```
# Create the .kaggle directory

!pip install kaggle
!mkdir -p ~/.kaggle


from google.colab import files
files.upload()

# To Copy the uploaded kaggle.json file to the correct location

!cp kaggle.json ~/.kaggle/

# Set secure permissions for API to work

!chmod 600 ~/.kaggle/kaggle.json

!kaggle datasets download -d vivek468/superstore-dataset-final
!unzip superstore-dataset-final.zip
```

Show hidden output

```
# Import required libraries
from sklearn.preprocessing import StandardScaler, OneHotEncoder
from sklearn.cluster import KMeans

from sklearn.compose import ColumnTransformer
from sklearn.model_selection import train_test_split, GridSearchCV, KFold
from sklearn.pipeline import Pipeline
from sklearn.linear_model import LinearRegression
from sklearn.tree import DecisionTreeRegressor
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error, r2_score, mean_absolute_error

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

```
# Explore the Dataset
# To correctly decode the dataset we need to specificly add different encoding format

df = pd.read_csv("Sample - Superstore.csv", encoding='latin-1')

# Retriving the coloumn names

df.columns
```

Show hidden output

```
#First and last 5 rows of the dataset

df.head()
df.tail()
```

Show hidden output

```
# Total no. of rows and columns of the dataset

df.shape
```

```
(9994, 21)
```

```
# Statistical Measures of the dataset

df.describe()
```

Show hidden output

```
# Find any missing or null values in the dataset

if df.isnull().any().any():
  print("There are missing or null values in the data set")
else:
  print("There are no missing or null values in the data set")
```

Show hidden output

```
# Find any duplicate values

if df.duplicated().any():
  print("There are duplicate values in the data set")
else:
  print("There are no duplicate values in the data set")
```

```
There are no duplicate values in the data set
```

```
# Check the retail store data whether it is completey based on USA or not

countries = df['Country'].unique()
if countries != "United States" :
  print("Countries in the data:\n", countries)
else:
  print("This data only based on USA retail stores.")
```

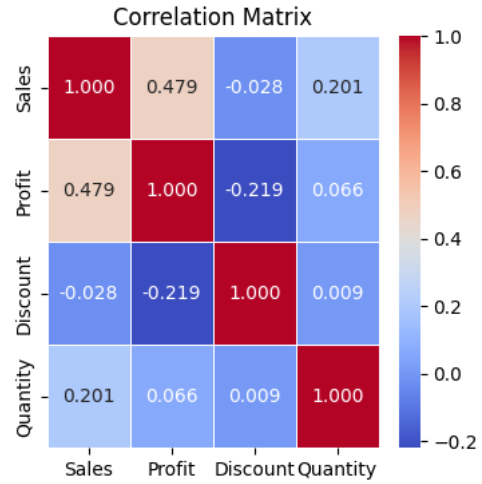
```
This data only based on USA retail stores.
```

```
# Categories and subcategories of the store products

categories_list = df.groupby('Category')['Sub-Category'].unique()
print(categories_list)
```

```
Category
Furniture                   [Bookcases, Chairs, Tables, Furnishings]
Office Supplies    [Labels, Storage, Art, Binders, Appliances, Pa...
Technology                  [Phones, Accessories, Machines, Copiers]
Name: Sub-Category, dtype: object
```

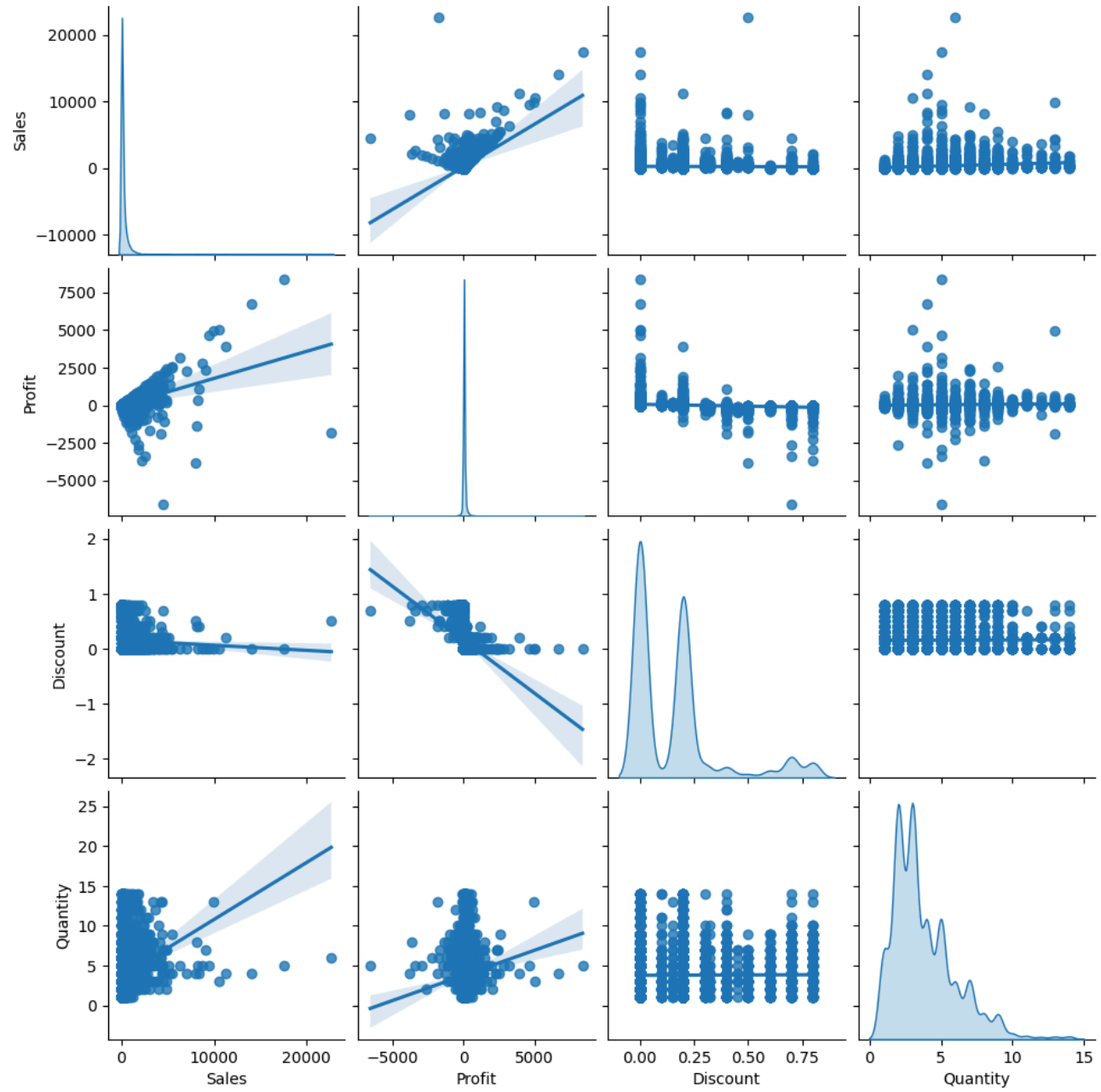
```
# Inorder to find the top 5 profitable products in each regions
# Find the relationship between multiple numerical varibles using heatmap and correlation matrix


data_to_correlate = df[[ 'Sales', 'Profit', 'Discount', 'Quantity']]
correlation_matrix = data_to_correlate.corr()
print(correlation_matrix)

plt.figure(figsize =(4,4))
sns.heatmap(correlation_matrix,
            annot=True,     # Show the correlation value numbers on the map
            cmap='coolwarm', # A color scheme
            fmt=".3f",      # Format numbers to three decimal places
            linewidths=.5,  # Add lines between cells
            cbar=True)      # Show the color bar
plt.title("Correlation Matrix")
```

