```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

df=pd.read_csv(r"C:\Users\induk\Downloads\Global_Superstore.csv")

df.head()
```

```
   Row ID         Order ID  Order Date   Ship Date     Ship Mode
Customer ID  \
0   32298   CA-2012-124891   7/31/2012   7/31/2012      Same Day
RH-19495
1   26341    IN-2013-77878    2/5/2013    2/7/2013  Second Class
JR-16210
2   25330    IN-2013-71249  10/17/2013  10/18/2013   First Class
CR-12730
3   13524  ES-2013-1579342   1/28/2013   1/30/2013   First Class
KM-16375
4   47221     SG-2013-4320   11/5/2013   11/6/2013      Same Day
RH-9495

      Customer Name      Segment            City             State  ...
\
0       Rick Hansen     Consumer  New York City          New York  ...

1     Justin Ritter    Corporate     Wollongong  New South Wales  ...

2      Craig Reiter     Consumer       Brisbane       Queensland  ...

3  Katherine Murray  Home Office         Berlin           Berlin  ...

4       Rick Hansen     Consumer          Dakar            Dakar  ...


         Product ID    Category Sub-Category  \
0   TEC-AC-10003033  Technology  Accessories
1   FUR-CH-10003950   Furniture       Chairs
2   TEC-PH-10004664  Technology       Phones
3   TEC-PH-10004583  Technology       Phones
4  TEC-SHA-10000501  Technology      Copiers

                                        Product Name     Sales
Quantity  \
0  Plantronics CS510 - Over-the-Head monaural Wir...  2309.650
7
1          Novimex Executive Leather Armchair, Black  3709.395
9
2                  Nokia Smart Phone, with Caller ID  5175.171
9
3                     Motorola Smart Phone, Cordless  2892.510
```

```
5
4                    Sharp Wireless Fax, High-Speed  2832.960
8

  Discount    Profit  Shipping Cost  Order Priority
0      0.0  762.1845         933.57        Critical
1      0.1 -288.7650         923.63        Critical
2      0.1  919.9710         915.49          Medium
3      0.1  -96.5400         910.16          Medium
4      0.0  311.5200         903.04        Critical

[5 rows x 24 columns]
```

```
df.tail()
```

```
       Row ID        Order ID Order Date  Ship Date       Ship Mode  \
51285   29002   IN-2014-62366 2014-06-19  6/19/2014        Same Day
51286   35398  US-2014-102288 2014-06-20  6/24/2014  Standard Class
51287   40470  US-2013-155768 2013-12-02  12/2/2013        Same Day
51288    9596  MX-2012-140767 2012-02-18  2/22/2012  Standard Class
51289    6147  MX-2012-134460 2012-05-22  5/26/2012    Second Class

      Customer ID     Customer Name      Segment      City       State
...  \
51285    KE-16420   Katrina Edelman    Corporate      Kure   Hiroshima
...
51286    ZC-21910  Zuschuss Carroll     Consumer   Houston       Texas
...
51287    LB-16795    Laurel Beltran  Home Office    Oxnard  California
...
51288    RB-19795        Ross Baird  Home Office  Valinhos   S‹o Paulo
...
51289    MC-18100     Mick Crebagga     Consumer  Tipitapa     Managua
...

            Product ID         Category Sub-Category  \
51285  OFF-FA-10000746  Office Supplies    Fasteners
51286  OFF-AP-10002906  Office Supplies   Appliances
51287  OFF-EN-10001219  Office Supplies    Envelopes
51288  OFF-BI-10000806  Office Supplies      Binders
51289  OFF-PA-10004155  Office Supplies         Paper

                                             Product Name   Sales
Quantity  \
51285                       Advantus Thumb Tacks, 12 Pack  65.100
5
51286  Hoover Replacement Belt for Commercial Guardsm...   0.444
1
51287       #10- 4 1/8" x 9 1/2" Security-Tint Envelopes  22.920
3
```

```
51288                              Acco Index Tab, Economy  13.440
2
51289            Eaton Computer Printout Paper, 8.5 x 11  61.380
3

      Discount   Profit  Shipping Cost  Order Priority
51285      0.0   4.5000           0.01          Medium
51286      0.8  -1.1100           0.01          Medium
51287      0.0  11.2308           0.01            High
51288      0.0   2.4000           0.00          Medium
51289      0.0   1.8000           0.00            High

[5 rows x 24 columns]
```

```
df.size
```

```
1230960
```

```
df.shape
```

```
(51290, 24)
```

```
df.ndim
```

```
2
```

```
df.nunique()
```

```
Row ID           51290
Order ID         25035
Order Date        1430
Ship Date         1464
Ship Mode            4
Customer ID       1590
Customer Name      795
Segment              3
City              3636
State             1094
Country            147
Postal Code        631
Market               7
Region              13
Product ID        10292
Category             3
Sub-Category        17
Product Name      3788
Sales            22995
Quantity            14
Discount            27
Profit           24575
Shipping Cost    10037
```

```
Order Priority        4
dtype: int64
```

```
df.dtypes
```

```
Row ID              int64
Order ID           object
Order Date         object
Ship Date          object
Ship Mode          object
Customer ID        object
Customer Name      object
Segment            object
City               object
State              object
Country            object
Postal Code       float64
Market             object
Region             object
Product ID         object
Category           object
Sub-Category       object
Product Name       object
Sales             float64
Quantity            int64
Discount          float64
Profit            float64
Shipping Cost     float64
Order Priority     object
dtype: object
```

```
df.sample()
```

```
      Row ID       Order ID Order Date  Ship Date       Ship Mode
Customer ID  \
2574   22535  IN-2014-63206  6/20/2014  6/26/2014  Standard Class
DV-13465

         Customer Name   Segment       City      State  ...
Product ID  \
2574  Dianna Vittorini  Consumer  Hyderabad  Telangana  ...  FUR-BO-
10000210

       Category Sub-Category                   Product Name    Sales
Quantity  \
2574  Furniture    Bookcases  Ikea Library with Doors, Pine  1094.22
3

     Discount  Profit  Shipping Cost  Order Priority
2574      0.0  328.23         111.17             Low
```

```
[1 rows x 24 columns]
```

```
df.info
```

```
<bound method DataFrame.info of        Row ID         Order ID  Order
Date   Ship Date       Ship Mode  \
0       32298   CA-2012-124891   7/31/2012   7/31/2012        Same Day

1       26341    IN-2013-77878    2/5/2013    2/7/2013    Second Class

2       25330    IN-2013-71249  10/17/2013  10/18/2013     First Class

3       13524  ES-2013-1579342   1/28/2013   1/30/2013     First Class

4       47221     SG-2013-4320   11/5/2013   11/6/2013        Same Day

...       ...              ...         ...         ...             ...

51285   29002    IN-2014-62366   6/19/2014   6/19/2014        Same Day

51286   35398   US-2014-102288   6/20/2014   6/24/2014  Standard Class

51287   40470   US-2013-155768   12/2/2013   12/2/2013        Same Day

51288    9596   MX-2012-140767   2/18/2012   2/22/2012  Standard Class

51289    6147   MX-2012-134460   5/22/2012   5/26/2012    Second Class


      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
...           ...               ...          ...            ...
51285    KE-16420   Katrina Edelman    Corporate           Kure
51286    ZC-21910  Zuschuss Carroll     Consumer        Houston
51287    LB-16795    Laurel Beltran  Home Office         Oxnard
51288    RB-19795        Ross Baird  Home Office       Valinhos
51289    MC-18100     Mick Crebagga     Consumer       Tipitapa

                  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
...                ...  ...               ...              ...
...
51285        Hiroshima  ...   OFF-FA-10000746  Office Supplies
Fasteners
51286            Texas  ...   OFF-AP-10002906  Office Supplies
Appliances
51287       California  ...   OFF-EN-10001219  Office Supplies
Envelopes
51288        S‹o Paulo  ...   OFF-BI-10000806  Office Supplies
Binders
51289          Managua  ...   OFF-PA-10004155  Office Supplies
Paper

                                              Product Name     Sales
Quantity  \
0      Plantronics CS510 - Over-the-Head monaural Wir...  2309.650
7
1               Novimex Executive Leather Armchair, Black  3709.395
9
2                      Nokia Smart Phone, with Caller ID  5175.171
9
3                         Motorola Smart Phone, Cordless  2892.510
5
4                         Sharp Wireless Fax, High-Speed  2832.960
8
...                                                  ...       ...
...
51285                      Advantus Thumb Tacks, 12 Pack    65.100
5
51286  Hoover Replacement Belt for Commercial Guardsm...     0.444
1
51287       #10- 4 1/8" x 9 1/2" Security-Tint Envelopes    22.920
3
51288                             Acco Index Tab, Economy    13.440
2
51289            Eaton Computer Printout Paper, 8.5 x 11    61.380
3

      Discount    Profit  Shipping Cost  Order Priority
0          0.0  762.1845         933.57        Critical
1          0.1 -288.7650         923.63        Critical
2          0.1  919.9710         915.49          Medium
3          0.1  -96.5400         910.16          Medium
4          0.0  311.5200         903.04        Critical
...        ...       ...            ...             ...
```

```
51285      0.0    4.5000           0.01          Medium
51286      0.8   -1.1100           0.01          Medium
51287      0.0   11.2308           0.01            High
51288      0.0    2.4000           0.00          Medium
51289      0.0    1.8000           0.00            High

[51290 rows x 24 columns]>
```

```
df.describe()
```

```
            Row ID   Postal Code         Sales      Quantity
Discount  \
count  51290.00000   9994.000000  51290.000000  51290.000000
51290.000000
mean   25645.50000  55190.379428    246.490581      3.476545
0.142908
std    14806.29199  32063.693350    487.565361      2.278766
0.212280
min        1.00000   1040.000000      0.444000      1.000000
0.000000
25%    12823.25000  23223.000000     30.758625      2.000000
0.000000
50%    25645.50000  56430.500000     85.053000      3.000000
0.000000
75%    38467.75000  90008.000000    251.053200      5.000000
0.200000
max    51290.00000  99301.000000  22638.480000     14.000000
0.850000

             Profit  Shipping Cost
count  51290.000000   51290.000000
mean      28.610982      26.375915
std      174.340972      57.296804
min    -6599.978000       0.000000
25%        0.000000       2.610000
50%        9.240000       7.790000
75%       36.810000      24.450000
max     8399.976000     933.570000
```

```
df.describe(include='object')
```

```
               Order ID Order Date   Ship Date       Ship Mode Customer
ID  \
count             51290      51290       51290           51290
51290
unique            25035       1430        1464               4
1590
top     CA-2014-100111  6/18/2014  11/22/2014  Standard Class    PO-
18850
freq                 14        135         130           30775
```

```
97

          Customer Name   Segment            City       State
Country  \
count              51290     51290           51290       51290
51290
unique               795         3            3636        1094
147
top     Muhammed Yedwab  Consumer  New York City  California  United
States
freq                 108     26518             915        2001
9994

       Market   Region       Product ID         Category Sub-Category
\
count   51290    51290            51290            51290        51290

unique      7       13            10292                3           17

top      APAC  Central  OFF-AR-10003651  Office Supplies      Binders

freq    11002    11117               35            31273         6152


       Product Name Order Priority
count         51290          51290
unique         3788              4
top          Staples         Medium
freq            227          29433

categorical_data = df.select_dtypes(exclude=[np.number])
categorical_data

              Order ID  Order Date   Ship Date       Ship Mode
Customer ID  \
0       CA-2012-124891   7/31/2012   7/31/2012        Same Day    RH-
19495
1        IN-2013-77878    2/5/2013    2/7/2013    Second Class    JR-
16210
2        IN-2013-71249  10/17/2013  10/18/2013     First Class    CR-
12730
3      ES-2013-1579342   1/28/2013   1/30/2013     First Class    KM-
16375
4         SG-2013-4320   11/5/2013   11/6/2013        Same Day     RH-
9495
...                ...         ...         ...             ...
...
51285    IN-2014-62366   6/19/2014   6/19/2014        Same Day    KE-
16420
51286   US-2014-102288   6/20/2014   6/24/2014  Standard Class    ZC-
```

```
21910
51287   US-2013-155768   12/2/2013   12/2/2013        Same Day    LB-
16795
51288   MX-2012-140767   2/18/2012   2/22/2012  Standard Class    RB-
19795
51289   MX-2012-134460   5/22/2012   5/26/2012    Second Class    MC-
18100

          Customer Name      Segment           City
State  \
0           Rick Hansen     Consumer  New York City         New York

1         Justin Ritter    Corporate     Wollongong  New South Wales

2          Craig Reiter     Consumer       Brisbane       Queensland

3      Katherine Murray  Home Office         Berlin           Berlin

4           Rick Hansen     Consumer          Dakar            Dakar

...                 ...          ...            ...              ...

51285   Katrina Edelman    Corporate           Kure        Hiroshima

51286  Zuschuss Carroll     Consumer        Houston            Texas

51287    Laurel Beltran  Home Office         Oxnard       California

51288        Ross Baird  Home Office       Valinhos        S‹o Paulo

51289     Mick Crebagga     Consumer       Tipitapa          Managua


             Country  Market      Region        Product ID
Category  \
0      United States      US        East   TEC-AC-10003033
Technology
1          Australia    APAC     Oceania   FUR-CH-10003950
Furniture
2          Australia    APAC     Oceania   TEC-PH-10004664
Technology
3            Germany      EU     Central   TEC-PH-10004583
Technology
4            Senegal  Africa      Africa  TEC-SHA-10000501
Technology
...              ...     ...         ...               ...
...
51285          Japan    APAC  North Asia   OFF-FA-10000746  Office
Supplies
51286  United States      US     Central   OFF-AP-10002906  Office
```

```
Supplies
51287  United States      US        West   OFF-EN-10001219  Office
Supplies
51288         Brazil   LATAM       South   OFF-BI-10000806  Office
Supplies
51289      Nicaragua   LATAM     Central   OFF-PA-10004155  Office
Supplies

      Sub-Category                                        Product Name
\
0      Accessories  Plantronics CS510 - Over-the-Head monaural Wir...

1           Chairs          Novimex Executive Leather Armchair, Black

2           Phones                   Nokia Smart Phone, with Caller ID

3           Phones                      Motorola Smart Phone, Cordless

4          Copiers                      Sharp Wireless Fax, High-Speed

...            ...                                                ...

51285    Fasteners                       Advantus Thumb Tacks, 12 Pack

51286   Appliances  Hoover Replacement Belt for Commercial Guardsm...

51287    Envelopes       #10- 4 1/8" x 9 1/2" Security-Tint Envelopes

51288      Binders                              Acco Index Tab, Economy

51289        Paper             Eaton Computer Printout Paper, 8.5 x 11


      Order Priority
0           Critical
1           Critical
2             Medium
3             Medium
4           Critical
...              ...
51285         Medium
51286         Medium
