<div style="border:1px solid #ccc; border-radius:8px; padding:8px; text-align:center;">* Ecommerce_project *</div>

<div style="border:1px solid #ccc; border-radius:8px; padding:8px; text-align:center;">* objective *</div>

- To analysis ecommerce order data

<div style="border:1px solid #ccc; border-radius:8px; padding:8px; text-align:center;">* important libraries *</div>

1. Pandas
2. Matplotlib
3. seaborn
4. Pandas Profiling
5. Numpy

```python
import pandas as pd
Ecommerce=pd.read_csv("/content/ecommerce_project.csv")
print(Ecommerce)
```

```
13284  37178   CA-2013-133935  19-09-2013  23-09-2013  Standard Class
13285  26149    IN-2011-67616  15-12-2011  20-12-2011    Second Class
13286  29909    IN-2014-78522  25-12-2014  29-12-2014    Second Class
13287   3583   MX-2011-139136  11-11-2011  16-11-2011    Standard Cla

        Customer_ID    Customer_Name      Segment             City  \
0         RH-19495      Rick Hansen     Consumer   New York City
1         JR-16210     Justin Ritter   Corporate       Wollongong
2         CR-12730      Craig Reiter    Consumer         Brisbane
3         KM-16375   Katherine Murray  Home Office         Berlin
4          RH-9495      Rick Hansen     Consumer            Dakar
...            ...              ...         ...              ...
13283     AG-10675      Anna Gayman     Consumer        Traralgon
13284     JW-15220        Jane Waco    Corporate        San Diego
13285     NC-18625       Noah Childs   Corporate       Wollongong
13286     DW-13195     David Wiener    Corporate        Toowoomba
13287          NaN              NaN         NaN              NaN

               State  ...      Product_ID        Category  Sub_Category  \
0           New York  ...  TEC-AC-10003033      Technology    Accessories
1    New South Wales  ...  FUR-CH-10003950       Furniture         Chairs
2         Queensland  ...  TEC-PH-10004664      Technology         Phones
3             Berlin  ...  TEC-PH-10004583      Technology         Phones
4              Dakar  ...  TEC-SHA-10000501     Technology        Copiers
...              ...  ...             ...             ...            ...
13283       Victoria  ...  FUR-CH-10001695       Furniture         Chairs
13284     California  ...  OFF-PA-10000100  Office Supplies          Paper
13285  New South Wales ...  OFF-ST-10001631  Office Supplies        Storage
13286     Queensland  ...  FUR-CH-10004580       Furniture         Chairs
13287            NaN  ...             NaN             NaN            NaN

                                 Product_Name      Sales  Quantity  \
0      Plantronics CS510 - Over-the-Head monaural Wir...  2309.650       7.0
1          Novimex Executive Leather Armchair, Black  3709.395       9.0
2              Nokia Smart Phone, with Caller ID  5175.171       9.0
3                 Motorola Smart Phone, Cordless  2892.510       5.0
4                 Sharp Wireless Fax, High-Speed  2832.960       8.0
...                                          ...        ...       ...
13283        Harbour Creations Chairmat, Set of Two   334.125       5.0
13284                                  Xerox 1945   368.910       9.0
13285                           Tenex Trays, Blue   341.523       7.0
13286            Office Star Bag Chairs, Adjustable   208.548       4.0
13287                                         NaN        NaN       NaN

       Discount  Shipping_Cost    Profit  Order_Priority
0           0.0         933.57  762.1845        Critical
1           0.1         923.63 -288.7650        Critical
2           0.1         915.49  919.9710          Medium
3           0.1         910.16  -96.5400          Medium
4           0.0         903.04  311.5200        Critical
...         ...            ...       ...             ...
13283       0.1          23.30  -33.5250          Medium
13284       0.0          23.30  180.7659            High
13285       0.1          23.29  129.0030          Medium
13286       0.1          23.29   48.5880          Medium
13287       NaN            NaN       NaN             NaN

[13288 rows x 24 columns]
```

```python
Ecommerce.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 13288 entries, 0 to 13287
Data columns (total 24 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   ROW_ID         13288 non-null  int64
 1   Order_ID       13288 non-null  object
 2   Order_Date     13288 non-null  object
 3   Ship_Date      13288 non-null  object
 4   Ship_Mode      13288 non-null  object
 5   Customer_ID    13287 non-null  object
 6   Customer_Name  13287 non-null  object
 7   Segment        13287 non-null  object
 8   City           13287 non-null  object
 9   State          13287 non-null  object
 10  Country        13287 non-null  object
 11  Market         13287 non-null  object
 12  Postal_Code    2247 non-null   float64
 13  Region         13287 non-null  object
 14  Product_ID     13287 non-null  object
 15  Category       13287 non-null  object
 16  Sub_Category   13287 non-null  object
 17  Product_Name   13287 non-null  object
 18  Sales          13287 non-null  float64
 19  Quantity       13287 non-null  float64
 20  Discount       13287 non-null  float64
 21  Shipping_Cost  13287 non-null  float64
 22  Profit         13287 non-null  float64
 23  Order_Priority 13287 non-null  object
dtypes: float64(6), int64(1), object(17)
memory usage: 2.4+ MB
```

```
Ecommerce["Order_Date"]=pd.to_datetime(Ecommerce["Order_Date"],errors="coerce")
```

```
/tmp/ipython-input-638580473.py:1: UserWarning: Parsing dates in %d-%m-%Y format when dayfirst=False (the default) was specified. Pass `dayfirst=True` or specify a format to silence this warning.
  Ecommerce["Order_Date"]=pd.to_datetime(Ecommerce["Order_Date"],errors="coerce")
```

```
Ecommerce["Ship_Date"]=pd.to_datetime(Ecommerce["Ship_Date"],errors="coerce")
```

```
/tmp/ipython-input-3371376093.py:1: UserWarning: Parsing dates in %d-%m-%Y format when dayfirst=False (the default) was specified. Pass `dayfirst=True` or specify a format to silence this warning.
  Ecommerce["Ship_Date"]=pd.to_datetime(Ecommerce["Ship_Date"],errors="coerce")
```

```
Ecommerce["Postal_Code"]=pd.to_numeric(Ecommerce["Postal_Code"],errors="coerce")
```

```
Ecommerce["Sales"]=pd.to_numeric(Ecommerce["Sales"],errors="coerce")
```

```
Ecommerce["Discount"]=pd.to_numeric(Ecommerce["Discount"],errors="coerce")
```

```
Ecommerce["Shipping_Cost"]=pd.to_numeric(Ecommerce["Shipping_Cost"],errors="coerce")
```

```
Ecommerce["Profit"]=pd.to_numeric(Ecommerce["Profit"],errors="coerce")
```

```
* Data Cleaning *
```

```
Ecommerce.duplicated()
```

	0
0	False
1	False
2	False
3	False
4	False
...	...
13283	False
13284	False
13285	False
13286	False
13287	False

13288 rows × 1 columns

dtype: bool

```
Ecommerce.duplicated().sum()
```

np.int64(0)

```
Ecommerce.isnull()
```

	ROW_ID	Order_ID	Order_Date	Ship_Date	Ship_Mode	Customer_ID	Customer_Name	Segment	City	State	...	Product_ID	Category	Sub_Category	Product_Name	Sales	Quantity	Discount	Shipping_Cost	Profit	Order_Priority
0	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
31106	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
31107	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
31108	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
31109	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
31110	False	False	False	False	False	False	False	False	False	False	...	False	True	True	True	True	True	True	True	True	True