```
              Sales    Profit  Discount  Quantity
Sales     1.000000  0.479064 -0.028190  0.200795
Profit    0.479064  1.000000 -0.219487  0.066253
Discount -0.028190 -0.219487  1.000000  0.008623
Quantity  0.200795  0.066253  0.008623  1.000000
Text(0.5, 1.0, 'Correlation Matrix')
```




```
# Find the relationship between multiple numerical varibles using pairplot
plt.figure(figsize =(4,4))
sns.pairplot(data_to_correlate,
             diag_kind= "kde",
             kind = "reg",
             height = 2.5)
```

```
<seaborn.axisgrid.PairGrid at 0x7ae33d19dee0>
<Figure size 400x400 with 0 Axes>
```




```
# Group categorical variables and Sum numerical variables
product_per_region = df.groupby(['Region', 'Category', 'Product Name'])[['Sales', 'Profit', 'Discount', 'Quantity']].sum()
print(product_per_region)
```

```
                                                                 Sales  \
Region  Category   Product Name
Central Furniture  12-1/2 Diameter Round Wall Clock                87.912
                   36X48 HARDFLOOR CHAIRMAT                       41.960
                   3M Hangers With Command Adhesive                2.960
                   3M Polarizing Light Filter Sleeves             59.680
                   3M Polarizing Task Lamp with Clamp Arm, Light Gray  1095.840
...                                                                  ...
West    Technology iKross Bluetooth Portable Keyboard + Cell Phone...   100.560
                   iOttie HLCRIO102 Car Mount                      31.984
                   iOttie XL Car Mount                             79.960
                   invisibleSHIELD by ZAGG Smudge-Free Screen Prot...  158.312
                   netTALK DUO VoIP Telephone Service             839.840

                                                                 Profit  \
Region  Category   Product Name
Central Furniture  12-1/2 Diameter Round Wall Clock              -52.7472
                   36X48 HARDFLOOR CHAIRMAT                      -55.5970
                   3M Hangers With Command Adhesive               -1.4060
```

```
                  3M Polarizing Light Filter Sleeves                    9.3250
                  3M Polarizing Task Lamp with Clamp Arm, Light Gray  284.9184
...                                                                        ...
West    Technology iKross Bluetooth Portable Keyboard + Cell Phone...   10.0560
                  iOttie HLCRIO102 Car Mount                           -5.9970
                  iOttie XL Car Mount                                 -17.9910
                  invisibleSHIELD by ZAGG Smudge-Free Screen Prot...   55.4092
                  netTALK DUO VoIP Telephone Service                  314.9400

                                                                      Discount  \
Region  Category   Product Name
Central Furniture  12-1/2 Diameter Round Wall Clock                        1.8
                  36X48 HARDFLOOR CHAIRMAT                                 1.2
                  3M Hangers With Command Adhesive                         0.6
                  3M Polarizing Light Filter Sleeves                       0.6
                  3M Polarizing Task Lamp with Clamp Arm, Light Gray       0.0
...                                                                        ...
West    Technology iKross Bluetooth Portable Keyboard + Cell Phone...      0.4
                  iOttie HLCRIO102 Car Mount                               0.4
                  iOttie XL Car Mount                                      0.2
                  invisibleSHIELD by ZAGG Smudge-Free Screen Prot...       0.6
                  netTALK DUO VoIP Telephone Service                       1.0

                                                                     Quantity
Region  Category   Product Name
Central Furniture  12-1/2 Diameter Round Wall Clock                       11
                  36X48 HARDFLOOR CHAIRMAT                                 5
                  3M Hangers With Command Adhesive                         2
                  3M Polarizing Light Filter Sleeves                       5
                  3M Polarizing Task Lamp with Clamp Arm, Light Gray       8
...                                                                      ...
West    Technology iKross Bluetooth Portable Keyboard + Cell Phone...      6
                  iOttie HLCRIO102 Car Mount                               2
                  iOttie XL Car Mount                                      5
                  invisibleSHIELD by ZAGG Smudge-Free Screen Prot...      11
                  netTALK DUO VoIP Telephone Service                      20

[5255 rows x 4 columns]
```

```
# Top five profitable products per region using sort and groupby method
top5_profit_per_region = product_per_region.sort_values(['Region', 'Profit'], ascending= [True, False]). groupby('Region').heac
print(top5_profit_per_region)
```

```
                      Zebra ZM400 Thermal Label Printer                  3343.5360
                      Hewlett Packard LaserJet 3310 Copier               3023.9496
South   Office Supplies Fellowes PB500 Electric Punch Plastic Comb Bind...  3812.9700
        Technology      HP Designjet T520 Inkjet Large Format Printer -...  1854.9894
                      Hewlett-Packard Deskjet 3050a All-in-One Color ...  1459.2000
                      Hewlett Packard LaserJet 3310 Copier               1439.9760
                      Cisco 9971 IP Video Phone Charcoal                 1416.8000
```

```
        Technology      Canon PC1060 Personal Laser Copier                      7
        Office Supplies Ibico EPK-21 Electric Binding System                   10
                        Honeywell Enviracaire Portable HEPA Air Cleaner...     13
East    Technology      Canon imageCLASS 2200 Advanced Copier                  11
                        Ativa V4110MDD Micro-Cut Shredder                      11
                        3D Systems Cube Printer, 2nd Generation, Magenta       11
                        Zebra ZM400 Thermal Label Printer                       6
                        Hewlett Packard LaserJet 3310 Copier                   18
South   Office Supplies Fellowes PB500 Electric Punch Plastic Comb Bind...     6
        Technology      HP Designjet T520 Inkjet Large Format Printer -...      8
                        Hewlett-Packard Deskjet 3050a All-in-One Color ...      8
                        Hewlett Packard LaserJet 3310 Copier                    5
                        Cisco 9971 IP Video Phone Charcoal                      7
West    Technology      Canon imageCLASS 2200 Advanced Copier                   4
        Office Supplies Fellowes PB500 Electric Punch Plastic Comb Bind...      8
        Technology      Canon PC1060 Personal Laser Copier                     12
                        Hewlett Packard LaserJet 3310 Copier                   13
                        Logitech Z-906 Speaker sys - home theater - 5.1-CH     13
```