51287           High
51288         Medium
51289           High

[51290 rows x 17 columns]

categorical_data.shape

(51290, 17)
```

```
numeric_data = df.select_dtypes(include=[np.number])
numeric_data
```

```
       Row ID  Postal Code     Sales  Quantity  Discount    Profit  \
0       32298      10024.0  2309.650         7       0.0  762.1845
1       26341          NaN  3709.395         9       0.1 -288.7650
2       25330          NaN  5175.171         9       0.1  919.9710
3       13524          NaN  2892.510         5       0.1  -96.5400
4       47221          NaN  2832.960         8       0.0  311.5200
...       ...          ...       ...       ...       ...       ...
51285   29002          NaN    65.100         5       0.0    4.5000
51286   35398      77095.0     0.444         1       0.8   -1.1100
51287   40470      93030.0    22.920         3       0.0   11.2308
51288    9596          NaN    13.440         2       0.0    2.4000
51289    6147          NaN    61.380         3       0.0    1.8000

       Shipping Cost
0             933.57
1             923.63
2             915.49
3             910.16
4             903.04
...              ...
51285          0.01
51286          0.01
51287          0.01
51288          0.00
51289          0.00

[51290 rows x 7 columns]
```

```
numeric_data.shape
```

```
(51290, 7)
```

```
df.isnull().head()
```

```
   Row ID  Order ID  Order Date  Ship Date  Ship Mode  Customer ID  \
0   False     False       False      False      False        False
1   False     False       False      False      False        False
2   False     False       False      False      False        False
3   False     False       False      False      False        False
4   False     False       False      False      False        False

   Customer Name  Segment   City  State  ...  Product ID  Category  \
0          False    False  False  False  ...       False     False
1          False    False  False  False  ...       False     False
2          False    False  False  False  ...       False     False
3          False    False  False  False  ...       False     False
4          False    False  False  False  ...       False     False
```

```
   Sub-Category  Product Name  Sales  Quantity  Discount  Profit  \
0         False         False  False     False     False   False
1         False         False  False     False     False   False
2         False         False  False     False     False   False
3         False         False  False     False     False   False
4         False         False  False     False     False   False

   Shipping Cost  Order Priority
0          False           False
1          False           False
2          False           False
3          False           False
4          False           False

[5 rows x 24 columns]
```

```
df.isnull().sum()
```

```
Row ID                0
Order ID              0
Order Date            0
Ship Date             0
Ship Mode             0
Customer ID           0
Customer Name         0
Segment               0
City                  0
State                 0
Country               0
Postal Code       41296
Market                0
Region                0
Product ID            0
Category              0
Sub-Category          0
Product Name          0
Sales                 0
Quantity              0
Discount              0
Profit                0
Shipping Cost         0
Order Priority        0
dtype: int64
```

```
df[df['Postal Code'].isnull()]
```

```
       Row ID         Order ID  Order Date   Ship Date       Ship Mode
\
1       26341    IN-2013-77878    2/5/2013    2/7/2013    Second Class
```

```
2       25330    IN-2013-71249  10/17/2013  10/18/2013     First Class
3       13524  ES-2013-1579342   1/28/2013   1/30/2013     First Class
4       47221     SG-2013-4320   11/5/2013   11/6/2013        Same Day
5       22732    IN-2013-42360   6/28/2013    7/1/2013    Second Class
...       ...              ...         ...         ...             ...
51283   24105    IN-2014-72327   5/30/2014   5/30/2014        Same Day
51284   24175    IN-2014-57662    8/5/2014   8/10/2014  Standard Class
51285   29002    IN-2014-62366   6/19/2014   6/19/2014        Same Day
51288    9596   MX-2012-140767   2/18/2012   2/22/2012  Standard Class
51289    6147   MX-2012-134460   5/22/2012   5/26/2012    Second Class


      Customer ID      Customer Name      Segment        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
5        JM-15655        Jim Mitchum    Corporate      Sydney
...           ...                ...         ...         ...
51283    KH-16330    Katharine Harms    Corporate     Lucknow
51284    DB-13270  Deborah Brumfield  Home Office  Townsville
51285    KE-16420    Katrina Edelman    Corporate        Kure
51288    RB-19795         Ross Baird  Home Office    Valinhos
51289    MC-18100      Mick Crebagga     Consumer    Tipitapa

                  State  ...        Product ID         Category Sub-
Category  \
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
5      New South Wales  ...   TEC-PH-10000030       Technology
Phones
...                ...  ...               ...              ...
...
51283    Uttar Pradesh  ...   OFF-PA-10000215  Office Supplies
Paper
```

```
51284       Queensland  ...   OFF-BI-10002424  Office Supplies
Binders
51285        Hiroshima  ...   OFF-FA-10000746  Office Supplies
Fasteners
51288        S‹o Paulo  ...   OFF-BI-10000806  Office Supplies
Binders
51289          Managua  ...   OFF-PA-10004155  Office Supplies
Paper

                                      Product Name     Sales Quantity
Discount  \
1      Novimex Executive Leather Armchair, Black  3709.395        9
0.1
2              Nokia Smart Phone, with Caller ID  5175.171        9
0.1
3                 Motorola Smart Phone, Cordless  2892.510        5
0.1
4                 Sharp Wireless Fax, High-Speed  2832.960        8
0.0
5              Samsung Smart Phone, with Caller ID  2862.675        5
0.1
...                                          ...       ...      ...
...
51283             Eaton Parchment Paper, Premium    26.940        2
0.0
51284                       Avery Binder, Economy    58.050        5
0.1
51285               Advantus Thumb Tacks, 12 Pack    65.100        5
0.0
51288                     Acco Index Tab, Economy    13.440        2
0.0
51289    Eaton Computer Printout Paper, 8.5 x 11    61.380        3
0.0

        Profit  Shipping Cost  Order Priority
1     -288.765         923.63        Critical
2      919.971         915.49          Medium
3      -96.540         910.16          Medium
4      311.520         903.04        Critical
5      763.275         897.35        Critical
...        ...            ...             ...
51283    1.860           0.01            High
51284   19.950           0.01          Medium
51285    4.500           0.01          Medium
51288    2.400           0.00          Medium
51289    1.800           0.00            High

[41296 rows x 24 columns]
```

```
df.dropna(how = 'any').shape
```

```
(9994, 24)
```

```
df.dropna(how = 'all').shape
```

```
(51290, 24)
```

```
df.columns
```

```
Index(['Row ID', 'Order ID', 'Order Date', 'Ship Date', 'Ship Mode',
       'Customer ID', 'Customer Name', 'Segment', 'City', 'State',
'Country',
       'Postal Code', 'Market', 'Region', 'Product ID', 'Category',
       'Sub-Category', 'Product Name', 'Sales', 'Quantity',
'Discount',
       'Profit', 'Shipping Cost', 'Order Priority'],
      dtype='object')
```

```
df['Order Date'] = pd.to_datetime(df['Order Date'])
```

```
category_sales = df.groupby('Category')
['Sales'].sum().sort_values(ascending=False)
category_sales
```

```
Category
Technology         4.744557e+06
Furniture          4.110874e+06
Office Supplies    3.787070e+06
Name: Sales, dtype: float64
```

```
category_sales = df.groupby('Category')['Sales'].sum()
category_sales
```

```
Category
Furniture          4.110874e+06
Office Supplies    3.787070e+06
Technology         4.744557e+06
Name: Sales, dtype: float64
```

```
profit = df.groupby('Category')
['Profit'].sum().sort_values(ascending=False)
profit
```

```
Category
Technology         663778.73318
Office Supplies    518473.83430
Furniture          285204.72380
Name: Profit, dtype: float64
```

```
profit = df.groupby('Category')['Profit'].sum()
profit
```

```
Category
Furniture          285204.72380
Office Supplies    518473.83430
Technology         663778.73318
Name: Profit, dtype: float64

loss=df[df['Profit'] < 0]
loss

       Row ID         Order ID Order Date   Ship Date       Ship Mode
\
1       26341    IN-2013-77878 2013-02-05    2/7/2013    Second Class

3       13524  ES-2013-1579342 2013-01-28   1/30/2013     First Class

9       40936   CA-2012-116638 2012-01-28   1/31/2012    Second Class

30        220   US-2011-128776 2011-12-28  12/30/2011    Second Class

32      32735   CA-2012-139731 2012-10-15  10/15/2012        Same Day

...       ...              ...        ...         ...             ...

51276   31558   US-2014-155299 2014-06-09   6/13/2014  Standard Class

51278   43818     NI-2011-4700 2011-09-08   9/10/2011    Second Class

51279   46231     ZI-2011-4350 2011-03-21   3/26/2011  Standard Class

51280   46582     TU-2014-6730 2014-11-29  11/30/2014     First Class

51286   35398   US-2014-102288 2014-06-20   6/24/2014  Standard Class


      Customer ID      Customer Name      Segment           City  \
1        JR-16210      Justin Ritter    Corporate     Wollongong
3        KM-16375   Katherine Murray  Home Office         Berlin
9        JH-15985        Joseph Holt     Consumer        Concord
30       RR-19525          Rick Reed    Corporate  Santo Domingo
32       JE-15745         Joel Eaton     Consumer       Amarillo
...           ...                ...          ...            ...
51276    Dl-13600       Dorris liebe    Corporate       Pasadena
51278     PO-8865  Patrick O'Donnell     Consumer          Lagos
51279      AS-285   Alejandro Savely    Corporate         Harare
51280     KF-6285     Karen Ferguson  Home Office         Midyat
51286    ZC-21910   Zuschuss Carroll     Consumer        Houston

                 State  ...        Product ID         Category Sub-
Category  \
1      New South Wales  ...   FUR-CH-10003950        Furniture
Chairs
```

```
3               Berlin  ...   TEC-PH-10004583       Technology
Phones
9       North Carolina  ...   FUR-TA-10000198        Furniture
Tables
30       Santo Domingo  ...   TEC-PH-10002815       Technology
Phones
32              Texas  ...   FUR-CH-10002024        Furniture
Chairs
...               ...  ...               ...              ...
...
51276           Texas  ...   OFF-AP-10002203  Office Supplies
Appliances
51278           Lagos  ...  OFF-STA-10001791  Office Supplies
Art
51279          Harare  ...  OFF-SAN-10001862  Office Supplies
Art
51280          Mardin  ...  OFF-BOS-10000350  Office Supplies
Art
51286           Texas  ...   OFF-AP-10002906  Office Supplies
Appliances

                                            Product Name     Sales
Quantity  \
1              Novimex Executive Leather Armchair, Black  3709.395
9
3                          Motorola Smart Phone, Cordless  2892.510
5
9      Chromcraft Bull-Nose Wood Oval Conference Tabl...  4297.644
13
30                            Samsung Smart Phone, VoIP  1696.640
5
32          HON 5400 Series Task Chairs for Big and Tall  2453.430
5
...                                                  ...       ...
...
51276  Eureka Disposable Bags for Sanitaire Vibra Gro...     1.624
2
51278                  Stanley Highlighters, Water Color     5.364
1
51279                   Sanford Highlighters, Easy-Erase     9.612
2
51280                                   Boston Pens, Blue    34.128
6
51286  Hoover Replacement Belt for Commercial Guardsm...     0.444
1

      Discount     Profit  Shipping Cost  Order Priority
1          0.1  -288.7650         923.63        Critical
3          0.1   -96.5400         910.16          Medium
```

```
9          0.4 -1862.3124         865.74        Critical
30         0.2  -148.4600         704.06        Critical
32         0.3  -350.4900         690.42            High
...        ...        ...            ...             ...
51276      0.8    -4.4660           0.02          Medium
51278      0.7    -6.4560           0.02            High
51279      0.7   -21.1680           0.02          Medium
51280      0.6   -49.5720           0.02          Medium
51286      0.8    -1.1100           0.01          Medium

[12544 rows x 24 columns]
```

```
loss.shape
```

```
(12544, 24)
```

```
loss.describe()
```

```
             Row ID   Postal Code         Sales      Quantity
Discount  \
count  12544.000000   1871.000000  12544.000000  12544.000000
12544.000000
mean   25638.463249  55991.122929    198.502681      3.427455
0.450812
std    14857.870020  26041.501999    423.779161      2.278014
0.187446
min       17.000000   1841.000000      0.444000      1.000000
0.002000
25%    12513.750000  33024.000000     21.420000      2.000000
0.400000
50%    25487.500000  60623.000000     64.543500      3.000000
0.470000
75%    38891.750000  77095.000000    206.685000      5.000000
0.600000
max    51280.000000  98198.000000  22638.480000     14.000000
0.850000

             Profit  Shipping Cost
count  12544.000000   12544.000000
mean     -73.393348      21.234884
std      186.222175      47.381359
min    -6599.978000       0.010000
25%      -65.340000       1.850000
50%      -21.206250       5.920000
75%       -6.899200      19.922500
max       -0.000900     923.630000
```

```
Total_loss=np.negative(loss['Profit'].sum())
print("Total loss = %.2d" %Total_loss)
```

```
Total loss = 920646
```

```python
loss.groupby(by='Segment').sum()
```

```
C:\Users\induk\AppData\Local\Temp\ipykernel_11236\1546001968.py:1:
FutureWarning: The default value of numeric_only in
DataFrameGroupBy.sum is deprecated. In a future version, numeric_only
will default to False. Either specify numeric_only or select only
columns which should be valid for the function.
  loss.groupby(by='Segment').sum()
```

```
                Row ID  Postal Code         Sales  Quantity  Discount
\
Segment

Consumer     167220474   57202260.0  1.290768e+06     22429  2932.554

Corporate     96783261   30034273.0  7.472977e+05     12769  1702.848

Home Office   57605148   17522858.0  4.519519e+05      7796  1019.582


                   Profit  Shipping Cost
Segment
Consumer    -477357.40794      139420.96
Corporate   -275089.09164       80217.70
Home Office -168199.65614       46731.72
```

```python
loss.groupby(by='Sub-Category').sum()
```

```
C:\Users\induk\AppData\Local\Temp\ipykernel_11236\2221067784.py:1:
FutureWarning: The default value of numeric_only in
DataFrameGroupBy.sum is deprecated. In a future version, numeric_only
will default to False. Either specify numeric_only or select only
columns which should be valid for the function.
  loss.groupby(by='Sub-Category').sum()
```

```
                Row ID  Postal Code         Sales  Quantity  Discount
\
Sub-Category

Accessories   16744866    5286382.0   95024.40830      2278   284.180

Appliances    10062035    4825871.0  157824.82480      1272   184.790

Art           26899971          0.0   42568.13230      2760   459.450

Binders       47120841   32609300.0   64478.33810      5317   913.690

Bookcases     16565636    6423506.0  297654.56950      2452   285.120

Chairs        25361375   15008025.0  370504.75040      3988   385.980
```

