

**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

CAPTURAS DE PANTALLA PYTHON BASE DE DATOS ORIGINAL

```
import pandas as pd
url = "https://raw.githubusercontent.com/johandrodriguez23/Mineria-de-Datos/main/ETL-ECommerce.csv"
df = pd.read_csv(url, encoding='unicode_escape')
```

df

	IDOrdenVenta	FechaDIAOrden	FechaMESOrden	FechaAÑOOrden	FechaDIAVenta	FechaMESVenta	FechaAÑOVenta	Antigüedad	ModoEntrega	CategoríaProducto	CostoEnvío	Prioridad	IDCliente	Nombre del Cliente	Rama	Ciudad	Departamento	País	Continente	
AU-2015-1	9	11	2015	17	11	2015	8	Primera Clase	Auto & Accesorios	Reproductores de Sonido para Carro	140	60	Medium	LS-001	Lane Daniels	Consumer	Brisbane	Queensland	Australia	Oceanic
AU-2015-2	30	6	2015	2	7	2015	2	Primera Clase	Auto & Accesorios	Altavoces para Carro	211	20	Medium	IZ-002	Alvarado Kitz	Home Office	Berlin	Berlin	Germany	Central
AU-2015-3	5	12	2015	13	12	2015	8	Primera Clase	Auto & Accesorios	Car Body Covers	117	12	Critical	EN-003	Moon Waleen	Consumer	Portlouis	Wellington	New Zealand	Oceanic
AU-2015-4	9	5	2015	16	5	2015	7	Primera Clase	Auto & Accesorios	Car & Motos Care	118	62	High	AN-004	Sanchez Bergman	Corporate	Kabul	Kabul	Afghanistan	Central Asia
AU-2015-5	9	7	2015	18	7	2015	9	Primera Clase	Auto & Accesorios	Llanta	250	0	Critical	ON-005	Rove Jackson	Corporate	Townsville	Queensland	Australia	Oceanic
FA-2015-30771	21	1	2015	27	1	2015	6	Standard Class	Fashion	Sports Wear	85	70	Medium	IN-0040977	Welch Felix	Corporate	Pasadena	Texas	United States	Central
FA-2015-30772	22	6	2015	24	6	2015	2	Standard Class	Fashion	Sports Wear	85	25	Medium	TT-0040978	Martinez Arnet	Corporate	Harare	Harare	Zimbabwe	Africa
FA-2015-30773	1	1	2015	7	1	2015	6	Standard Class	Fashion	Sports Wear	85	8	Medium	ON-0040979	Mccoy Doston	Home Office	Townsville	Queensland	Australia	Oceanic
FA-2015-30774	7	12	2015	14	12	2015	7	Standard Class	Fashion	Sports Wear	85	83	Medium	RN-0040980	Bentley Zypem	Consumer	Houston	Texas	United States	Central
FA-2015-30775	1	12	2015	6	12	2015	5	Standard Class	Fashion	Sports Wear	85	83	Medium	RZ-0040981	Mckure Schwarz	Home Office	Valinhos	São Paulo	Brazil	South

51290 rows x 25 columns

Utilizamos el comando path para guardar los cambios

```
[35] path = "C:\ETL-ECommerce"
```

```
[36] df.to_csv(path)
```

Eliminación de columnas que no vamos a utilizar

```
df.drop(columns = ['FechaDIAOrden', 'FechaDIAVenta', 'Rama'], inplace=True)
```

IDOrdenVenta	FechaMESOrden	FechaAÑOOrden	FechaMESVenta	FechaAÑOVenta	Antigüedad	ModoEntrega	CategoríaProducto	Producto	Ventas	GananciaVenta	CostoEnvío	Prioridad	IDCliente
AU-2015-1	9	11	17	11	8	Primera Clase	Auto & Accesorios	Reproductores de Sonido para Carro	140	2	5%	Medium	LS-001
AU-2015-2	30	6	2	7	2	Primera Clase	Auto & Accesorios	Altavoces para Carro	211	3	3%	Medium	IZ-002
AU-2015-3	5	12	13	12	8	Primera Clase	Auto & Accesorios	Car Body Covers	117	5	1%	Critical	EN-003
AU-2015-4	9	5	16	5	7	Primera Clase	Auto & Accesorios	Car & Motos Care	118	2	5%	High	AN-004
AU-2015-5	9	7	18	7	9	Primera Clase	Auto & Accesorios	Llanta	250	1	4%	Critical	ON-005
FA-2015-30771	21	1	27	1	6	Standard Class	Fashion	Sports Wear	85	5	4%	Medium	IN-0040977
FA-2015-30772	22	6	24	6	2	Standard Class	Fashion	Sports Wear	85	1	3%	Medium	TT-0040978
FA-2015-30773	1	1	7	1	6	Standard Class	Fashion	Sports Wear	85	1	5%	Medium	ON-0040979
FA-2015-30774	7	12	14	12	7	Standard Class	Fashion	Sports Wear	85	3	4%	Medium	RN-0040980
FA-2015-30775	1	12	6	12	5	Standard Class	Fashion	Sports Wear	85	3	3%	Medium	RZ-0040981

