% | |
| AU- 2015- 4 | 9 | 5 | 16 | 5 | 7 | Primera Clase | Auto & Accessories | Car & Motos Care | 118 | 2 | 5% | |
| AU- 2015- 5 | 9 | 7 | 18 | 7 | 9 | Primera Clase | Auto & Accessories | Llanta | 250 | 1 | 4% | |
| | | | | | | | | | | | | |
| FA- 2015- 30771 | 21 | 1 | 27 | 1 | 6 | Standard Class | Fashion | Sports Wear | 85 | 5 | 4% | |
| FA- 2015- | 22 | 6 | 24 | 6 | 2 | Standard Class | Fashion | Sports Wear | 85 | 1 | 3% | |

Validación de Datos NAN

| df.describe(include='0') | | | | | | | | | | | | | | |
|--------------------------|----------------|------------|-------------|--------|----------|---------------|-----------|-----------|--------------------------|----------|---------------------|--------------|------------------|------------|
| | FechaANOVenta | Antigüedad | ModoEntrega | Ventas | Cantidad | GananciaVenta | Prioridad | IdCliente | Nombre del Cliente | Rama | Ciudad | Departamento | Pais | Continente |
| count | 51290 | 51290 | 51290 | 51290 | 51290 | 51290 | 51290 | 51290 | 51290 | 51290 | 51290 | 51290 | 51290 | 51290 |
| unique | 4 | 4 | 42 | 5 | 154 | 17 | 4 | 51290 | 795 | 3 | 3636 | 1094 | 147 | 1: |
| top | Standard Class | Fashion | Sports Wear | 3% | \$ 17 | \$2 | Medium | LS-001 | Mcclain O'Donnell | Consumer | New York City | California | United States | Centra |
| freq | 30775 | 30775 | 2827 | 10360 | 1154 | 6835 | 29433 | 1 | 108 | 26518 | 915 | 2001 | 9994 | 1111 |
| 7: | | | | | | | | | | | | | | |

Instrucción 0

| <pre>df.describe(include='0')</pre> | | | | | | | | | | | | | | |
|-------------------------------------|----------------|------------|-------------|--------|----------|---------------|-----------|-----------|--------------------------|----------|---------------------|--------------|------------------|------------|
| | FechaANOVenta | Antigüedad | ModoEntrega | Ventas | Cantidad | GananciaVenta | Prioridad | IdCliente | Nombre del Cliente | Rama | Ciudad | Departamento | Pais | Continente |
| count | 51290 | 51290 | 51290 | 51290 | 51290 | 51290 | 51290 | 51290 | 51290 | 51290 | 51290 | 51290 | 51290 | 51290 |
| unique | 4 | 4 | 42 | 5 | 154 | 17 | 4 | 51290 | 795 | 3 | 3636 | 1094 | 147 | 13 |
| top | Standard Class | Fashion | Sports Wear | 3% | \$ 17 | \$2 | Medium | LS-001 | Mcclain O'Donnell | Consumer | New York City | California | United States | Central |
| freq | 30775 | 30775 | 2827 | 10360 | 1154 | 6835 | 29433 | 1 | 108 | 26518 | 915 | 2001 | 9994 | 11117 |

Validación de datos Vacíos

| <pre>df.isna().sum()</pre> | |
|-----------------------------|---|
| IDOrdenVenta | 0 |
| FechaDIAOrden | 0 |
| FechaMESOrden | 0 |
| FechaANOOrden | 0 |
| FechaDIAVenta | 0 |
| FechaMESVenta | 0 |
| FechaANOVenta | 0 |
| Antigüedad | 0 |
| ModoEntrega | 0 |
| CategoriaProducto | 0 |
| Producto | 0 |
| Ventas | 0 |
| Cantidad | 0 |
| PorcentajeDescuentoAplicado | 0 |
| GananciaVenta | 0 |
| CostoEnvio | 0 |
| Prioridad | 0 |
| IdCliente | 0 |
| Nombre del Cliente | 0 |
| Rama | 0 |
| Ciudad | 0 |
| Denartamento | А |

Link del Repositorio GitHub

https://github.com/JohanRodriguez23/Mineria-de-Datos

https://drive.google.com/drive/folders/1PPVEmlgoqhe_x5JBrjWGJILs2st3PLFW