31111 rows × 24 columns

```
Ecommerce.isnull().sum()
```

	0
ROW_ID	0
Order_ID	0
Order_Date	0
Ship_Date	0
Ship_Mode	0
Customer_ID	1
Customer_Name	1
Segment	1
City	1
State	1
Country	1
Market	1
Postal_Code	11041
Region	1
Product_ID	1
Category	1
Sub_Category	1
Product_Name	1
Sales	1
Quantity	1
Discount	1
Shipping_Cost	1
Profit	1
Order_Priority	1

dtype: int64

Ecommerce.drop_duplicates()

	ROW_ID	Order_ID	Order_Date	Ship_Date	Ship_Mode	Customer_ID	Customer_Name	Segment	City	State	...	Product_ID	Category	Sub_Category	Product_Name	Sales	Quantity	Discount	Shipping_Cost	Profit	Order_Priority
0	26341	CA-2012-124891	2012-07-31	2012-07-31	Same Day	RH-19495	Rick Hansen	Consumer	New York City	New York	...	TEC-AC-10003033	Technology	Accessories	Plantronics CS510 - Over-the-Head monaural Wir...	2309.650	7.0	0.0	933.57	762.1845	Critical
1	26341	IN-2013-77878	2013-02-05	2013-02-07	Second Class	JR-16210	Justin Ritter	Corporate	Wollongong	New South Wales	...	FUR-CH-10003950	Furniture	Chairs	Novimex Executive Leather Armchair, Black	3709.395	9.0	0.1	923.63	-288.7650	Critical
2	25330	IN-2013-71249	2013-10-17	2013-10-18	First Class	CR-12730	Craig Reiter	Consumer	Brisbane	Queensland	...	TEC-PH-10004664	Technology	Phones	Nokia Smart Phone, with Caller ID	5175.171	9.0	0.1	915.49	919.9710	Medium
3	13524	ES-2013-1579342	2013-01-28	2013-01-30	First Class	KM-16375	Katherine Murray	Home Office	Berlin	Berlin	...	TEC-PH-10004583	Technology	Phones	Motorola Smart Phone, Cordless	2892.510	5.0	0.1	910.16	-96.5400	Medium
4	47221	SG-2013-4320	2013-11-05	2013-11-06	Same Day	RH-9495	Rick Hansen	Consumer	Dakar	Dakar	...	TEC-SHA-10000501	Technology	Copiers	Sharp Wireless Fax, High-Speed	2832.960	8.0	0.0	903.04	311.5200	Critical
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
13283	23773	ID-2012-37047	2012-11-22	2012-11-29	Standard Class	AG-10675	Anna Gayman	Consumer	Traralgon	Victoria	...	FUR-CH-10001695	Furniture	Chairs	Harbour Creations Chairmat, Set of Two	334.125	5.0	0.1	23.30	-33.5250	Medium
13284	37178	CA-2013-133935	2013-09-19	2013-09-23	Standard Class	JW-15220	Jane Waco	Corporate	San Diego	California	...	OFF-PA-10000100	Office Supplies	Paper	Xerox 1945	368.910	9.0	0.0	23.30	180.7659	High
13285	26149	IN-2011-67616	2011-12-15	2011-12-20	Second Class	NC-18625	Noah Childs	Corporate	Wollongong	New South Wales	...	OFF-ST-10001631	Office Supplies	Storage	Tenex Trays, Blue	341.523	7.0	0.1	23.29	129.0030	Medium
13286	29909	IN-2014-78522	2014-12-25	2014-12-29	Second Class	DW-13195	David Wiener	Corporate	Toowoomba	Queensland	...	FUR-CH-10004580	Furniture	Chairs	Office Star Bag Chairs, Adjustable	208.548	4.0	0.1	23.29	48.5880	Medium
13287	3583	MX-2011-139136	2011-11-11	2011-11-16	Standard Cla	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

13288 rows × 24 columns

Ecommerce.dropna()

	ROW_ID	Order_ID	Order_Date	Ship_Date	Ship_Mode	Customer_ID	Customer_Name	Segment	City	State	...	Product_ID	Category	Sub_Category	Product_Name	Sales	Quantity	Discount	Shipping_Cost	Profit	Order_Priority	
0	26341	CA-2012-124891	2012-07-31	2012-07-31	Same Day	RH-19495	Rick Hansen	Consumer	New York City	New York	...	TEC-AC-10003033	Technology	Accessories	Plantronics CS510 - Over-the-Head monaural Wir...	2309.6500	7.0	0.00	933.57	762.1845	Critical	
8	40155	CA-2014-135909	2014-10-14	2014-10-21	Standard Class	JW-15220	Jane Waco	Corporate	Sacramento	California	...	OFF-BI-10003527	Office Supplies	Binders	Fellowes PB500 Electric Punch Plastic Comb Bin...	5083.9600	5.0	0.20	867.69	1906.4850	Low	
9	40936	CA-2012-116638	2012-01-28	2012-01-31	Second Class	JH-15985	Joseph Holt	Consumer	Concord	North Carolina	...	FUR-TA-10000198	Furniture	Tables	Chromcraft Bull-Nose Wood Oval Conference Tabl...	4297.6440	13.0	0.40	865.74	-1862.3124	Critical	
10	34577	CA-2011-102988	2011-04-05	2011-04-09	Second Class	GM-14695	Greg Maxwell	Corporate	Alexandria	Virginia	...	OFF-SU-10002881	Office Supplies	Supplies	Martin Yale Chadless Opener Electric Letter Op...	4164.0500	5.0	0.00	846.54	83.2810	High	
16	36178	CA-2014-143567	2014-11-03	2014-11-06	Second Class	TB-21175	Thomas Boland	Corporate	Henderson	Kentucky	...	TEC-AC-10004145	Technology	Accessories	Logitech diNovo Edge Keyboard	2249.9100	9.0	0.00	780.70	517.4793	Critical	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
13248	37342	CA-2012-113215	2012-09-03	2012-09-08	Standard Class	CP-12085	Cathy Prescott	Corporate	Tucson	Arizona	...	FUR-FU-10002937	Furniture	Furnishings	GE 48" Fluorescent Tube, Cool White Energy Sav...	238.1520	3.0	0.20	23.39	89.3070	Medium	
13264	38622	US-2014-128447	2014-11-11	2014-11-18	Standard Class	MC-17845	Michael Chen	Consumer	Pasco	Washington	...	OFF-AP-10004540	Office Supplies	Appliances	Eureka The Boss Lite 10-Amp Upright Vacuum, Blue	400.8000	5.0	0.00	23.36	112.2240	Medium	
13267	33700	US-2013-110170	2013-09-28	2013-10-04	Standard Class	HM-14860	Harry Marie	Corporate	Huntsville	Texas	...	FUR-BO-10000780	Furniture	Bookcases	O'Sullivan Plantations 2-Door Library in Landv...	956.6648	7.0	0.32	23.35	-225.0976	Medium	
13275	34788	CA-2012-157322	2012-07-02	2012-07-06	Standard Class	RH-19600	Rob Haberlin	Consumer	Carol Stream	Illinois	...	OFF-ST-10003208	Office Supplies	Storage	Adjustable Depth Letter/Legal Cart	435.5040	3.0	0.20	23.32	48.9942	Medium	
13284	37178	CA-2013-133935	2013-09-19	2013-09-23	Standard Class	JW-15220	Jane Waco	Corporate	San Diego	California	...	OFF-PA-10000100	Office Supplies	Paper	Xerox 1945	368.9100	9.0	0.00	23.30	180.7659	High	