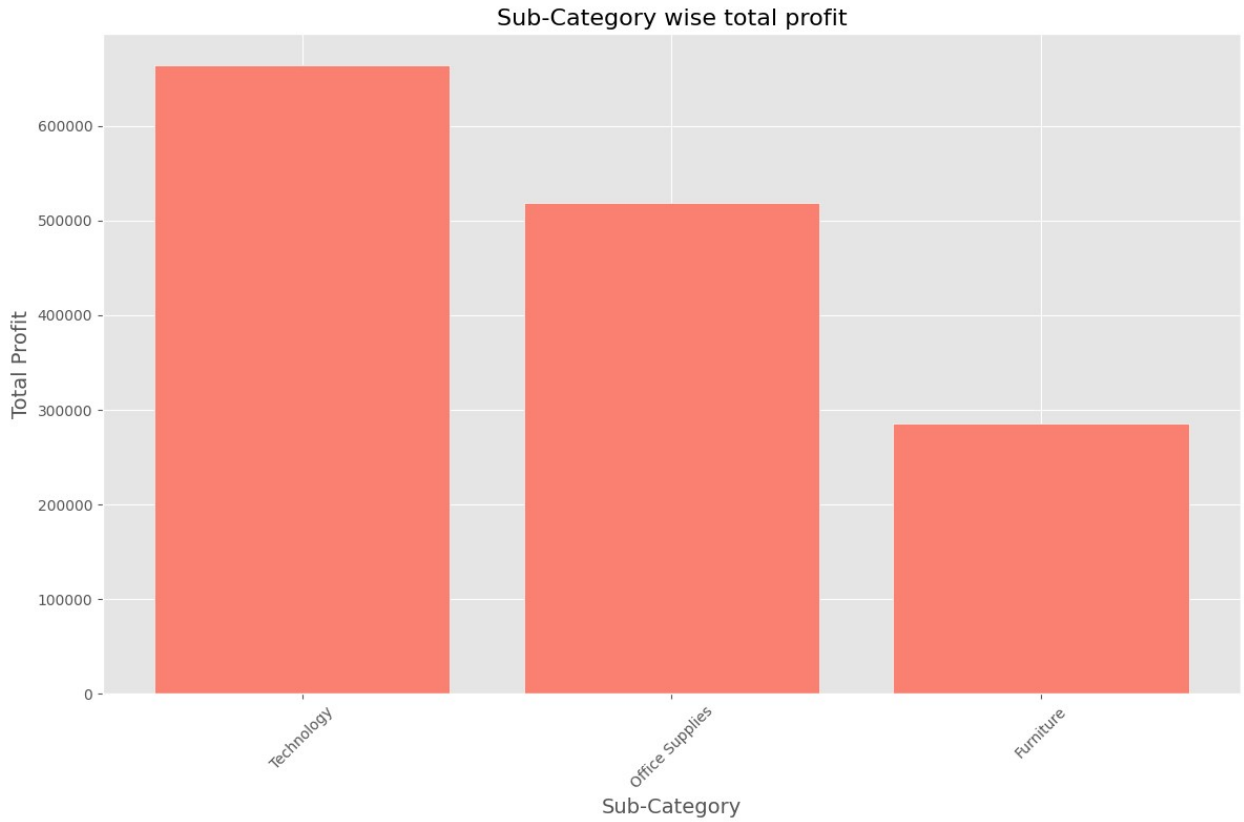
```
Copiers        11172787           0.0  233307.31018       1702   199.704

Envelopes      13140057           0.0   25435.83450       1951   260.800

Fasteners      14884085      701930.0   14035.13530       2051   282.860

Furnishings    18282079    10970913.0   69833.24050       2928   362.160

Labels         12690157           0.0    8952.49990       1828   243.360

Machines       12404081     2236261.0  195009.45090       1471   207.380

Paper          12773900           0.0   22855.24630       1750   244.890

Phones         20762031     6105294.0  261349.10340       2684   328.000

Storage        35566023     8606475.0  240588.90780       4614   533.040

Supplies       14573818     1761430.0   46921.61690       2171   265.980

Tables         12605141    10224004.0  343674.25570       1777   213.600
```