```
# Average discount given on these most profitable products using aggregation method
disc_on_top5_profit = top5_profit_per_region.groupby(['Region', 'Product Name']).agg({
    'Discount': 'mean',
    'Sales': 'sum',
    'Profit': 'sum'
}).reset_index()

print(disc_on_top5_profit)
```

```
     Region                                    Product Name  Discount  \
0   Central               Canon PC1060 Personal Laser Copier       0.0
1   Central             Canon imageCLASS 2200 Advanced Copier       0.0
2   Central   GBC Ibimaster 500 Manual ProClick Binding System       0.8
3   Central  Honeywell Enviracaire Portable HEPA Air Cleane...       0.0
4   Central               Ibico EPK-21 Electric Binding System       0.8
5      East   3D Systems Cube Printer, 2nd Generation, Magenta       0.0
6      East                  Ativa V4110MDD Micro-Cut Shredder       0.0
7      East             Canon imageCLASS 2200 Advanced Copier       0.6
8      East              Hewlett Packard LaserJet 3310 Copier       0.8
9      East                  Zebra ZM400 Thermal Label Printer       0.0
10    South                 Cisco 9971 IP Video Phone Charcoal       0.0
11    South  Fellowes PB500 Electric Punch Plastic Comb Bin...       0.0
12    South  HP Designjet T520 Inkjet Large Format Printer ...       0.5
13    South               Hewlett Packard LaserJet 3310 Copier       0.0
14    South  Hewlett-Packard Deskjet 3050a All-in-One Color...       0.0
15     West                 Canon PC1060 Personal Laser Copier       0.6
16     West             Canon imageCLASS 2200 Advanced Copier       0.0
17     West  Fellowes PB500 Electric Punch Plastic Comb Bin...       0.6
18     West               Hewlett Packard LaserJet 3310 Copier       0.6
19     West  Logitech Z-906 Speaker sys - home theater - 5....       0.0

        Sales      Profit
0    4899.930   2302.9671
1   17499.950   8399.9760
2   10653.720   3804.9000
3    3908.450   1289.7885
4   11339.940   1700.9910
5   14299.890   3717.9714
6    7699.890   3772.9461
7   30099.914  10079.9712
8    8639.856   3023.9496
9    6965.700   3343.5360
10   3080.000   1416.8000
11   7625.940   3812.9700
12  11374.935   1854.9894
13   2999.950   1439.9760
14   3040.000   1459.2000
15   6719.904   2267.9676
16  13999.960   6719.9808
17   8134.336   3050.3760
18   6239.896   2183.9636
19   4289.870   1715.9480
```

```
# Top 5 loss-making products per region using sort and groupby method
top5_loss_per_region = product_per_region.sort_values(['Region', 'Profit'], ascending= [True, True]). groupby('Region').head(5)
print(top5_loss_per_region)
```

```
West    Technology      Lexmark MX611dhe Monochrome Laser Printer           3399.9800
```

```
West    Technology      Lexmark MX611dhe Monochrome Laser Printer          -3399.9800
                        Zebra GK420t Direct Thermal/Thermal Transfer Pr...  -938.2800
        Furniture       O'Sullivan 4-Shelf Bookcase in Odessa Pine          -802.0974
                        Hon 2090 "Pillow Soft" Series Mid Back Swivel/T...  -547.9110
                        Atlantic Metals Mobile 4-Shelf Bookcases, Custo...  -491.7150

                                                                           Discount  \
Region  Category        Product Name
Central Office Supplies GBC DocuBind P400 Electric Binding System              1.60
                        Fellowes PB500 Electric Punch Plastic Comb Bind...     1.60
                        3.6 Cubic Foot Counter Height Office Refrigerator      2.40
        Technology      Lexmark MX611dhe Monochrome Laser Printer              0.90
        Office Supplies Ibico Hi-Tech Manual Binding System                    3.20
East    Technology      Cubify CubeX 3D Printer Double Head Print              1.40
        Office Supplies Martin Yale Chadless Opener Electric Letter Opener     0.40
        Furniture       Riverside Furniture Oval Coffee Table, Oval End...     1.20
        Office Supplies GBC Ibimaster 500 Manual ProClick Binding System       2.50
        Technology      Cisco 9971 IP Video Phone Charcoal                     0.70
South   Technology      Cubify CubeX 3D Printer Triple Head Print              0.50
        Furniture       Chromcraft Bull-Nose Wood Oval Conference Table...     0.80
        Office Supplies GBC Ibimaster 500 Manual ProClick Binding System       1.40
        Technology      Cisco TelePresence System EX90 Videoconferencin...     0.50
        Office Supplies GBC DocuBind P400 Electric Binding System              0.70
West    Technology      Lexmark MX611dhe Monochrome Laser Printer              0.70
                        Zebra GK420t Direct Thermal/Thermal Transfer Pr...     0.70
        Furniture       O'Sullivan 4-Shelf Bookcase in Odessa Pine             2.25
                        Hon 2090 "Pillow Soft" Series Mid Back Swivel/T...     0.80
                        Atlantic Metals Mobile 4-Shelf Bookcases, Custo...     1.00

                                                                           Quantity
Region  Category        Product Name
Central Office Supplies GBC DocuBind P400 Electric Binding System               16
                        Fellowes PB500 Electric Punch Plastic Comb Bind...      12
                        3.6 Cubic Foot Counter Height Office Refrigerator        9
        Technology      Lexmark MX611dhe Monochrome Laser Printer               13
        Office Supplies Ibico Hi-Tech Manual Binding System                     18
East    Technology      Cubify CubeX 3D Printer Double Head Print                7
        Office Supplies Martin Yale Chadless Opener Electric Letter Opener       8
        Furniture       Riverside Furniture Oval Coffee Table, Oval End...      23
        Office Supplies GBC Ibimaster 500 Manual ProClick Binding System        17
        Technology      Cisco 9971 IP Video Phone Charcoal                       9
South   Technology      Cubify CubeX 3D Printer Triple Head Print                4
        Furniture       Chromcraft Bull-Nose Wood Oval Conference Table...      20
        Office Supplies GBC Ibimaster 500 Manual ProClick Binding System        13
        Technology      Cisco TelePresence System EX90 Videoconferencin...       6
        Office Supplies GBC DocuBind P400 Electric Binding System                4
West    Technology      Lexmark MX611dhe Monochrome Laser Printer                5
                        Zebra GK420t Direct Thermal/Thermal Transfer Pr...       6
        Furniture       O'Sullivan 4-Shelf Bookcase in Odessa Pine               26
                        Hon 2090 "Pillow Soft" Series Mid Back Swivel/T...      15
                        Atlantic Metals Mobile 4-Shelf Bookcases, Custo...      14
```