51290 rows x 22 columns

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

D	E	F	G	H	I	J	N	O	P	Q	R	S	T
FechaANOOrden	FechaDIAVe	FechaMESVe	FechaANOVenta	Antigüedad	ModoEntreg	CategoríaPro	PorcentajeD	GananciaVer	CostoEnvío	Prioridad	IdCliente	Nombre del	Rama
2015	17	11	2015	8	Primera Clas	Auto & Acce	5%	\$ 46	0	\$ 4	60	Medium	LS-001
2015	2	7	2015	2	Primera Clas	Auto & Acce	3%	\$ 112	1	\$ 11	20	Medium	IZ-002
2015	13	12	2015	8	Primera Clas	Auto & Acce	1%	\$ 31	15	\$ 3	12	Critical	EN-003
2015	16	5	2015	7	Primera Clas	Auto & Acce	5%	\$ 26	20	\$ 2	62	High	AN-004
2015	18	7	2015	9	Primera Clas	Auto & Acce	4%	\$ 160	0	\$ 16	0	Critical	ON-005
2015	5	3	2015	8	Primera Clas	Auto & Acce	4%	\$ 24	0	\$ 2	40	Critical	TO-006
2015	10	4	2015	1	Primera Clas	Auto & Acce	5%	\$ 54	0	\$ 5	40	High	OM-007
2015	6	4	2015	7	Primera Clas	Auto & Acce	2%	\$ 22	60	\$ 2	26	Critical	AN-008
2015	16	2	2015	7	Primera Clas	Auto & Acce	3%	\$ 116	35	\$ 11	64	Critical	EN-009
2015	1	5	2015	10	Primera Clas	Auto & Acce	2%	\$ 54	40	\$ 5	44	Critical	TT-0010
2015	26	11	2015	10	Primera Clas	Auto & Acce	1%	\$ 122	56	\$ 12	26	Critical	ED-0011
2015	2	9	2015	1	Primera Clas	Auto & Acce	4%	\$ 18	28	\$ 1	83	High	ON-0012
2015	16	7	2015	7	Primera Clas	Auto & Acce	2%	\$ 35	64	\$ 3	56	Critical	WN-0013
2015	27	7	2015	5	Primera Clas	Auto & Acce	4%	\$ 140	0	\$ 14	0	High	AN-0014
2015	21	10	2015	9	Primera Clas	Auto & Acce	1%	\$ 18	NULL	\$ 1	80	Medium	EY-0015
2015	5	3	2015	10	Primera Clas	Auto & Acce	1%	\$ 27	0	\$ 2	70	Critical	RN-0016
2015	8	5	2015	4	Primera Clas	Auto & Acce	5%	\$ 22	60	\$ 2	26	High	CK-0017
2015	19	6	2015	7	Primera Clas	Auto & Acce	5%	\$ 93	25	\$ 9	33	High	RE-0018
2015	20	5	2015	7	Primera Clas	Auto & Acce	5%	\$ 46	0	\$ 4	60	Critical	ON-0019
2015	2	8	2015	6	Primera Clas	Auto & Acce	2%	\$ 122	56	\$ 12	26	Critical	ED-0020
2015	13	12	2015	4	Primera Clas	Auto & Acce	1%	\$ 31	15	\$ 3	12	Critical	AM-0021
2015	16	2	2015	7	Primera Clas	Auto & Acce	3%	\$ 30	92	\$ 3	9	High	KE-0022
2015	26	7	2015	6	Primera Clas	Auto & Acce	2%	\$ 150	0	\$ 15	0	High	LL-0023
2015	24	11	2015	3	Primera Clas	Auto & Acce	3%	\$ 18	0	\$ 1	80	Critical	CO-0024

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

IDOrden	Venta	FechaDIAOrden	FechaMESOrden	FechaDIAVenta	FechaMESVenta	Antigüedad	ModoEntrega	CategoríaProducto	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='O')

	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

Instrucción 0

df.describe(include='O')

	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

```
df.isna().sum()
```

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
Departamento	0