Sub-Category	Profit	Shipping Cost
Accessories	-39857.49820	10645.11
Appliances	-63991.69040	16528.92
Art	-21660.62770	4897.23
Binders	-52884.06130	7275.25
Bookcases	-101446.29730	30938.35
Chairs	-96084.89690	41210.07
Copiers	-71547.49982	23797.50
Envelopes	-10863.56550	2925.86
Fasteners	-6273.29990	1515.93
Furnishings	-28514.49690	7512.67
Labels	-3760.77010	1020.00
Machines	-78672.74030	18653.09
Paper	-10299.49370	2412.81
Phones	-96417.66010	30129.14
Storage	-76063.97800	27538.13
Supplies	-18184.43100	4491.99
Tables	-144123.14860	34878.33

Data Visualisaton

```
profit = df.groupby('Category')
['Profit'].sum().sort_values(ascending=False)
```

```
plt.figure(figsize=(12, 8))
plt.bar(profit.index, profit.values, color='salmon',
edgecolor='white')
plt.xlabel('Sub-Category', fontsize=14)
plt.ylabel('Total Profit', fontsize=14)
plt.title('Sub-Category wise total profit', fontsize=16)
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```




```
Total_loss = np.negative(loss['Profit'].sum())
print(f"Total loss = {Total_loss:.2f}")
```

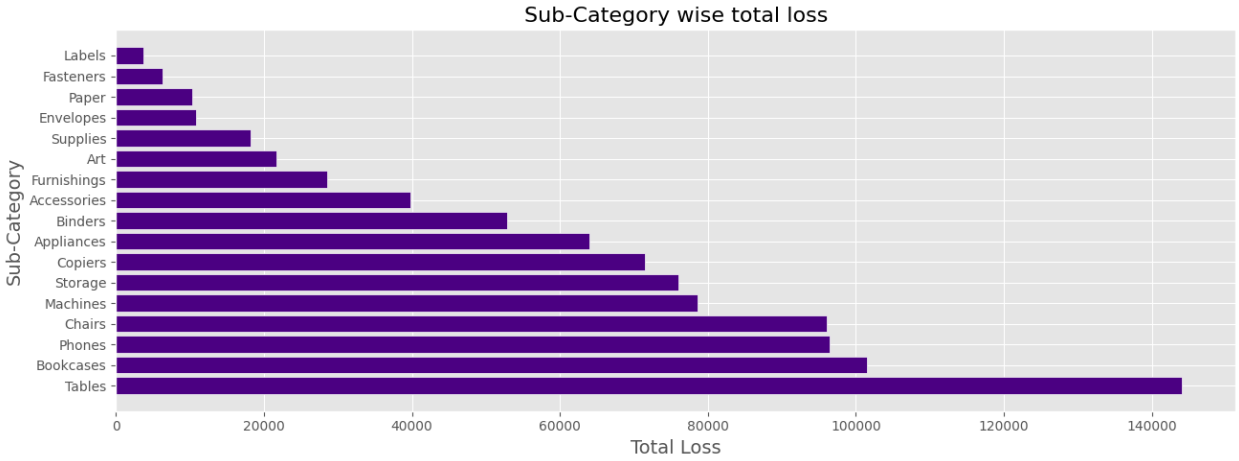
```
Total loss = 920646.16
```