```
# Average discount given on these Top 5 loss-making  products using aggregation method
disc_on_top5_loss = top5_loss_per_region.groupby(['Region', 'Product Name']).agg({
    'Discount': 'mean',
    'Sales': 'sum',
    'Profit': 'sum'
}).reset_index()

print(disc_on_top5_loss)
```

```
     Region                                    Product Name  Discount  \
0   Central  3.6 Cubic Foot Counter Height Office Refrigerator      2.40
1   Central  Fellowes PB500 Electric Punch Plastic Comb Bin...      1.60
2   Central          GBC DocuBind P400 Electric Binding System      1.60
3   Central                Ibico Hi-Tech Manual Binding System      3.20
4   Central          Lexmark MX611dhe Monochrome Laser Printer      0.90
5      East                 Cisco 9971 IP Video Phone Charcoal      0.70
6      East          Cubify CubeX 3D Printer Double Head Print      1.40
7      East   GBC Ibimaster 500 Manual ProClick Binding System      2.50
8      East  Martin Yale Chadless Opener Electric Letter Op...      0.40
9      East  Riverside Furniture Oval Coffee Table, Oval En...      1.20
10    South  Chromcraft Bull-Nose Wood Oval Conference Tabl...      0.80
11    South  Cisco TelePresence System EX90 Videoconferenci...      0.50
12    South          Cubify CubeX 3D Printer Triple Head Print      0.50
13    South          GBC DocuBind P400 Electric Binding System      0.70
14    South   GBC Ibimaster 500 Manual ProClick Binding System      1.40
15     West  Atlantic Metals Mobile 4-Shelf Bookcases, Cust...      1.00
16     West  Hon 2090 "Pillow Soft" Series Mid Back Swivel/...      0.80
17     West          Lexmark MX611dhe Monochrome Laser Printer      0.70
18     West         O'Sullivan 4-Shelf Bookcase in Odessa Pine      2.25
19     West  Zebra GK420t Direct Thermal/Thermal Transfer P...      0.70

        Sales     Profit
```

```
0     530.316 -1378.8216
1    6100.752 -1525.1880
2    8710.336 -3048.6176
3    1829.940 -1189.4610
4   14279.916 -1189.9930
5    1188.000  -950.4000
6    6299.979 -9239.9692
7    5402.958 -1065.3720
8    5329.984 -1199.2464
9    3958.530 -1187.5590
10   6611.760 -2865.0960
11  22638.480 -1811.0784
12   7999.980 -3839.9904
13   1633.188 -1306.5504
14   2967.822 -1978.5480
15   2261.889  -491.7150
16   3371.760  -547.9110
17   2549.985 -3399.9800
18   1808.651  -802.0974
19    703.710  -938.2800
```

```
# Overall top 5 profitable products and loss-making products across the country

top5_profit_overall = (product_per_region.sort_values('Profit', ascending= False)
                    .head(5) )

print('\033[1m', "\n Overall Top 5 Profit-Making products across the country:\n", '\033[0m', top5_profit_overall)

#filter the small profits out to get actual loss-making products

top5_loss_overall =  (product_per_region[product_per_region['Profit'] < 0]
                    .sort_values('Profit', ascending= True)
                    .head(5) )
print('\033[1m', "\n  Overall Top 5 Loss-Making products across the country:\n", '\033[0m', top5_loss_overall)
```

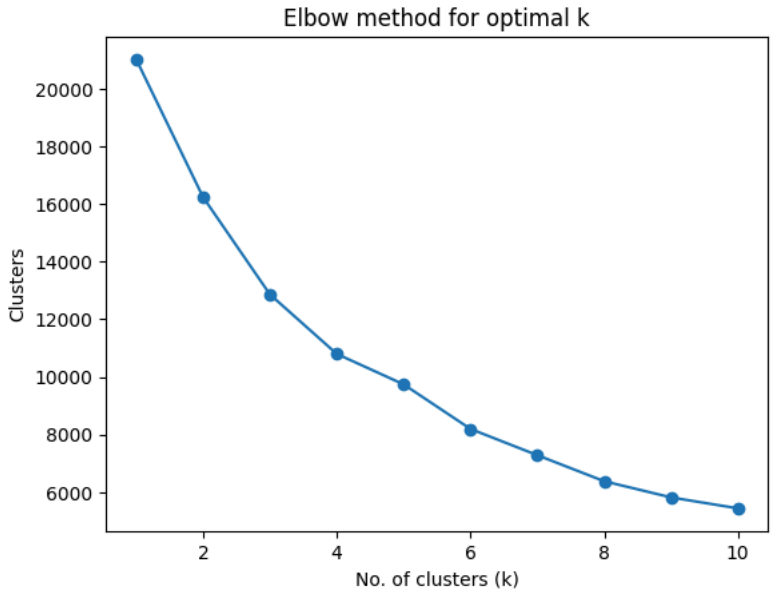
```
Central Office Supplies GBC Ibimaster 500 Manual ProClick Binding System    10653.720

                                                                                Profit  \
Region  Category        Product Name
East    Technology      Canon imageCLASS 2200 Advanced Copier               10079.9712
Central Technology      Canon imageCLASS 2200 Advanced Copier                8399.9760
West    Technology      Canon imageCLASS 2200 Advanced Copier                6719.9808
South   Office Supplies Fellowes PB500 Electric Punch Plastic Comb Bind...   3812.9700
Central Office Supplies GBC Ibimaster 500 Manual ProClick Binding System     3804.9000

                                                                            Discount  \
Region  Category        Product Name
East    Technology      Canon imageCLASS 2200 Advanced Copier                    0.6
Central Technology      Canon imageCLASS 2200 Advanced Copier                    0.0
West    Technology      Canon imageCLASS 2200 Advanced Copier                    0.0
South   Office Supplies Fellowes PB500 Electric Punch Plastic Comb Bind...      0.0
Central Office Supplies GBC Ibimaster 500 Manual ProClick Binding System        0.8

                                                                            Quantity
Region  Category        Product Name
East    Technology      Canon imageCLASS 2200 Advanced Copier                     11
Central Technology      Canon imageCLASS 2200 Advanced Copier                      5
West    Technology      Canon imageCLASS 2200 Advanced Copier                      4
South   Office Supplies Fellowes PB500 Electric Punch Plastic Comb Bind...        6
Central Office Supplies GBC Ibimaster 500 Manual ProClick Binding System         18

  Overall Top 5 Loss-Making products across the country:
                                                                                Sales  \
Region  Category        Product Name
East    Technology      Cubify CubeX 3D Printer Double Head Print          6299.979
South   Technology      Cubify CubeX 3D Printer Triple Head Print          7999.980
West    Technology      Lexmark MX611dhe Monochrome Laser Printer          2549.985
Central Office Supplies GBC DocuBind P400 Electric Binding System          8710.336
South   Furniture       Chromcraft Bull-Nose Wood Oval Conference Table...  6611.760

                                                                              Profit  \
Region  Category        Product Name
East    Technology      Cubify CubeX 3D Printer Double Head Print         -9239.9692
South   Technology      Cubify CubeX 3D Printer Triple Head Print         -3839.9904
```