2247 rows × 24 columns

* Data Analysis *

```
describe=Ecommerce.describe()
print(describe)
```

```
              ROW_ID                     Order_Date  \
count   13288.000000                          13288
mean    23959.759859  2013-05-15 10:36:59.385912064
min         8.000000            2011-01-01 00:00:00
25%     12961.250000            2012-06-25 00:00:00
50%     23280.500000            2013-07-12 00:00:00
75%     34409.750000            2014-05-24 00:00:00
max     51284.000000            2014-12-31 00:00:00
std     13704.506078                            NaN

                          Ship_Date    Postal_Code          Sales  \
count                         13288    2247.000000   13287.000000
mean    2013-05-19 01:07:11.306441984   55073.732977     707.098782
min             2011-01-03 00:00:00    1040.000000      46.000000
25%             2012-06-28 00:00:00   22153.000000     284.070000
50%             2013-07-15 00:00:00   55407.000000     485.904000
75%             2014-05-28 00:00:00   90032.000000     846.606000
max             2015-01-05 00:00:00   99301.000000   22638.480000
std                             NaN   32670.456753     752.975965

            Quantity      Discount  Shipping_Cost        Profit
count   13287.000000  13287.000000   13287.000000  13287.000000
mean        4.520358      0.109747      82.986266     91.168504
min         1.000000      0.000000      23.290000  -6599.978000
25%         3.000000      0.000000      33.040000      8.497500
50%         4.000000      0.000000      51.530000     65.300000
75%         6.000000      0.200000      92.050000    164.060000
max        14.000000      0.850000     933.570000   8399.976000
std         2.534694      0.165307      90.816599    317.393382
```

```
info=Ecommerce.info()
print(info)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 13288 entries, 0 to 13287
Data columns (total 24 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   ROW_ID         13288 non-null  int64
 1   Order_ID       13288 non-null  object
 2   Order_Date     13288 non-null  datetime64[ns]
 3   Ship_Date      13288 non-null  datetime64[ns]
 4   Ship_Mode      13288 non-null  object
 5   Customer_ID    13287 non-null  object
 6   Customer_Name  13287 non-null  object
 7   Segment        13287 non-null  object
 8   City           13287 non-null  object
 9   State          13287 non-null  object
 10  Country        13287 non-null  object
 11  Market         13287 non-null  object
 12  Postal_Code    2247 non-null   float64
 13  Region         13287 non-null  object
 14  Product_ID     13287 non-null  object
 15  Category       13287 non-null  object
 16  Sub_Category   13287 non-null  object
 17  Product_Name   13287 non-null  object
 18  Sales          13287 non-null  float64
 19  Quantity       13287 non-null  float64
```

```
 20  Discount       13287 non-null  float64
 21  Shipping_Cost  13287 non-null  float64
 22  Profit         13287 non-null  float64