```
loss_by_segment = loss.groupby('Segment')
['Profit'].sum().sort_values(ascending=True)

loss_by_subcategory = loss.groupby('Sub-Category')
['Profit'].sum().sort_values(ascending=True)

plt.figure(figsize=(14, 6))
```

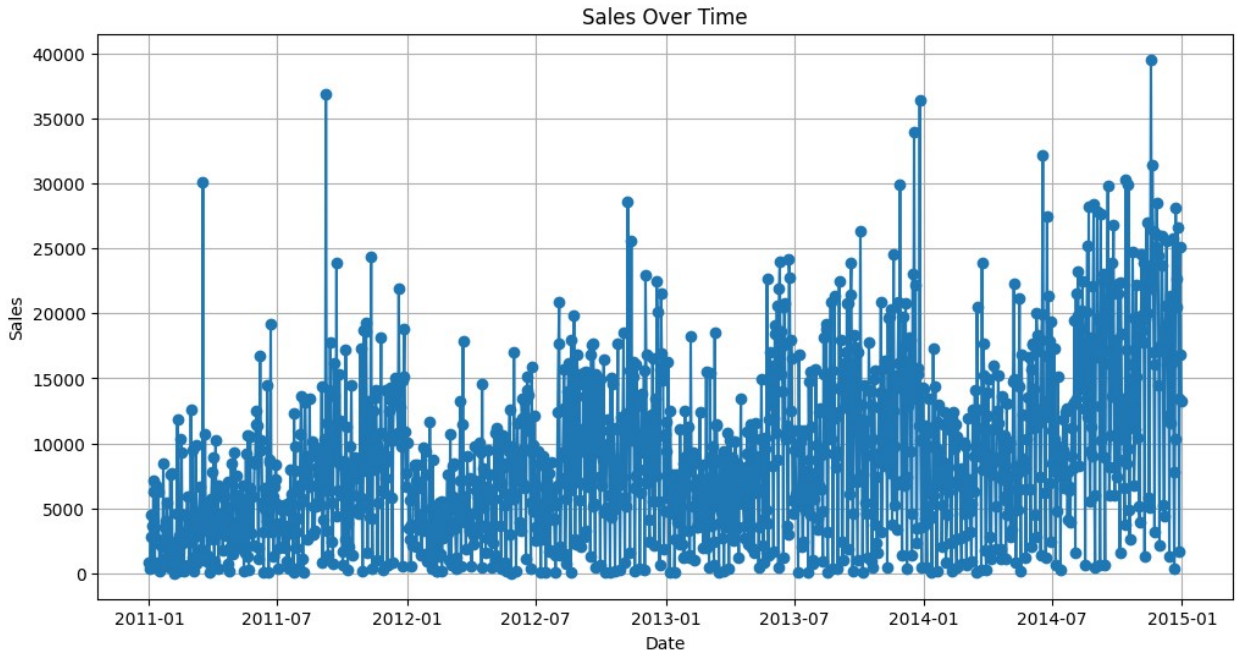
```
<Figure size 1400x600 with 0 Axes>
```

```
<Figure size 1400x600 with 0 Axes>
```

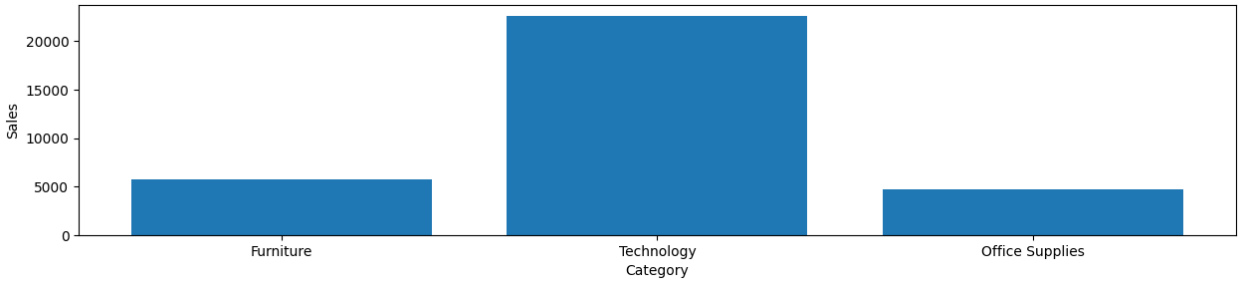
```
plt.subplot(1, 2, 2)
plt.barh(loss_by_subcategory.index,
np.negative(loss_by_subcategory.values), color='indigo',
edgecolor='white')
plt.xlabel('Total Loss', fontsize=14)
plt.ylabel('Sub-Category', fontsize=14)
plt.title('Sub-Category wise total loss', fontsize=16)

plt.tight_layout()
plt.show()
```



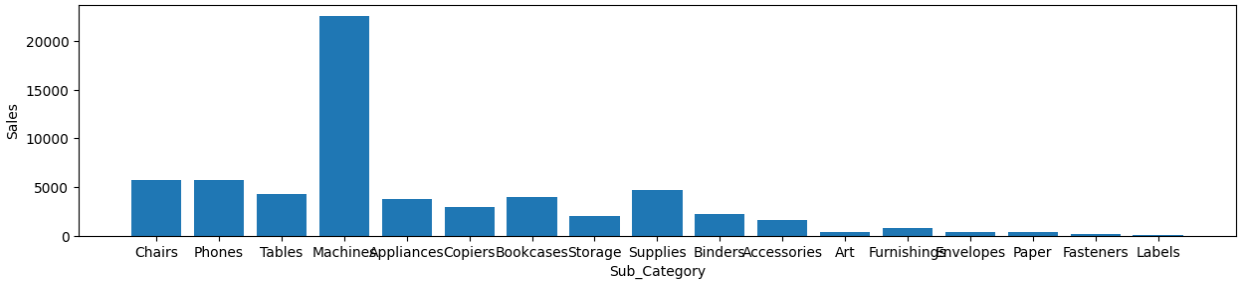

```
sales_over_time = df.groupby('Order Date')
['Sales'].sum().reset_index()

plt.figure(figsize=(12, 6))
plt.plot(sales_over_time['Order Date'], sales_over_time['Sales'],
marker='o', linestyle='-')
plt.title('Sales Over Time')
plt.xlabel('Date')
plt.ylabel('Sales')
plt.grid(True)
plt.show()
```



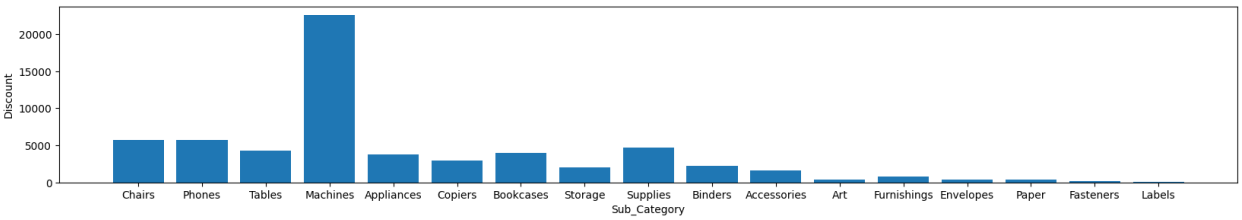

```
plt.rcParams['figure.figsize']=(15,3)
plt.bar(loss['Category'],loss['Sales']);
plt.rcParams.update({'font.size':10});
plt.xlabel('Category');
plt.ylabel('Sales');
```



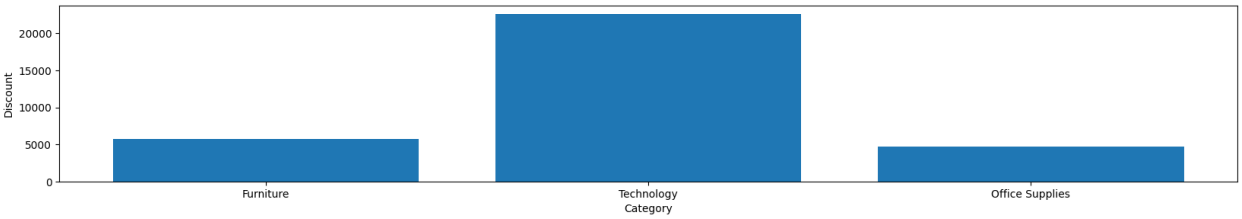

```
plt.rcParams['figure.figsize']=(15,3)
plt.bar(loss['Sub-Category'],loss['Sales']);
plt.rcParams.update({'font.size':10});
plt.xlabel('Sub_Category');
plt.ylabel('Sales');
```




```
plt.rcParams['figure.figsize']=(20,3)
plt.bar(loss['Sub-Category'],loss['Sales']);
plt.rcParams.update({'font.size':10});
plt.xlabel('Sub_Category');
plt.ylabel('Discount');
```