```
West    Technology      Lexmark MX611dhe Monochrome Laser Printer               0.7
Central Office Supplies GBC DocuBind P400 Electric Binding System              1.6
South   Furniture       Chromcraft Bull-Nose Wood Oval Conference Table...     0.8

                                                                          Quantity
Region  Category        Product Name
East    Technology      Cubify CubeX 3D Printer Double Head Print                 7
South   Technology      Cubify CubeX 3D Printer Triple Head Print                 4
West    Technology      Lexmark MX611dhe Monochrome Laser Printer                 5
Central Office Supplies GBC DocuBind P400 Electric Binding System                16
South   Furniture       Chromcraft Bull-Nose Wood Oval Conference Table           20
```

```
# ML k-means algorithm to understand hidden patterns

X = product_per_region[['Sales', 'Profit', 'Discount', 'Quantity']]
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)
```

```
# To find optimal no. of clusters

inertia = []
K = range(1,11)

for k in K:
  kmeans = KMeans(n_clusters=k, random_state=42)
  kmeans.fit_predict(X_scaled)
  inertia.append(kmeans.inertia_)
```

```
# Plot using elbow method for optimal k

plt.plot(range(1,11), inertia, marker = 'o')
plt.xlabel('No. of clusters (k)')
plt.ylabel('Clusters')
plt.title('Elbow method for optimal k')
plt.show()
```




```
kmeans = KMeans(n_clusters=4, random_state=42)
clusters = kmeans.fit_predict(X_scaled)

product_per_region['Clusters'] = clusters
cluster_summary = product_per_region.groupby('Clusters')[['Sales', 'Profit', 'Discount', 'Quantity']].mean()
print(cluster_summary)
```

```
                Sales       Profit  Discount   Quantity
Clusters
0          748.329491   110.462234  0.311622  12.600144
1         7638.955632  1793.171956  0.403509  12.368421
2          552.488491  -152.831469  1.565377  14.746575
3          187.313057    21.331943  0.184147   4.358098
```

```
# Find clusters with maximum average profit using the index of cluster_summary
profit_clusters_index = cluster_summary['Profit'].idxmax()

# Filter products that belong to those clusters
max_profit_products = product_per_region[product_per_region['Clusters'] == profit_clusters_index]
print("\n Products with Maximum Average Profit Per Region \n", max_profit_products)
```

```
 Products with Maximum Average Profit Per Region
                                                                      Sales  \
Region  Category        Product Name
Central Furniture       HON 5400 Series Task Chairs for Big and Tall         6939.702
        Office Supplies GBC Ibimaster 500 Manual ProClick Binding System    10653.720
                        Honeywell Enviracaire Portable HEPA Air Cleaner...   3908.450
                        Ibico EPK-21 Electric Binding System                11339.940
        Technology      Canon PC1060 Personal Laser Copier                   4899.930
                        Canon imageCLASS 2200 Advanced Copier               17499.950
                        Lexmark MX611dhe Monochrome Laser Printer           14279.916
East    Furniture       DMI Eclipse Executive Suite Bookcases                7214.112
                        HON 5400 Series Task Chairs for Big and Tall         8201.466
                        Riverside Palais Royal Lawyers Bookcase, Royale...  11717.034
        Office Supplies Adjustable Depth Letter/Legal Cart                   4064.704
                        Deluxe Rollaway Locking File with Drawer             4325.152
                        Dual Level, Single-Width Filing Carts                4496.740
                        Fellowes PB500 Electric Punch Plastic Comb Bind...   5592.356
                        GBC DocuBind P400 Electric Binding System            4355.168
                        GBC DocuBind TL300 Electric Binding System           8790.502
                        Honeywell Enviracaire Portable HEPA Air Cleaner...   4329.360
        Technology      3D Systems Cube Printer, 2nd Generation, Magenta    14299.890
                        Apple iPhone 5                                       7797.960
                        Ativa V4110MDD Micro-Cut Shredder                    7699.890
                        Bady BDG101FRU Card Printer                          3999.950
                        Canon imageCLASS 2200 Advanced Copier               30099.914
                        Canon imageCLASS MF7460 Monochrome Digital Lase...   3991.980
                        HP Designjet T520 Inkjet Large Format Printer -...   6999.960
                        Hewlett Packard 610 Color Digital Copier / Printer   3699.926
                        Hewlett Packard LaserJet 3310 Copier                 8639.856
                        Hewlett-Packard Desktjet 6988DT Refurbished Pri...   3404.500
                        Plantronics CS510 - Over-the-Head monaural Wire...   5081.230
                        Plantronics Savi W720 Multi-Device Wireless Hea...   4557.060
                        Sharp AL-1530CS Digital Copier                       3999.920
                        Zebra ZM400 Thermal Label Printer                    6965.700
South   Office Supplies Fellowes PB500 Electric Punch Plastic Comb Bind...   7625.940
                        GBC DocuBind TL300 Electric Binding System           8342.007
        Technology      Cisco 9971 IP Video Phone Charcoal                   3080.000
                        Cisco TelePresence System EX90 Videoconferencin...  22638.480
                        HP Designjet T520 Inkjet Large Format Printer -...  11374.935
                        Hewlett Packard LaserJet 3310 Copier                 2999.950
                        Hewlett-Packard Deskjet 3050a All-in-One Color ...   3040.000
                        Samsung Galaxy Mega 6.3                              7307.826
West    Furniture       Bretford Rectangular Conference Table Tops           7710.665
                        Global Troy Executive Leather Low-Back Tilter       10019.600
                        GuestStacker Chair with Chrome Finish Legs           8030.016
                        SAFCO Arco Folding Chair                             5744.960
        Office Supplies Fellowes PB300 Plastic Comb Binding Machine          3724.704
                        Fellowes PB500 Electric Punch Plastic Comb Bind...   8134.336
                        High Speed Automatic Electric Letter Opener         13100.240
                        Ibico EPK-21 Electric Binding System                 4535.976
                        Tennsco 6- and 18-Compartment Lockers                5144.298
        Technology      Canon PC1060 Personal Laser Copier                   6719.904
                        Canon PC940 Copier                                   3149.930
                        Canon imageCLASS 2200 Advanced Copier               13999.960
                        Hewlett Packard LaserJet 3310 Copier                 6239.896
                        Logitech P710e Mobile Speakerphone                   7467.210
                        Logitech Z-906 Speaker sys - home theater - 5.1-CH   4289.870
```