 23  Order_Priority 13287 non-null  object
dtypes: datetime64[ns](2), float64(6), int64(1), object(15)
memory usage: 2.4+ MB
None
```

```
sample=Ecommerce.sample()
print(sample)

       ROW_ID      Order_ID Order_Date  Ship_Date Ship_Mode Customer_ID  \
514     20851  IN-2013-71333 2013-09-13 2013-09-13  Same Day    JE-15610

    Customer_Name    Segment      City         State  ...   Product_ID  \
514       Jim Epp  Corporate  Amravati  Maharashtra  ... TEC-PH-10004006

       Category Sub_Category           Product_Name  Sales Quantity  \
514  Technology       Phones  Nokia Audio Dock, Full Size  840.15     5.0

     Discount Shipping_Cost Profit  Order_Priority
514       0.0        286.67  260.4        Critical

[1 rows x 24 columns]
```

```
sorting_values=Ecommerce.sort_values(by="Order_Date",ascending=True)
print(sorting_values)

       ROW_ID       Order_ID Order_Date  Ship_Date      Ship_Mode  \
9352    42433    AG-2011-2040 2011-01-01 2011-01-06  Standard Class
12965   49550    CA-2011-1510 2011-01-02 2011-01-06  Standard Class
9787    26362    IN-2011-65159 2011-01-03 2011-01-07   Second Class
3904    46681    IZ-2011-4680 2011-01-03 2011-01-07  Standard Class
2709    30450    ID-2011-80230 2011-01-03 2011-01-09  Standard Class
...       ...            ...        ...        ...            ...
6560    26335   IN-2014-30390 2014-12-31 2015-01-03    First Class
6607    15693 ES-2014-1695428 2014-12-31 2015-01-02   Second Class
13007   22880   IN-2014-78627 2014-12-31 2015-01-04  Standard Class
656      1783  MX-2014-116267 2014-12-31 2015-01-03   Second Class
6643     1787  MX-2014-116267 2014-12-31 2015-01-03   Second Class

      Customer_ID    Customer_Name    Segment            City  \
9352     TB-11280  Toby Braunhardt   Consumer     Constantine
12965     MM-7260  Magdelene Morse   Consumer  St. Catharines
9787     LB-16735     Larry Blacks   Consumer          Manila
3904      LW-6990  Lindsay Williams  Corporate          Mosul
2709     KL-16645     Ken Lonsdale   Consumer        Papakura
...          ...              ...        ...            ...
6560     JD-16150  Justin Deggeller  Corporate        Bangkok
6607     RD-19480      Rick Duston   Consumer           Caen
13007    MG-17695   Maureen Gnade    Consumer          Jilin
656      EB-13975       Erica Bern  Corporate       São Paulo
6643     EB-13975       Erica Bern  Corporate       São Paulo

                State  ...       Product_ID        Category  Sub_Category  \
9352      Constantine  ...  OFF-TEN-10000025  Office Supplies      Storage
12965         Ontario  ...  TEC-OKI-10002750      Technology     Machines
9787   National Capital  ... OFF-ST-10003020  Office Supplies      Storage
3904          Ninawa  ...  FUR-NOV-10002791       Furniture        Chairs
2709         Auckland  ...   TEC-CO-10004182      Technology       Copiers
...             ...  ...              ...             ...          ...
6560          Bangkok  ...    FUR-CH-10003861      Furniture        Chairs