```
plt.rcParams['figure.figsize']=(20,3)
plt.bar(loss['Category'],loss['Sales']);
plt.rcParams.update({'font.size':10});
plt.xlabel('Category');
plt.ylabel('Discount');
```

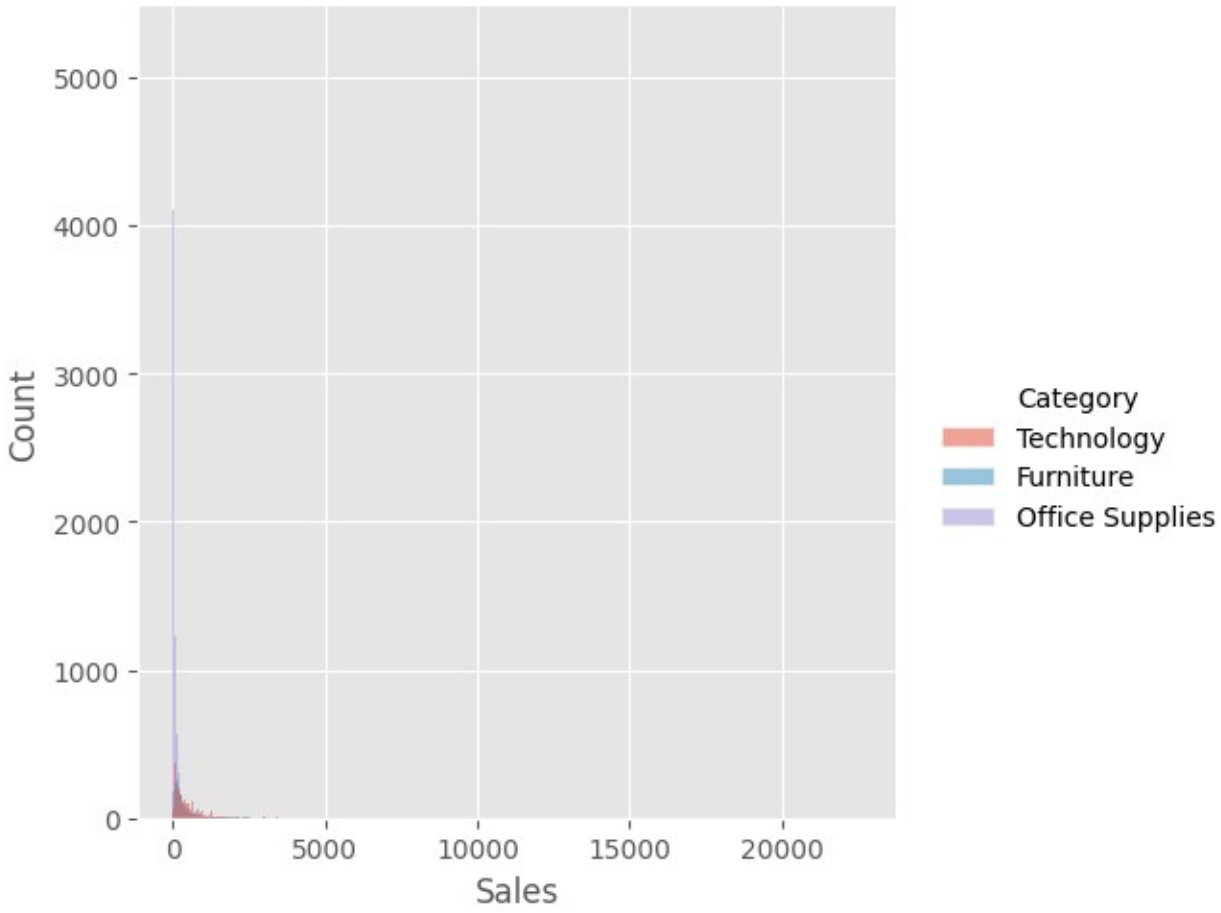



```
df.corr()
sns.heatmap(df.corr(),cmap='Greens',annot=True);
plt.rcParams['figure.figsize']=(25,5)
```

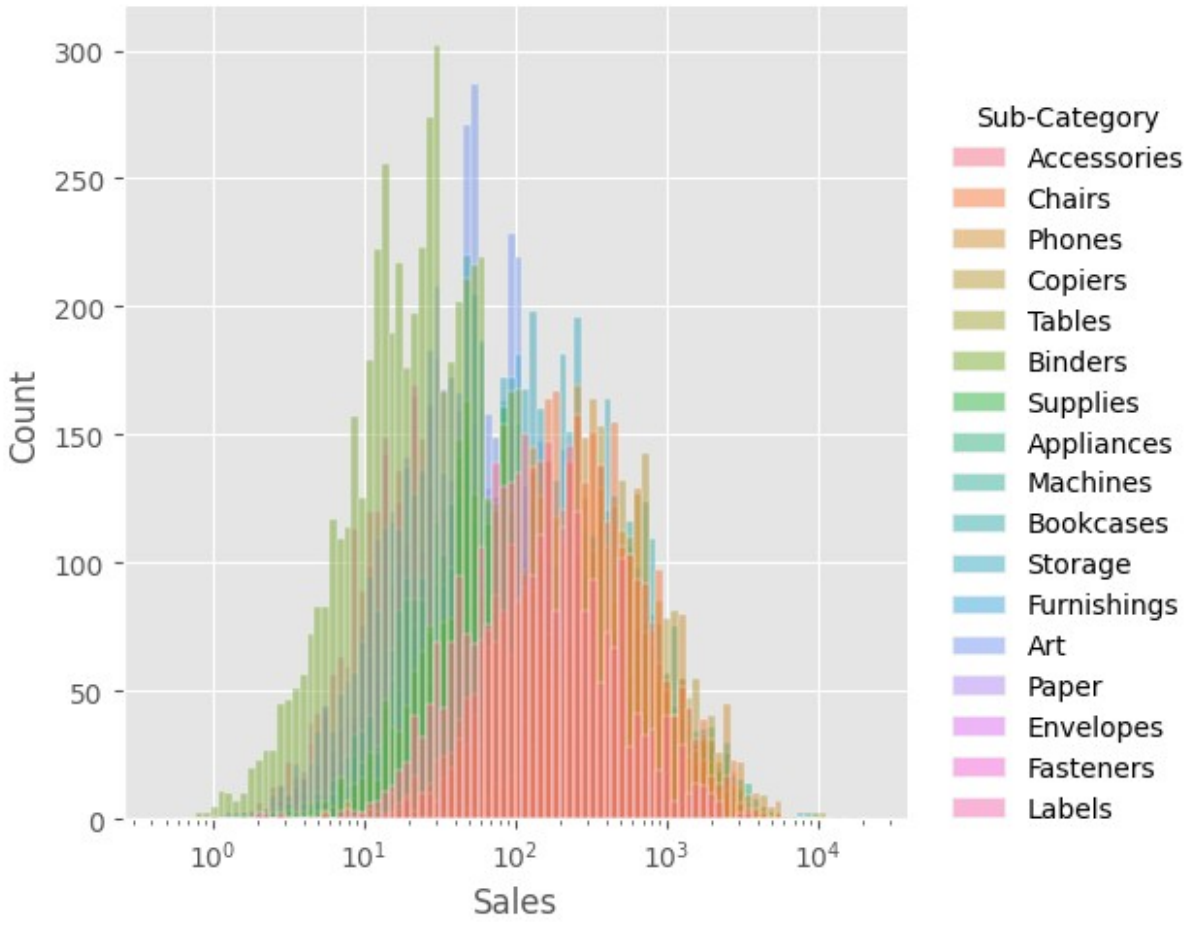
```
C:\Users\induk\AppData\Local\Temp\ipykernel_11236\57406746.py:1:
FutureWarning: The default value of numeric_only in DataFrame.corr is
deprecated. In a future version, it will default to False. Select only
valid columns or specify the value of numeric_only to silence this
warning.
  df.corr()
C:\Users\induk\AppData\Local\Temp\ipykernel_11236\57406746.py:2:
FutureWarning: The default value of numeric_only in DataFrame.corr is
deprecated. In a future version, it will default to False. Select only
```

```
valid columns or specify the value of numeric_only to silence this
warning.
  sns.heatmap(df.corr(),cmap='Greens',annot=True);
```

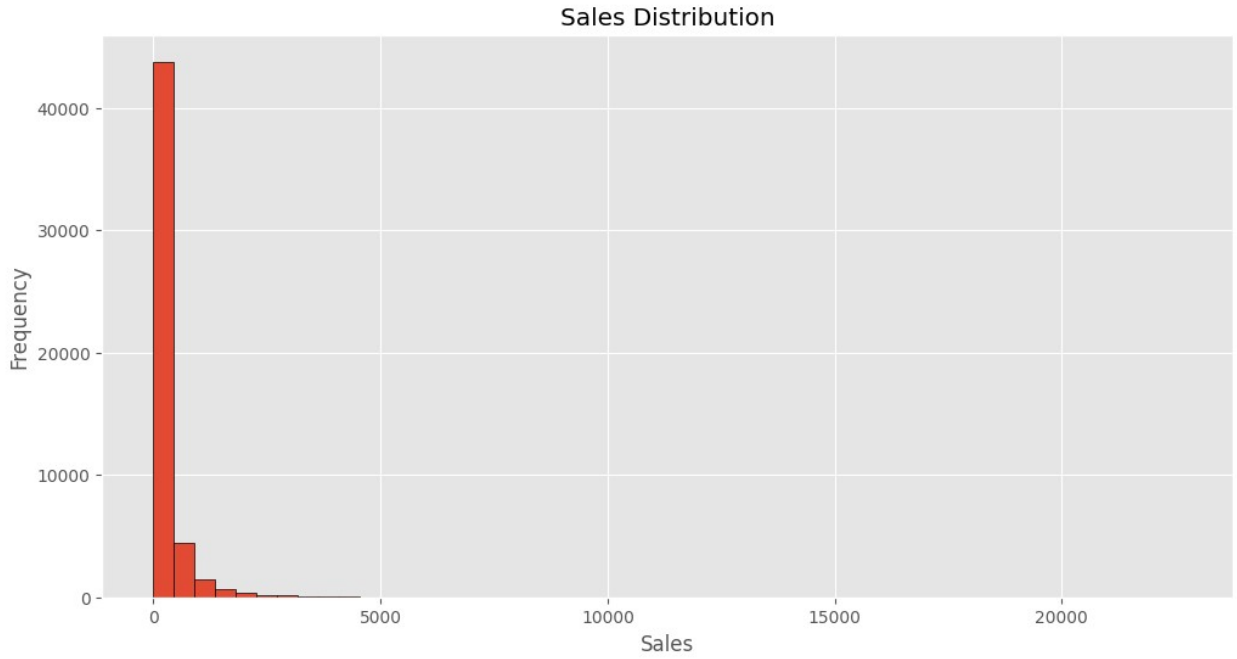