```
# Summary of maximum profitable products

max_profit_summary = max_profit_products.groupby(['Region', 'Product Name'])['Profit'].sum().reset_index()
print(max_profit_summary)
```

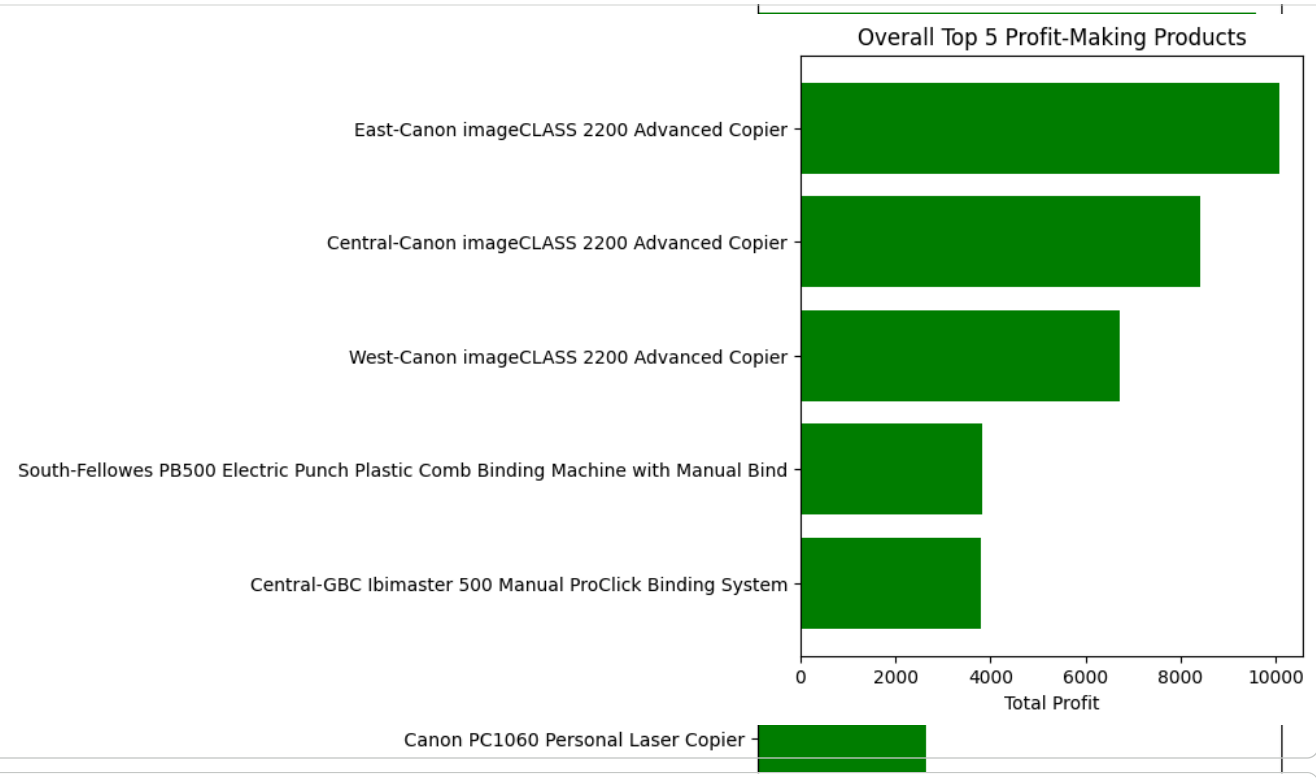
```
     Region                                    Product Name      Profit
0   Central              Canon PC1060 Personal Laser Copier   2302.9671
1   Central            Canon imageCLASS 2200 Advanced Copier   8399.9760
2   Central   GBC Ibimaster 500 Manual ProClick Binding System   3804.9000
3   Central       HON 5400 Series Task Chairs for Big and Tall    210.2940
4   Central  Honeywell Enviracaire Portable HEPA Air Cleane...   1289.7885
5   Central              Ibico EPK-21 Electric Binding System   1700.9910
6   Central         Lexmark MX611dhe Monochrome Laser Printer  -1189.9930
7      East   3D Systems Cube Printer, 2nd Generation, Magenta   3717.9714
8      East                 Adjustable Depth Letter/Legal Cart   1101.4622
9      East                                      Apple iPhone 5   1156.6974
10     East                  Ativa V4110MDD Micro-Cut Shredder   3772.9461
11     East                          Bady BDG101FRU Card Printer   1150.0855
```

11	East	Bady BDG101FRO Card Printer	1159.9855
12	East	Canon imageCLASS 2200 Advanced Copier	10079.9712
13	East	Canon imageCLASS MF7460 Monochrome Digital Las...	1995.9900
14	East	DMI Eclipse Executive Suite Bookcases	-90.1764
15	East	Deluxe Rollaway Locking File with Drawer	1031.3824
16	East	Dual Level, Single-Width Filing Carts	1169.1524
17	East	Fellowes PB500 Electric Punch Plastic Comb Bin...	2414.8810
18	East	GBC DocuBind P400 Electric Binding System	1415.4296
19	East	GBC DocuBind TL300 Electric Binding System	2610.2409
20	East	HON 5400 Series Task Chairs for Big and Tall	-210.2940
21	East	HP Designjet T520 Inkjet Large Format Printer ...	2239.9872
22	East	Hewlett Packard 610 Color Digital Copier / Pri...	1224.9755
23	East	Hewlett Packard LaserJet 3310 Copier	3023.9496
24	East	Hewlett-Packard Desktjet 6988DT Refurbished Pr...	1668.2050
25	East	Honeywell Enviracaire Portable HEPA Air Cleane...	1106.3920
26	East	Plantronics CS510 - Over-the-Head monaural Wir...	1323.0995
27	East	Plantronics Savi W720 Multi-Device Wireless He...	1721.5560
28	East	Riverside Palais Royal Lawyers Bookcase, Royal...	-493.3488
29	East	Sharp AL-1530CS Digital Copier	1194.9761
30	East	Zebra ZM400 Thermal Label Printer	3343.5360
31	South	Cisco 9971 IP Video Phone Charcoal	1416.8000
32	South	Cisco TelePresence System EX90 Videoconferenci...	-1811.0784
33	South	Fellowes PB500 Electric Punch Plastic Comb Bin...	3812.9700
34	South	GBC DocuBind TL300 Electric Binding System	260.1271
35	South	HP Designjet T520 Inkjet Large Format Printer ...	1854.9894
36	South	Hewlett Packard LaserJet 3310 Copier	1439.9760
37	South	Hewlett-Packard Deskjet 3050a All-in-One Color...	1459.2000
38	South	Samsung Galaxy Mega 6.3	1091.9740
39	West	Bretford Rectangular Conference Table Tops	180.5424
40	West	Canon PC1060 Personal Laser Copier	2267.9676
41	West	Canon PC940 Copier	1480.4671
42	West	Canon imageCLASS 2200 Advanced Copier	6719.9808
43	West	Fellowes PB300 Plastic Comb Binding Machine	1257.0876
44	West	Fellowes PB500 Electric Punch Plastic Comb Bin...	3050.3760
45	West	Global Troy Executive Leather Low-Back Tilter	626.2250
46	West	GuestStacker Chair with Chrome Finish Legs	803.0016
47	West	Hewlett Packard LaserJet 3310 Copier	2183.9636
48	West	High Speed Automatic Electric Letter Opener	524.0096
49	West	Ibico EPK-21 Electric Binding System	1644.2913
50	West	Logitech P710e Mobile Speakerphone	1418.7699
51	West	Logitech Z-906 Speaker sys - home theater - 5....	1715.9480
52	West	Okidata MB760 Printer	881.3700
53	West	Plantronics Savi W720 Multi-Device Wireless He...	1670.9220
54	West	SAFCO Arco Folding Chair	646.3080
55	West	Tennsco 6- and 18-Compartment Lockers	795.5100
56	West	Zebra GX420t Direct Thermal/Thermal Transfer P...	621.2115