6607   Lower Normandy  ...   OFF-ST-10002159  Office Supplies      Storage
13007           Jilin  ...    TEC-AC-10003750      Technology   Accessories
656         São Paulo  ...    TEC-CO-10000137      Technology       Copiers
6643        São Paulo  ...    FUR-CH-10004844       Furniture        Chairs

                       Product_Name   Sales Quantity Discount  \
9352             Tenex Lockers, Blue  408.3000    2.0    0.000
12965      Okidata Inkjet, Wireless  314.2200    1.0    0.000
9787        Tenex Lockers, Industrial 337.6395    3.0    0.450
3904      Novimex Swivel Stool, Set of Two 666.8400  4.0  0.000
2709         Hewlett Wireless Fax, Laser  912.4560  4.0    0.400
...                           ...        ...    ...      ...
6560   Novimex Rocking Chair, Set of Two  292.7592  3.0  0.270
6607        Fellowes Lockers, Wire Frame  557.2800  3.0  0.100
13007     Memorex Memory Card, Erganomic  300.2400  3.0  0.000
656            Canon Wireless Fax, Color 1264.4660  5.0  0.002
6643       Novimex Swivel Stool, Adjustable 333.9000 3.0  0.000

      Shipping_Cost   Profit  Order_Priority
9352          35.46  106.1400         Medium
12965         24.10    3.1200         Medium
9787          33.75 -122.8005           High
3904          81.26  253.3200           High
2709         107.10 -319.4640            Low
```

```
loc=Ecommerce.loc[:,["Customer_Name","City"]]
print(loc)

        Customer_Name           City
0         Rick Hansen  New York City
1       Justin Ritter     Wollongong
2        Craig Reiter       Brisbane
3     Katherine Murray         Berlin
4         Rick Hansen          Dakar
...              ...            ...
13283    Anna Gayman      Traralgon
13284      Jane Waco      San Diego
13285     Noah Childs     Wollongong
13286    David Wiener      Toowoomba
13287            NaN            NaN

[13288 rows x 2 columns]
```

```
iloc=Ecommerce.iloc[:,1:5:1]
print(iloc)

          Order_ID Order_Date  Ship_Date      Ship_Mode
0     CA-2012-124891 2012-07-31 2012-07-31       Same Day
1      IN-2013-77878 2013-02-05 2013-02-07   Second Class
2      IN-2013-71249 2013-10-17 2013-10-18    First Class
3    ES-2013-1579342 2013-01-28 2013-01-30    First Class
4       SG-2013-4320 2013-11-05 2013-11-06       Same Day
...            ...        ...        ...            ...
13283  ID-2012-37047 2012-11-22 2012-11-29  Standard Class
13284 CA-2013-133935 2013-09-19 2013-09-23  Standard Class
13285  IN-2011-67616 2011-12-15 2011-12-20   Second Class
13286  IN-2014-78522 2014-12-25 2014-12-29   Second Class
13287 MX-2011-139136 2011-11-11 2011-11-16  Standard Cla
```

```
[13288 rows x 4 columns]
```

```
pivot_table=pd.pivot_table(Ecommerce,values="Profit",index="Order_Priority",aggfunc="mean")
print(pivot_table)
```

```
                 Profit
Order_Priority
Critical       62.963381
High           74.293889
Low            72.728055
Medium        116.493412
```

```