```
plt.style.use('ggplot')

sns.displot(data=df, x='Sales', hue='Category');
```



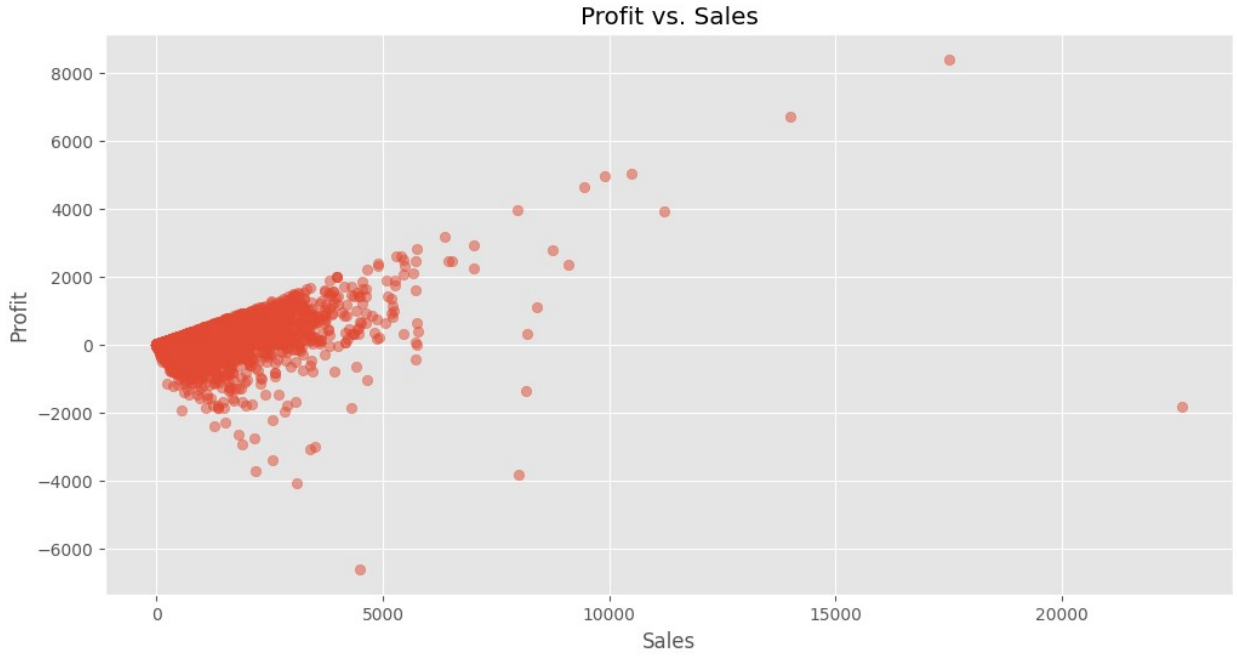

```
sns.displot(data=df, x='Sales', log_scale=True, hue='Sub-Category',
alpha=0.5);
```



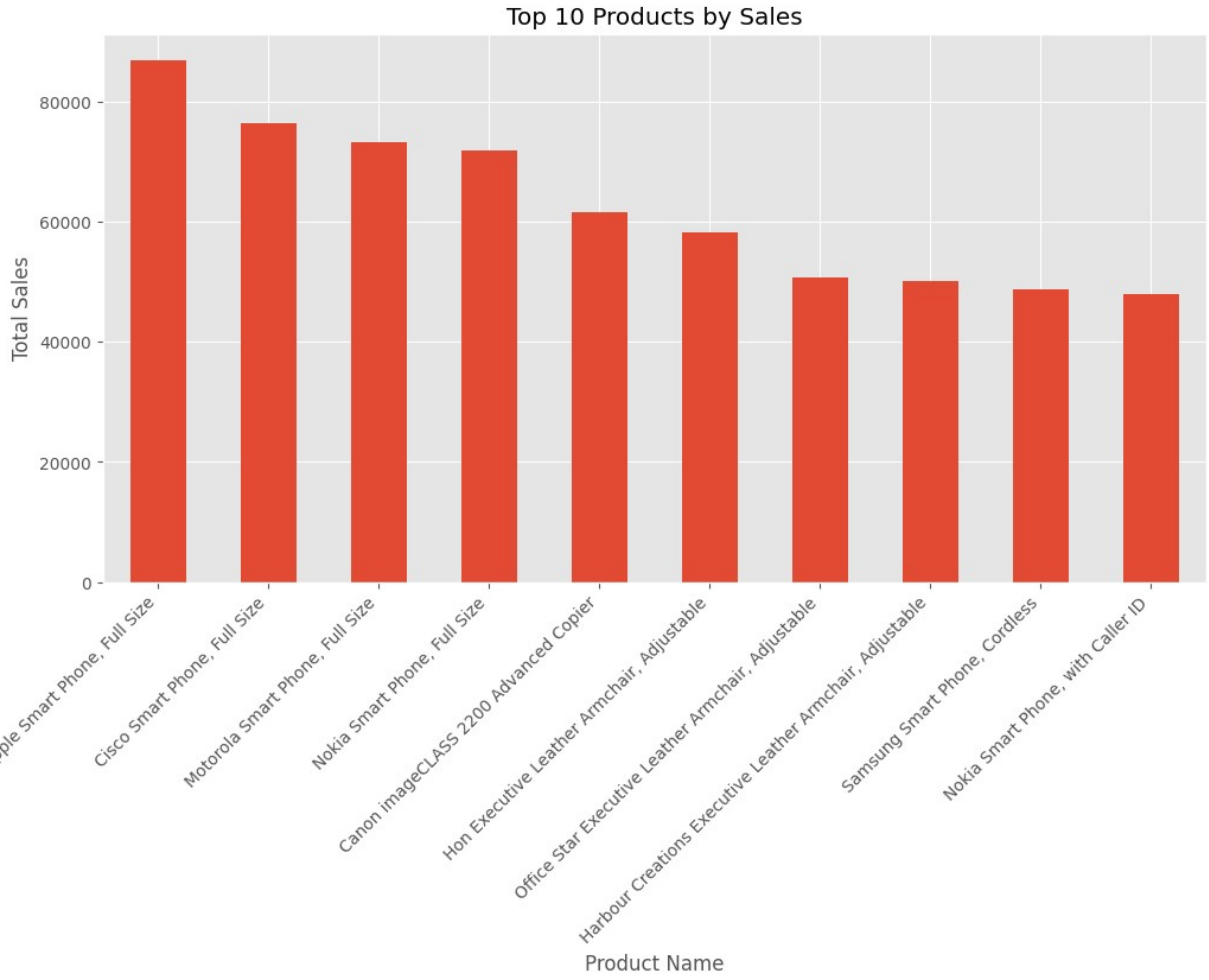