```
# Plot the Top 5 profit-making products per region

regions = df['Region'].unique()


fig, axes = plt.subplots(len(regions), 1 , figsize=(10,4*len(regions)))

axes = axes if len(regions)>1 else [axes]

for ax, region in zip(axes, regions):
  regional_data = max_profit_summary[max_profit_summary['Region'] == region].sort_values('Profit', ascending=False).head(5)
  ax.barh(regional_data['Product Name'], regional_data['Profit'], color = 'Green')
  ax.set_title(f"Top 5 Profit-Making Products in {region}")
  ax.invert_yaxis()

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

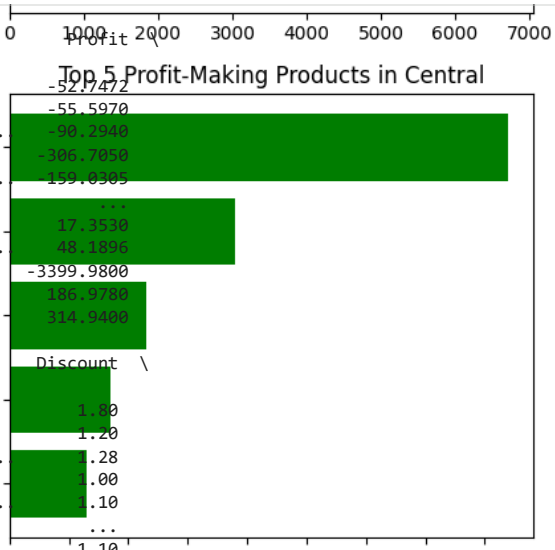
```
# Plot the Overall Top 5 Profit-Making Products across the country

max_profit_overall_summary = max_profit_products.sort_values('Profit', ascending= False).head()

plt.figure(figsize=(5,6))
plt.barh(max_profit_overall_summary.index.get_level_values('Region') + '-'+ max_profit_overall_summary.index.get_level_values('
plt.xlabel('Total Profit')
plt.title('Overall Top 5 Profit-Making Products')
```

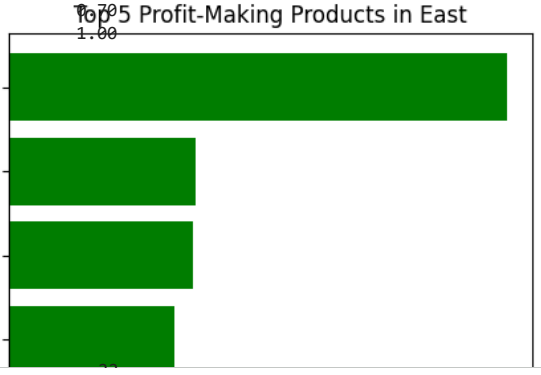
```
plt.gca().invert_yaxis()
plt.show()
```



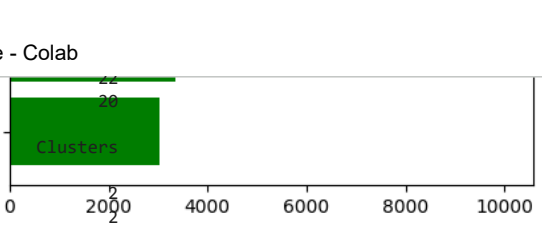

```
# Find clusters with negative average profit using the index of cluster_summary
loss_clusters_index = cluster_summary['Profit'].idxmin()

# Filter products that belong to those clusters
loss_products = product_per_region[product_per_region['Clusters'] == loss_clusters_index]
print(loss_products)
```

```
Region  Category        Product Name                                    Profit  \
Central Furniture       12-1/2 Diameter Round Wall Clock               -52.7472
                        36X48 HARDFLOOR CHAIRMAT                       -55.5970
                        Atlantic Metals Mobile 5-Shelf Bookcases, Custo..  -90.2940
                        BPI Conference Tables                         -306.7050
                        Chromcraft Bull-Nose Wood Round Conference Tabl..  -159.0305
...                                                                        ...
West    Office Supplies Wilson Jones Impact Binders                     17.3530
                        Wilson Jones Leather-Like Binders with DublLock..  48.1896
        Technology      Lexmark MX611dhe Monochrome Laser Printer      -3399.9800
                        Panasonic KX TS3282W Corded phone               186.9780
                        netTALK DUO VoIP Telephone Service              314.9400

                                                                Discount  \
Region  Category        Product Name
Central Furniture       12-1/2 Diameter Round Wall Clock            1.80
                        36X48 HARDFLOOR CHAIRMAT                    1.20
                        Atlantic Metals Mobile 5-Shelf Bookcases, Custo..  1.28
                        BPI Conference Tables                       1.00
                        Chromcraft Bull-Nose Wood Round Conference Tabl..  1.10
...                                                                      ...
West    Office Supplies Wilson Jones Impact Binders                 1.10
                        Wilson Jones Leather-Like Binders with DublLock...  1.70
        Technology      Lexmark MX611dhe Monochrome Laser Printer    1.00
                        Panasonic KX TS3282W Corded phone            1.00
```





```
                        Panasonic KX TS3282W Corded phone                  22
                        netTALK DUO VoIP Telephone Service                 20
                                    Hewlett Packard LaserJet 3310 Copier
                                                                           Clusters
Region  Category        Product Name                                        2
Central Furniture       12-1/2 Diameter Round Wall Clock                    2
                        36X48 HARDFLOOR CHAIRMAT                            2
                        Atlantic Metals Mobile 5-Shelf Bookcases, Custo...  2
                        BPI Conference Tables                               2
                        Chromcraft Bull-Nose Wood Round Conference Tabl...  2
...                                                                        ...
West    Office Supplies Wilson Jones Impact Binders                         2
                        Wilson Jones Leather-Like Binders with DublLock...  2
        Technology      Lexmark MX611dhe Monochrome Laser Printer           2
                        Panasonic KX TS3282W Corded phone                   2
                        netTALK DUO VoIP Telephone Service                  2

[292 rows x 5 columns]
```



