**Entregable 1**

**Elección del Dataset**

**DATASET PROPIO**

**Ventas de los años 2018 al 2020**

**Realizado por:**

Carlos, Parra

Imagen que contiene vajilla, plato, dibujo

Descripción generada automáticamente

Diciembre 2022

# CONTENIDO

[DESCRIPCIÓN DEL DATASET 2](#_Toc121151089)

[ALCANCE 2](#_Toc121151090)

[TABLAS Y TIPOS DE DATOS 2](#_Toc121151091)

[VARIABLES POTENCIALMENTE INTERESANTES 3](#_Toc121151092)

[HIPOTESIS 7](#_Toc121151093)

# DESCRIPCIÓN DEL DATASET

Para este proyecto, se utilizará un dataset propio, con información real, por lo que se agradece hacer un uso racional de la información y tratarla con la mayor discreción.

Solo debe ser utilizada para propósitos específicos de esta investigación.

La información es confidencial, y todos los derechos a la misma que han sido o serán divulgados permanecerán como propiedad del autor. El Receptor no obtendrá derecho alguno, de ningún tipo, sobre la información, ni tampoco ningún derecho de utilizarla, excepto para el objeto del presente acuerdo.

# ALCANCE

Se utilizará esta información para la realización de análisis, modelos estadísticos, proyecciones, etc., y todo lo relacionado con la Comisión 32795 del curso de Data Science impartido por la CODERHOUSE.

# TABLAS Y TIPOS DE DATOS

Las tablas utilizadas para el modelo se encuentran almacenadas en una carpeta compartida de Google Drive, llamada “DATASET” están en formato .xlsx y son accesibles mediante el siguiente URL <https://drive.google.com/drive/folders/1JBjZRPlZTm-Ax9WZx4quXsWlJqPbkKqm?usp=sharing>

Las tablas utilizadas son las siguientes:

![Tabla

Descripción generada automáticamente]()

# VARIABLES POTENCIALMENTE INTERESANTES

TABLA: ARTICULOS (df\_Articulos)

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 1490 entries, 0 to 1489

Data columns (total 9 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 Artículo 1490 non-null int64

1 Artículos 1490 non-null object

2 Stock Disponible 1490 non-null int64

3 Costo Vigente $ 1490 non-null int64

4 Costo Reposición $ 1490 non-null int64

5 Pendientes recepción 1490 non-null int64

6 Unnamed: 6 148 non-null datetime64[ns]

7 Pendientes de Entrega 1490 non-null int64

8 Stock Futuro 1490 non-null int64

dtypes: datetime64[ns](1), int64(7), object(1)

memory usage: 104.9+ KB

TABLA: LISTA DE PRECIO FINAL (df\_Lista\_Final)

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 1355 entries, 0 to 1354

Data columns (total 3 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 Código artículo 1355 non-null int64

1 Descripción 1355 non-null object

2 Precio Unitario 1355 non-null int64

dtypes: int64(2), object(1)

memory usage: 31.9+ KB

TABLA: LISTA DE PRECIO MAYORISTA (df\_Lista\_Mayorista)

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 1352 entries, 0 to 1351

Data columns (total 3 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 Código artículo 1352 non-null int64

1 Descripción 1352 non-null object

2 Precio Unitario 1352 non-null int64

dtypes: int64(2), object(1)

memory usage: 31.8+ KB

TABLA: Ventas (df\_Ventas)

<class 'pandas.core.frame.DataFrame'>

Index: 0 entries

Data columns (total 25 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 Nombre Doc 0 non-null object

1 Número del Documento 0 non-null object

2 Código del Cliente 0 non-null object

3 Nombre del Cliente 0 non-null object

4 Cod. Vendedor 0 non-null object

5 Nombre del Vendedor 0 non-null object

6 Fecha 0 non-null object

7 Dia 0 non-null object

8 Mes 0 non-null object

9 Año 0 non-null object

10 Afecto 0 non-null object

11 Total 0 non-null object

12 RUT Cliente 0 non-null object

13 Cod. Producto 0 non-null object

14 Desc. Producto 0 non-null object

15 Cantidad 0 non-null object

16 Precio Unitario 0 non-null object

17 Total Detalle 0 non-null object

18 Total Neto Linea 0 non-null object

19 Descripción Detallada 0 non-null object

20 Costo Vigente 0 non-null object

21 Nom Tipo Doc Asociado 0 non-null object

22 Num Tipo Doc Asociado 0 non-null object

23 Fecha del Doc Asociado 0 non-null object

24 Línea 0 non-null object

dtypes: object(25)

memory usage: 0.0+ bytes

TABLA: VENTAS 2018 (df\_Ventas\_2018)

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 82461 entries, 0 to 82460

Data columns (total 25 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 Nombre Doc 82461 non-null object

1 Número del Documento 82461 non-null int64

2 Código del Cliente 82461 non-null object

3 Nombre del Cliente 82461 non-null object

4 Cod. Vendedor 82461 non-null object

5 Nombre del Vendedor 79797 non-null object

6 Fecha 82461 non-null datetime64[ns]

7 Dia 82461 non-null int64

8 Mes 82461 non-null int64

9 Año 82461 non-null int64

10 Afecto 82461 non-null int64

11 Total 82461 non-null int64

12 RUT Cliente 82461 non-null object

13 Cod. Producto 82461 non-null int64

14 Desc. Producto 82461 non-null object

15 Cantidad 82461 non-null float64

16 Precio Unitario 82461 non-null int64

17 Total Detalle 82461 non-null int64

18 Total Neto Linea 82461 non-null int64

19 Descripción Detallada 80788 non-null object

20 Costo Vigente 82461 non-null float64

21 Nom Tipo Doc Asociado 75185 non-null object

22 Num Tipo Doc Asociado 75185 non-null object

23 Fecha del Doc Asociado 75185 non-null datetime64[ns]

24 Línea 82461 non-null int64

dtypes: datetime64[ns](2), float64(2), int64(11), object(10)

memory usage: 15.7+ MB

TABLA: VENTAS 2019 (df\_Ventas\_2019)