```
plt.figure(figsize=(12, 6))
plt.hist(df['Sales'], bins=50, edgecolor='k')
plt.title('Sales Distribution')
plt.xlabel('Sales')
plt.ylabel('Frequency')
plt.grid(True)
plt.show()
```



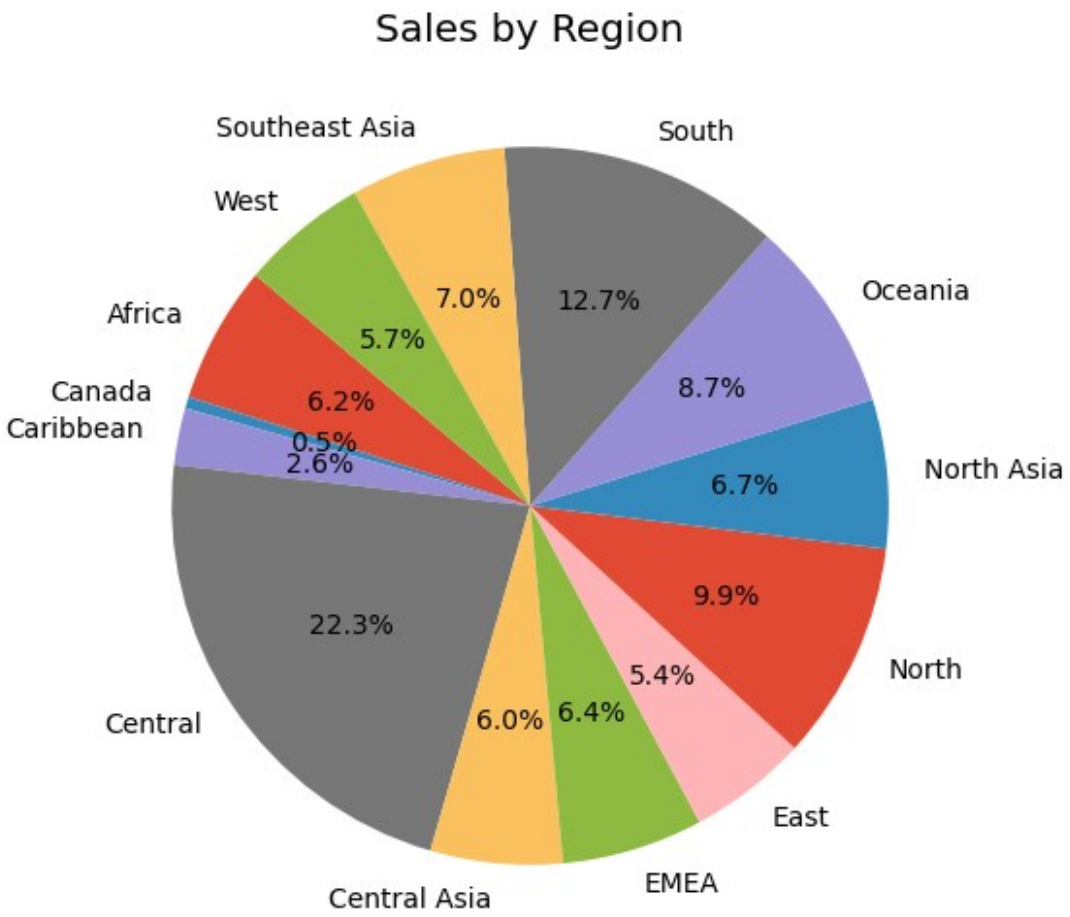

```
plt.figure(figsize=(12, 6))
plt.scatter(df['Sales'], df['Profit'], alpha=0.5)
plt.title('Profit vs. Sales')
plt.xlabel('Sales')
plt.ylabel('Profit')
plt.grid(True)
plt.show()
```



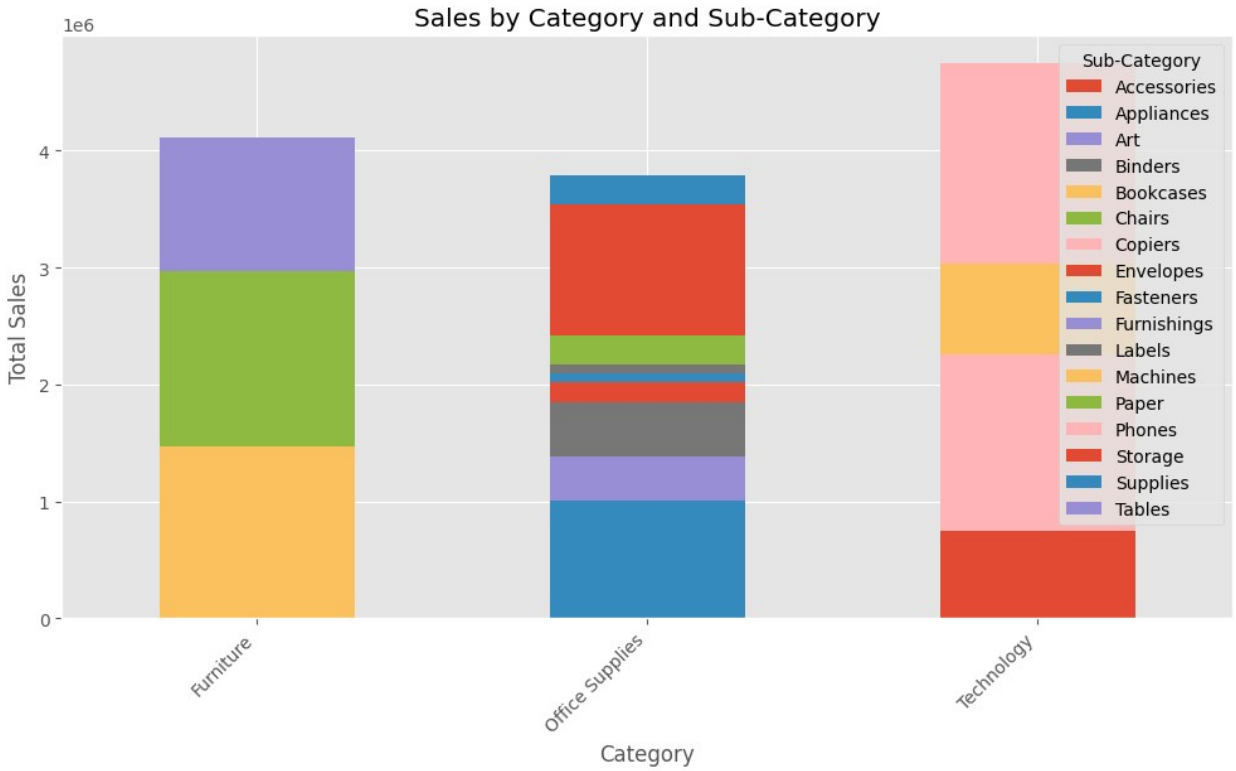
```
top_products = df.groupby('Product Name')
['Sales'].sum().sort_values(ascending=False).head(10)

plt.figure(figsize=(12, 6))
top_products.plot(kind='bar')
plt.title('Top 10 Products by Sales')
plt.xlabel('Product Name')
plt.ylabel('Total Sales')
plt.xticks(rotation=45, ha='right')
plt.grid(True)
plt.show()
```



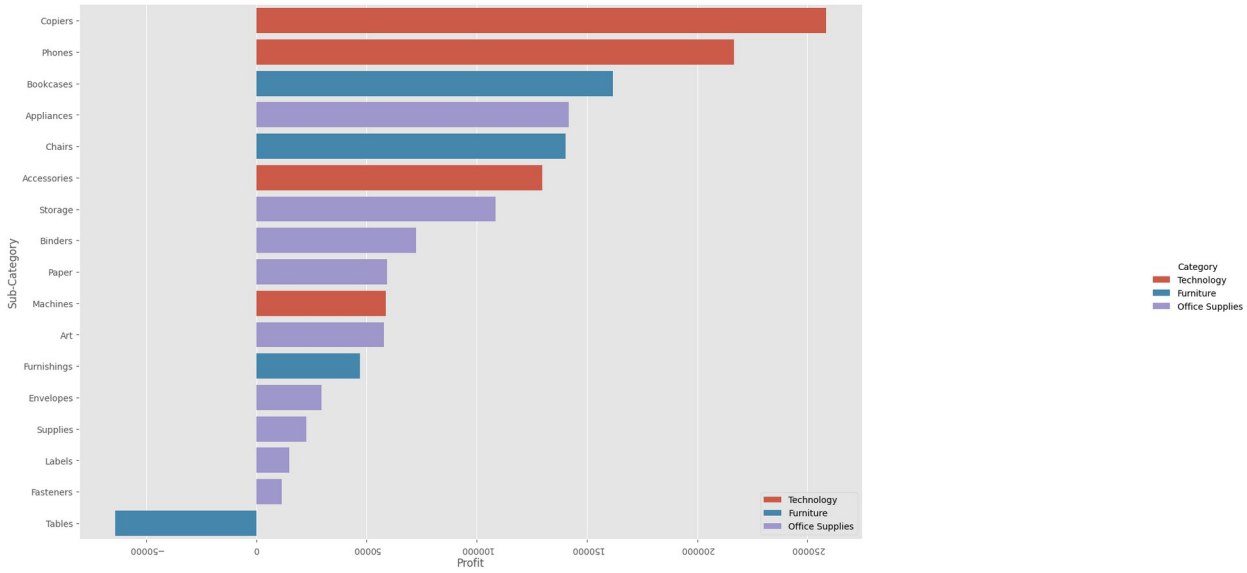

```
region_sales = df.groupby('Region')['Sales'].sum()

plt.figure(figsize=(12, 6))
region_sales.plot(kind='pie', autopct='%1.1f%%', startangle=140)
plt.title('Sales by Region')
plt.ylabel('')
plt.show()
```




```
category_sales = df.groupby(['Category', 'Sub-Category'])
['Sales'].sum().unstack()

category_sales.plot(kind='bar', stacked=True, figsize=(12, 6))
plt.title('Sales by Category and Sub-Category')
plt.xlabel('Category')
plt.ylabel('Total Sales')
plt.xticks(rotation=45, ha='right')
plt.legend(title='Sub-Category')
plt.grid(True)
plt.show()
```




```
Profits = df.groupby(["Category", 'Sub-Category'], as_index=False) \
['Profit'].sum().sort_values('Profit', ascending=False)

f = sns.catplot(data=Profits,
            y='Sub-Category', x='Profit',
            kind='bar', hue='Category', dodge=False)

f.fig.set_size_inches(25, 10)

plt.xticks(rotation=180)

plt.legend(loc='lower right');
```




```
f = sns.catplot(data=Profits,
            y='Sub-Category', x='Profit',
            kind='bar', hue='Sub-Category', dodge=False)

f.fig.set_size_inches(25, 10)

plt.xticks(rotation=180)

plt.legend(loc='lower right');
```