groupby=Ecommerce.groupby("Order_Priority")["Profit"].sum()
print(groupby)
```

```
Order_Priority
Critical    114656.31604
High        367011.81038
Low          50764.18214
Medium      678923.60632
Name: Profit, dtype: float64
```

```
sub_category_groupby=Ecommerce.groupby("Sub_Category")["Profit"].sum().reset_index()
print(sub_category_groupby)
```

```
     Sub_Category        Profit
0     Accessories  103725.24000
1      Appliances  133632.83030
2             Art   18409.04350
3         Binders   37139.58140
4        Bookcases  166306.95460
5          Chairs  133393.49220
6          Copiers  250390.83178
7        Envelopes    8809.19130
8        Fasteners     746.29080
9       Furnishings   26307.00480
10          Labels    1561.92660
11        Machines   60518.83970
12           Paper   16531.10430
13          Phones  200606.50740
14         Storage   91716.60920
15        Supplies    5062.44850
16          Tables  -43501.98150
```

```
correlation=Ecommerce["Profit"].corr(Ecommerce['Sales'])
print("correlation:",correlation)
```

```
correlation: 0.44580375788827203
```

```
* Data Visualization *
```

```
import matplotlib.pyplot as plt
```

```
plt.figure(figsize=(10,5))
plt.bar(sub_category_groupby["Sub_Category"],sub_category_groupby["Profit"])
plt.title("subcategory wise sales")
plt.xlabel("subcategory")
plt.ylabel("sales")
plt.show()
```
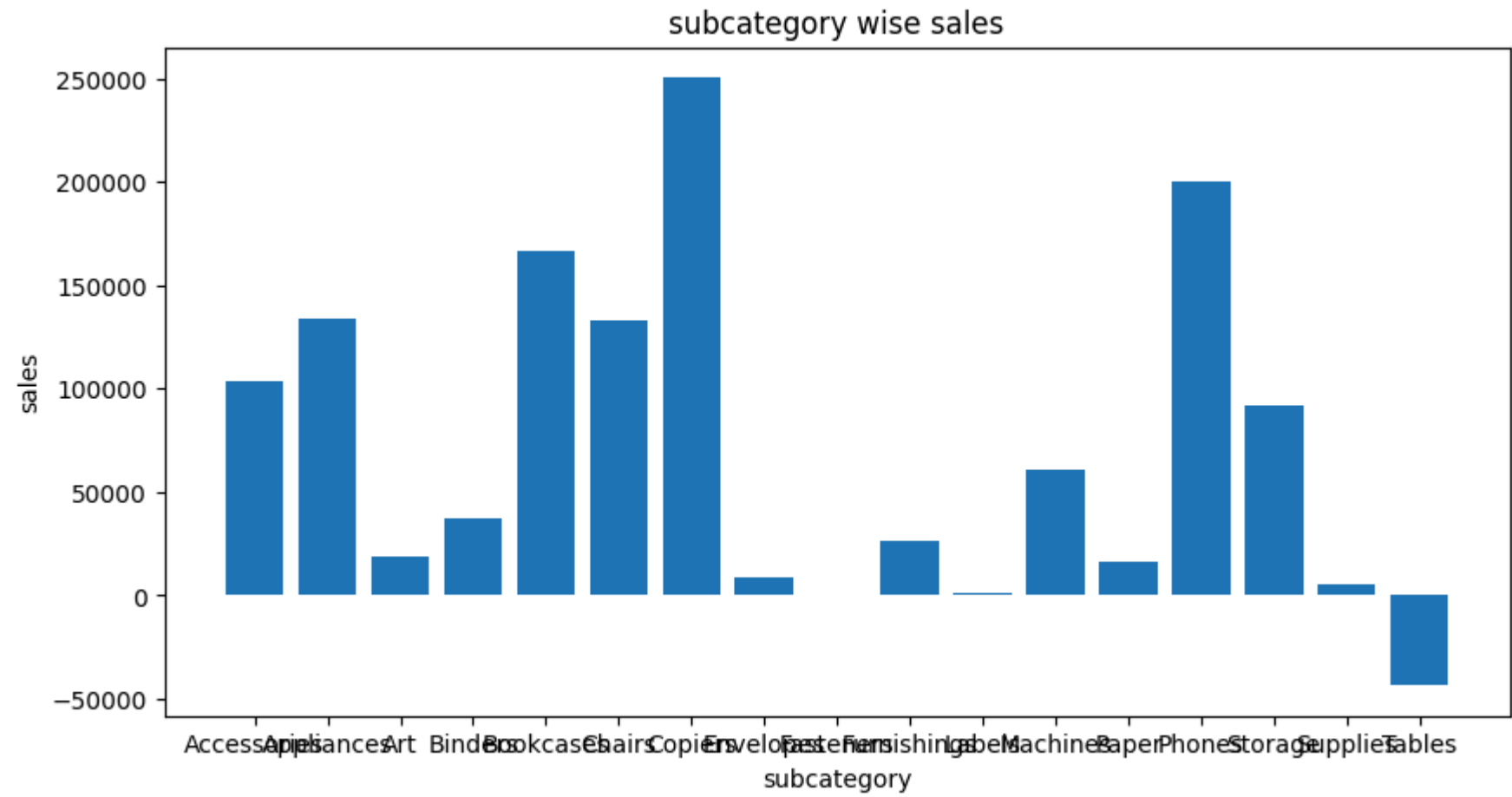


```
state_wise_sales=Ecommerce.groupby("State")["Sales"].sum().reset_index()
plt.bar(state_wise_sales["State"],state_wise_sales["Sales"])
plt.title("state_wise_sales")
plt.xlabel("State")
plt.ylabel("Sales")
plt.show()
```
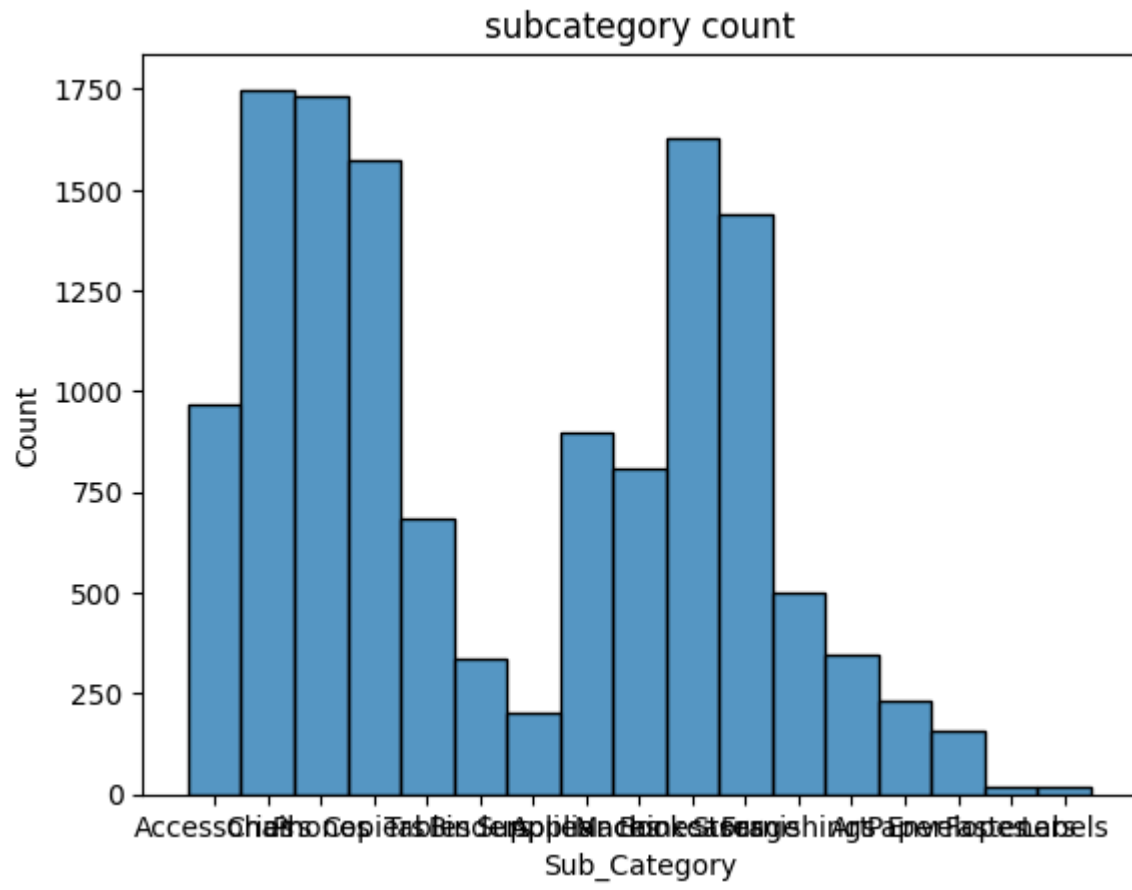
state_wise_sales
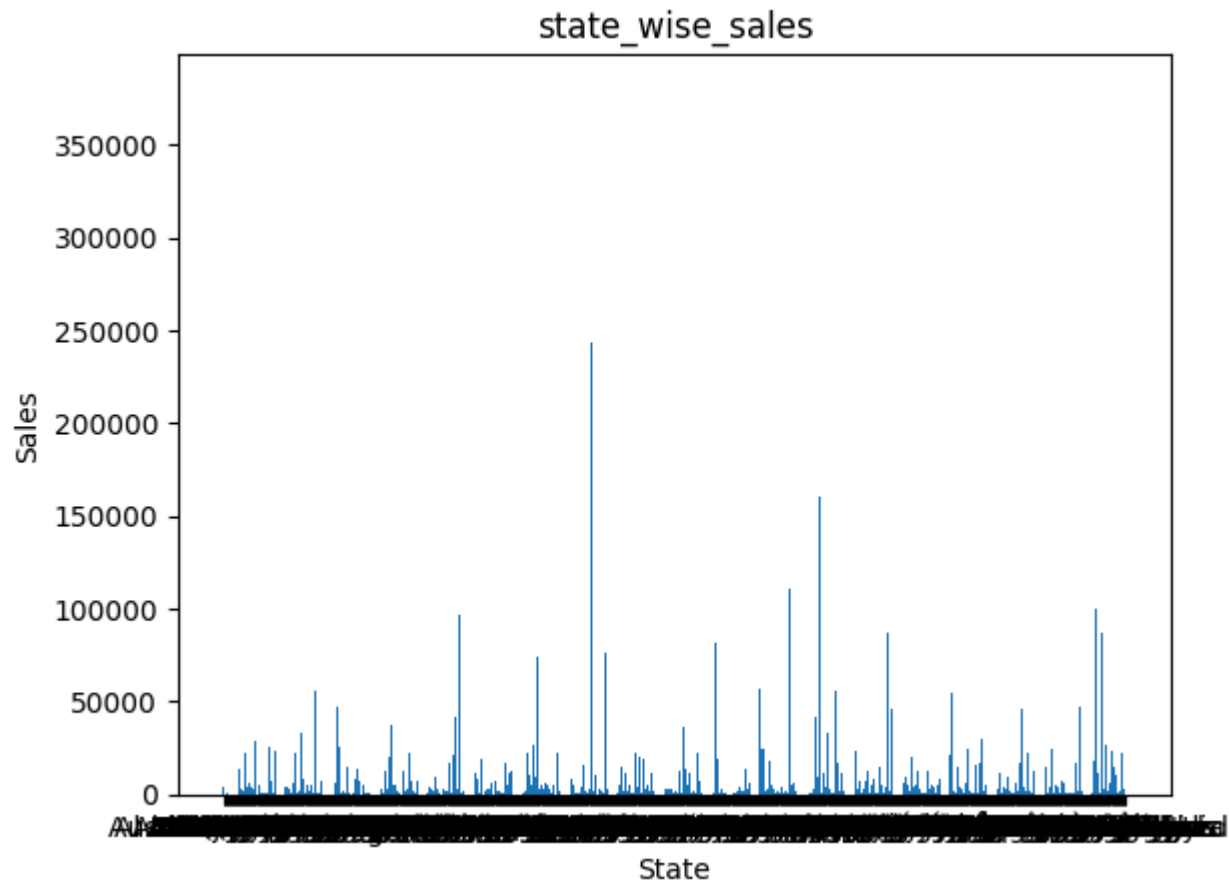

```
cols=["Sales","Quantity","Shipping_Cost","Profit"]
sns.pairplot(Ecommerce[cols])
plt.title("pairplot of columns")
plt.show()
```
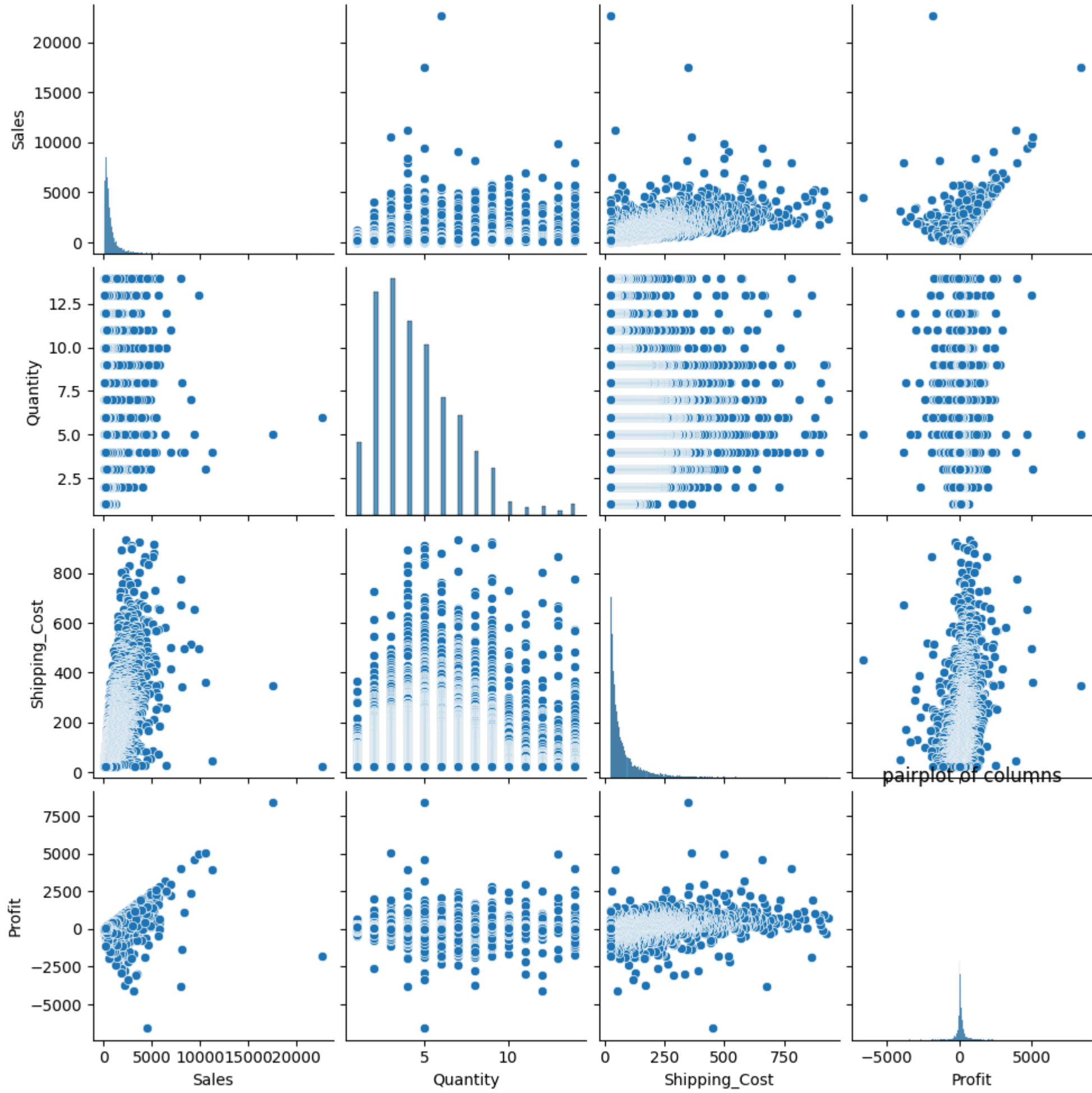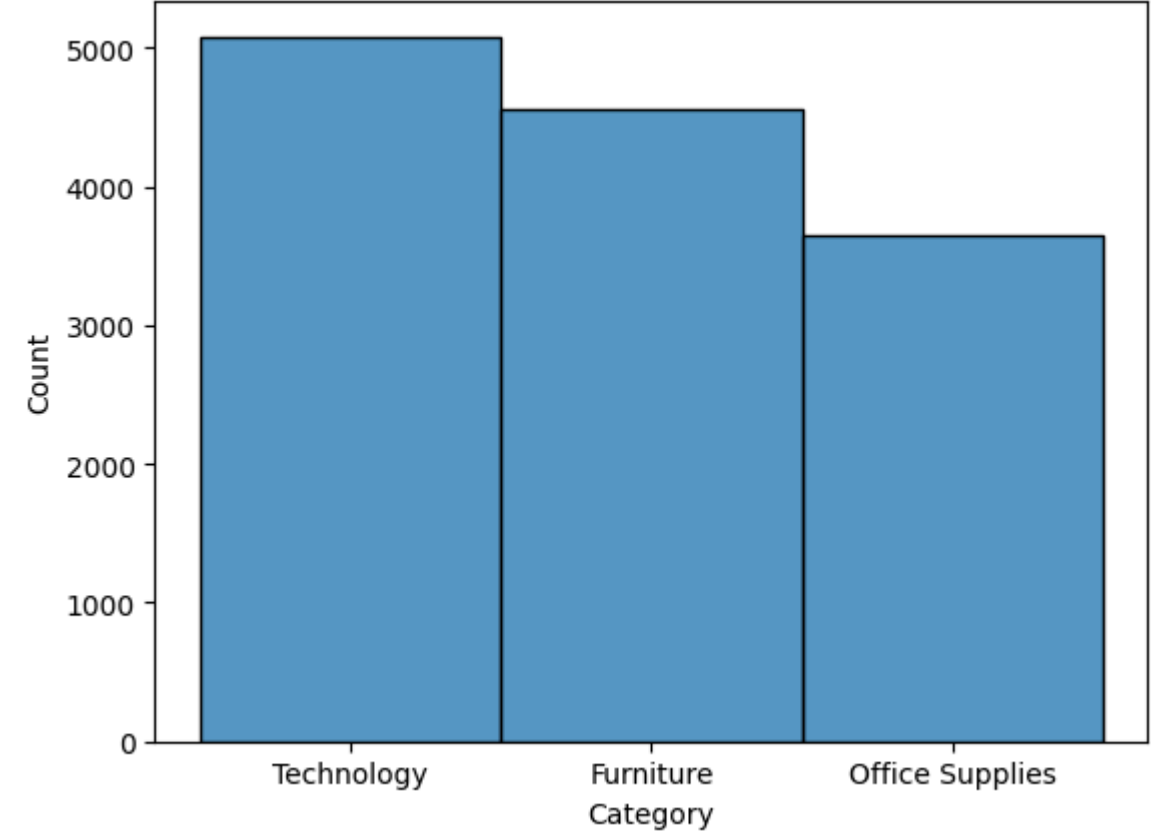


```
import seaborn as sns
sns.histplot(Ecommerce["Sub_Category"])
plt.title("subcategory count")
```
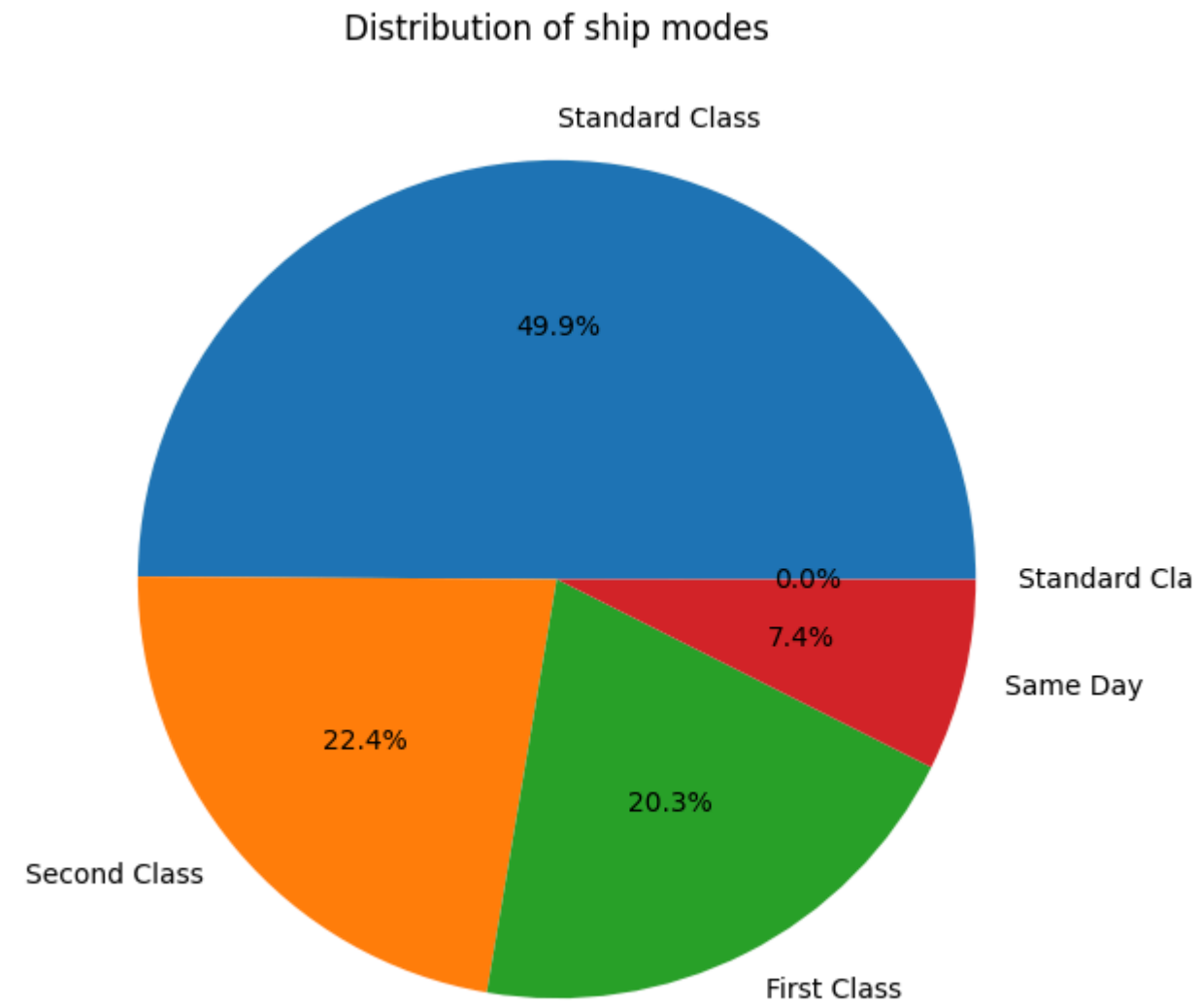
Text(0.5, 1.0, 'subcategory count')

```
sns.histplot(Ecommerce["Category"])
```
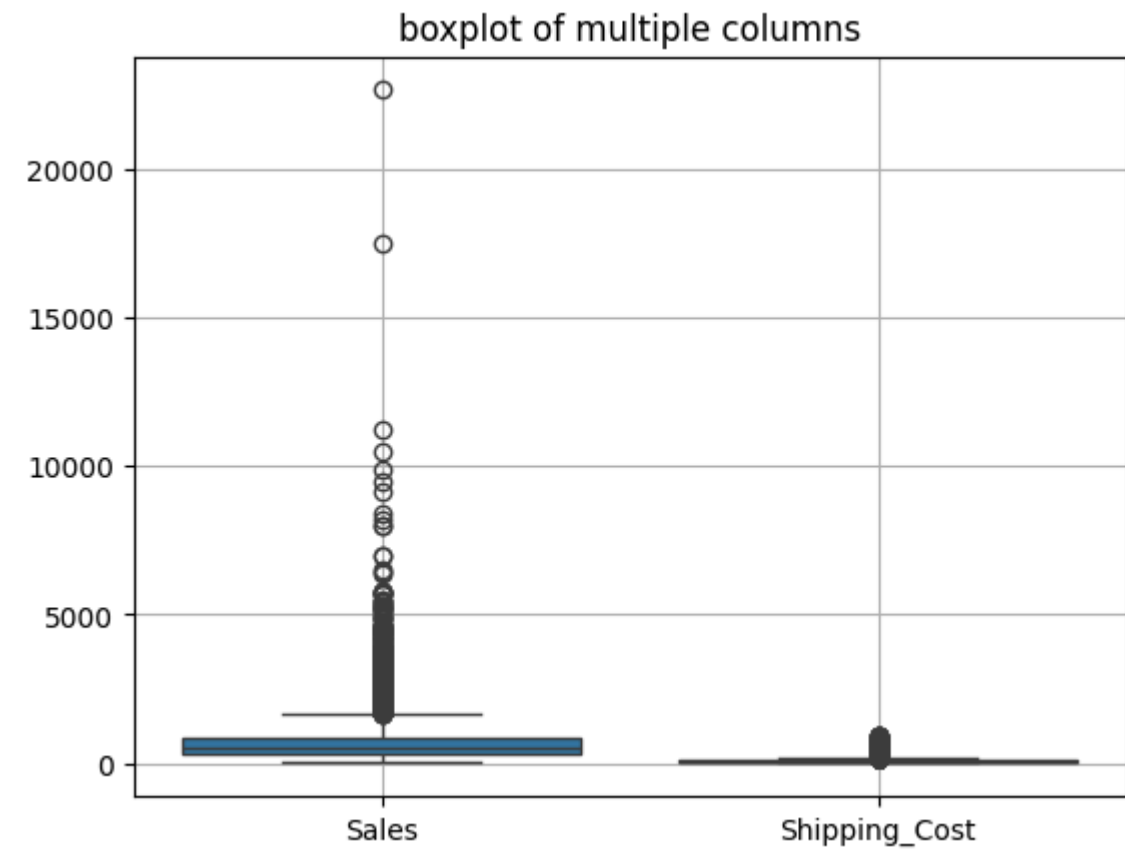
```
<Axes: xlabel='Category', ylabel='Count'>
```



```
ship_mode_value_counts=Ecommerce["Ship_Mode"].value_counts()
plt.figure(figsize=(7,7))
plt.pie(ship_mode_value_counts,labels=ship_mode_value_counts.index,autopct='%1.1f%%')
plt.title("Distribution of ship modes")
plt.show()
```



```
# to find outliers boxplot :
sns.boxplot(data=Ecommerce[['Sales','Category','Shipping_Cost']])
plt.title("boxplot of multiple columns")
plt.grid()
plt.show()
```



```
import numpy as np
outliers=np.where(Ecommerce['Sales']>10000)
print(outliers)
```

```
(array([], dtype=int64),)
```

```
outlier_index=Ecommerce[7395:7400]
print(outlier_index)
```

```
Empty DataFrame