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 73169 entries, 0 to 73168

Data columns (total 25 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 Nombre Doc 73169 non-null object

1 Número del Documento 73169 non-null int64

2 Código del Cliente 73169 non-null object

3 Nombre del Cliente 73169 non-null object

4 Cod. Vendedor 73169 non-null object

5 Nombre del Vendedor 73165 non-null object

6 Fecha 73169 non-null datetime64[ns]

7 Dia 73169 non-null int64

8 Mes 73169 non-null int64

9 Año 73169 non-null int64

10 Afecto 73169 non-null int64

11 Total 73169 non-null int64

12 RUT Cliente 73169 non-null object

13 Cod. Producto 73169 non-null int64

14 Desc. Producto 73169 non-null object

15 Cantidad 73169 non-null int64

16 Precio Unitario 73169 non-null int64

17 Total Detalle 73169 non-null int64

18 Total Neto Linea 73169 non-null int64

19 Descripción Detallada 71675 non-null object

20 Costo Vigente 73169 non-null float64

21 Nom Tipo Doc Asociado 66512 non-null object

22 Num Tipo Doc Asociado 66512 non-null object

23 Fecha del Doc Asociado 66512 non-null datetime64[ns]

24 Línea 73169 non-null int64

dtypes: datetime64[ns](2), float64(1), int64(12), object(10)

memory usage: 14.0+ MB

TABLA: VENTAS 2020 (df\_Ventas\_2020)

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 50781 entries, 0 to 50780

Data columns (total 25 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 Nombre Doc 50781 non-null object

1 Número del Documento 50781 non-null int64

2 Código del Cliente 50781 non-null object

3 Nombre del Cliente 50781 non-null object

4 Cod. Vendedor 50781 non-null object

5 Nombre del Vendedor 50779 non-null object

6 Fecha 50781 non-null datetime64[ns]

7 Dia 50781 non-null int64

8 Mes 50781 non-null int64

9 Año 50781 non-null int64

10 Afecto 50781 non-null int64

11 Total 50781 non-null int64

12 RUT Cliente 50781 non-null object

13 Cod. Producto 50781 non-null int64

14 Desc. Producto 50781 non-null object

15 Cantidad 50781 non-null int64

16 Precio Unitario 50781 non-null int64

17 Total Detalle 50781 non-null int64

18 Total Neto Linea 50781 non-null int64

19 Descripción Detallada 50084 non-null object

20 Costo Vigente 50781 non-null float64

21 Nom Tipo Doc Asociado 43623 non-null object

22 Num Tipo Doc Asociado 43623 non-null object

23 Fecha del Doc Asociado 43623 non-null datetime64[ns]

24 Línea 50781 non-null int64

dtypes: datetime64[ns](2), float64(1), int64(12), object(10)

memory usage: 9.7+ MB

TABLA: VENTAS COMPLETO (df\_ventas\_total)

<class 'pandas.core.frame.DataFrame'>

Int64Index: 206411 entries, 0 to 50780

Data columns (total 25 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 Nombre Doc 206411 non-null object

1 Número del Documento 206411 non-null int64

2 Código del Cliente 206411 non-null object

3 Nombre del Cliente 206411 non-null object

4 Cod. Vendedor 206411 non-null object

5 Nombre del Vendedor 203741 non-null object

6 Fecha 206411 non-null datetime64[ns]

7 Dia 206411 non-null int64

8 Mes 206411 non-null int64

9 Año 206411 non-null int64

10 Afecto 206411 non-null int64

11 Total 206411 non-null int64

12 RUT Cliente 206411 non-null object

13 Cod. Producto 206411 non-null int64

14 Desc. Producto 206411 non-null object

15 Cantidad 206411 non-null float64

16 Precio Unitario 206411 non-null int64

17 Total Detalle 206411 non-null int64

18 Total Neto Linea 206411 non-null int64

19 Descripción Detallada 202547 non-null object

20 Costo Vigente 206411 non-null float64

21 Nom Tipo Doc Asociado 185320 non-null object

22 Num Tipo Doc Asociado 185320 non-null object

23 Fecha del Doc Asociado 185320 non-null datetime64[ns]

24 Línea 206411 non-null int64

dtypes: datetime64[ns](2), float64(2), int64(11), object(10)

memory usage: 40.9+ MB

# HIPOTESIS

Productos más vendidos

Mejores vendedores

Mejores meses del año

Top Ventas

Mejores clientes

Comportamiento futuro de la venta

Inferir que productos son susceptibles a descuentos por temporadas

Describir si el mercado dependiendo del rubro varia con respecto las estaciones climáticas

Cuáles son las categorías de productos más vendidos

Definir el Top ten de clientes, productos y proveedores

Adicionalmente podrían surgir otras interrogantes que puedan ser de utilidad al modelo las cuales podrían ser descritas en virtud de mejorar el análisis general de la información.