Columns: [ROW_ID, Order_ID, Order_Date, Ship_Date, Ship_Mode, Customer_ID, Customer_Name, Segment, City, State, Country, Market, Postal_Code, Region, Product_ID, Category, Sub_Category, Product_Name, Sales, Quantity, Discount, Shipping_Cost, Profit, Order_Priority]
Index: []

[0 rows x 24 columns]
```

```
drop_outliers=Ecommerce.drop(outliers[0],inplace=True)
print(drop_outliers)
```

```
!pip install ydata.profiling
```

```
import pandas as pd
from ydata_profiling import ProfileReport
df=pd.read_csv('/content/electric_vehicles_spec_2025.csv')
profile= ProfileReport (df,title='electric_vehicles_spec_2025',explorative=True)
print(profile)
profile.to_file('data_profiling_report.html')
```

```
!pip install ydata.profiling
```

```
Requirement already satisfied: requests<3,>=2.32.0 in /usr/local/lib/python3.12/dist-packages (from ydata.profiling) (2.32.4)
Requirement already satisfied: tqdm<5,>=4.66.3 in /usr/local/lib/python3.12/dist-packages (from ydata.profiling) (4.67.1)
Requirement already satisfied: seaborn<0.14,>=0.10.1 in /usr/local/lib/python3.12/dist-packages (from ydata.profiling) (0.13.2)
Collecting multimethod<2,>=1.4 (from ydata.profiling)
  Downloading multimethod-1.12-py3-none-any.whl.metadata (9.6 kB)
Requirement already satisfied: statsmodels<1,>=0.13.2 in /usr/local/lib/python3.12/dist-packages (from ydata.profiling) (0.14.5)
Requirement already satisfied: typeguard<5,>=4 in /usr/local/lib/python3.12/dist-packages (from ydata.profiling) (4.4.4)
Collecting imagehash==4.3.2 (from ydata.profiling)
  Downloading ImageHash-4.3.2-py2.py3-none-any.whl.metadata (8.4 kB)
Requirement already satisfied: wordcloud>=1.9.4 in /usr/local/lib/python3.12/dist-packages (from ydata.profiling) (1.9.4)
Collecting dacite<2,>=1.9 (from ydata.profiling)
  Downloading dacite-1.9.2-py3-none-any.whl.metadata (17 kB)
Requirement already satisfied: numba<0.63,>=0.60 in /usr/local/lib/python3.12/dist-packages (from ydata.profiling) (0.60.0)
Requirement already satisfied: PyWavelets in /usr/local/lib/python3.12/dist-packages (from imagehash==4.3.2->ydata.profiling) (1.9.0)
Requirement already satisfied: pillow in /usr/local/lib/python3.12/dist-packages (from imagehash==4.3.2->ydata.profiling) (11.3.0)
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.12/dist-packages (from jinja2<3.2,>=3.1.6->ydata.profiling) (3.0.3)
Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.12/dist-packages (from matplotlib<=3.10,>=3.5->ydata.profiling) (1.3.3)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.12/dist-packages (from matplotlib<=3.10,>=3.5->ydata.profiling) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.12/dist-packages (from matplotlib<=3.10,>=3.5->ydata.profiling) (4.60.1)
Requirement already satisfied: kiwisolver>=1.3.1 in /usr/local/lib/python3.12/dist-packages (from matplotlib<=3.10,>=3.5->ydata.profiling) (1.4.9)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.12/dist-packages (from matplotlib<=3.10,>=3.5->ydata.profiling) (25.0)
Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.12/dist-packages (from matplotlib<=3.10,>=3.5->ydata.profiling) (3.2.5)
Requirement already satisfied: python-dateutil>=2.7 in /usr/local/lib/python3.12/dist-packages (from matplotlib<=3.10,>=3.5->ydata.profiling) (2.9.0.post0)
Requirement already satisfied: llvmlite<0.44,>=0.43.0dev0 in /usr/local/lib/python3.12/dist-packages (from numba<0.63,>=0.60->ydata.profiling) (0.43.0)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.12/dist-packages (from pandas!=1.4.0,<3.0,>1.5->ydata.profiling) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.12/dist-packages (from pandas!=1.4.0,<3.0,>1.5->ydata.profiling) (2025.2)
Requirement already satisfied: joblib>=0.14.1 in /usr/local/lib/python3.12/dist-packages (from phik<0.13,>=0.12.5->ydata.profiling) (1.5.2)
Requirement already satisfied: annotated-types>=0.6.0 in /usr/local/lib/python3.12/dist-packages (from pydantic<3,>=2->ydata.profiling) (0.7.0)
Requirement already satisfied: pydantic-core==2.41.4 in /usr/local/lib/python3.12/dist-packages (from pydantic<3,>=2->ydata.profiling) (2.41.4)
Requirement already satisfied: typing-extensions>=4.14.1 in /usr/local/lib/python3.12/dist-packages (from pydantic<3,>=2->ydata.profiling) (4.15.0)
Requirement already satisfied: typing-inspection>=0.4.2 in /usr/local/lib/python3.12/dist-packages (from pydantic<3,>=2->ydata.profiling) (0.4.2)
Requirement already satisfied: charset_normalizer<4,>=2 in /usr/local/lib/python3.12/dist-packages (from requests<3,>=2.32.0->ydata.profiling) (3.4.4)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.12/dist-packages (from requests<3,>=2.32.0->ydata.profiling) (3.11)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.12/dist-packages (from requests<3,>=2.32.0->ydata.profiling) (2.5.0)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.12/dist-packages (from requests<3,>=2.32.0->ydata.profiling) (2025.11.12)
Requirement already satisfied: patsy>=0.5.6 in /usr/local/lib/python3.12/dist-packages (from statsmodels<1,>=0.13.2->ydata.profiling) (1.0.2)
Requirement already satisfied: attrs>=19.3.0 in /usr/local/lib/python3.12/dist-packages (from visions<0.8.2,>=0.7.5->visions[type_image_path]<0.8.2,>=0.7.5->ydata.profiling) (25.4.0)
Requirement already satisfied: networkx>=2.4 in /usr/local/lib/python3.12/dist-packages (from visions<0.8.2,>=0.7.5->visions[type_image_path]<0.8.2,>=0.7.5->ydata.profiling) (3.6)
Collecting puremagic (from visions<0.8.2,>=0.7.5->visions[type_image_path]<0.8.2,>=0.7.5->ydata.profiling)
  Downloading puremagic-1.30-py3-none-any.whl.metadata (5.8 kB)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.12/dist-packages (from python-dateutil>=2.7->matplotlib<=3.10,>=3.5->ydata.profiling) (1.17.0)
Downloading ydata_profiling-4.18.0-py2.py3-none-any.whl (398 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 398.7/398.7 kB 7.3 MB/s eta 0:00:00
Downloading ImageHash-4.3.2-py2.py3-none-any.whl (296 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 296.7/296.7 kB 25.0 MB/s eta 0:00:00
```