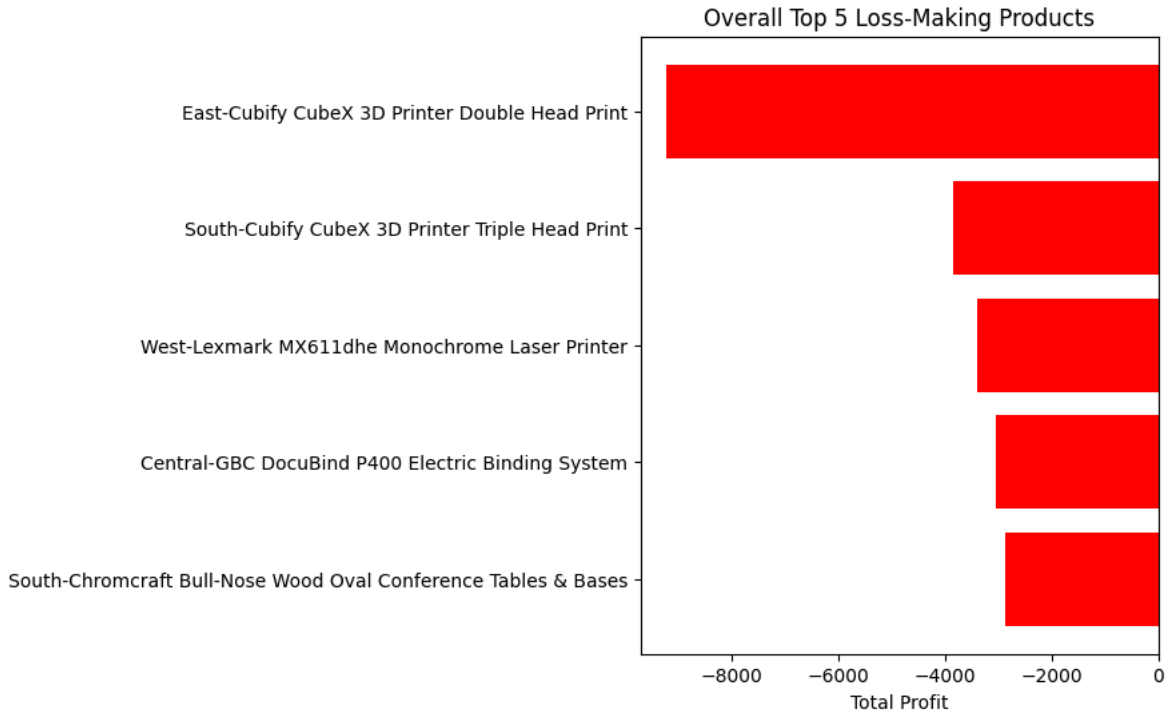

```
# Summary of negative profitable products

loss_summary = loss_products.groupby(['Region', 'Product Name'])['Profit'].sum().reset_index()
print(loss_summary)
```

```
      Region                                     Product Name     Profit
0    Central                  12-1/2 Diameter Round Wall Clock   -52.7472
1    Central  3.6 Cubic Foot Counter Height Office Refrigerator -1378.8216
2    Central                             36X48 HARDFLOOR CHAIRMAT   -55.5970
3    Central   APC 7 Outlet Network SurgeArrest Surge Protector  -241.4400
4    Central  Acco 6 Outlet Guardian Premium Plus Surge Supp...  -135.0184
..       ...                                              ...        ...
287     West         Wilson Jones Custom Binder Spines & Labels    -5.9840
288     West            Wilson Jones Easy Flow II Sheet Lifters     3.9600
289     West                        Wilson Jones Impact Binders    17.3530
290     West  Wilson Jones Leather-Like Binders with DublLoc...    48.1896
291     West                netTALK DUO VoIP Telephone Service   314.9400

[292 rows x 3 columns]
```

```
# Plot the Top 5 loss-making products per region

fig, axes = plt.subplots(len(regions), 1, figsize = (10, 4*len(regions)))

axes = axes if len(regions)>1 else [axes]

for ax, region in zip(axes, regions):
  regional_data = loss_summary[loss_summary['Region'] == region].sort_values('Profit').head()
  ax.barh(regional_data['Product Name'], regional_data['Profit'], color = 'Red')
  ax.set_title(f'Top 5 Loss-Making Products in {region}')
  ax.invert_yaxis()

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

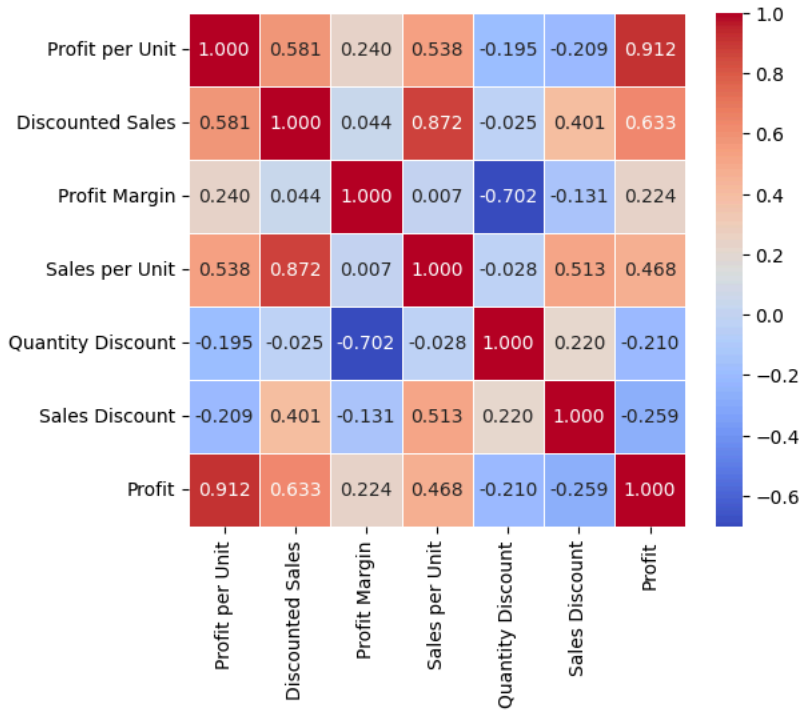
Show hidden output

```
# Plotting Overall Top 5 Loss-Making Products across the country

loss_overall_summary = loss_products.sort_values('Profit').head()

plt.figure(figsize=(5,6))
plt.barh(loss_overall_summary.index.get_level_values('Region')+ '-'+ loss_overall_summary.index.get_level_values('Product Name'
plt.xlabel('Total Profit')
plt.title('Overall Top 5 Loss-Making Products')
plt.gca().invert_yaxis()
plt.show()
```




```
# Feature engineering

df['Profit Margin'] = df['Profit']/df['Sales'] #Profit Margin a better indicator for profit
df['Discounted Sales'] = df['Sales']*(1-df['Discount'])
df['Profit per Unit'] = df['Profit'] / df['Quantity']
df['Sales per Unit'] = df['Sales']/df['Quantity']
df['Sales Discount'] = df['Sales']* df['Discount']
df['Quantity Discount'] = df['Quantity']* df['Discount']

c = df[['Profit per Unit', 'Discounted Sales', 'Profit Margin','Sales per Unit','Quantity Discount', 'Sales Discount', 'Profit'
print(c.corr())

plt.figure(figsize=(6,5))
sns.heatmap(c.corr(), annot = True, cmap = 'coolwarm', fmt='.3f', linewidths=0.5, cbar=True )
plt.show()
```

```
                  Profit per Unit  Discounted Sales  Profit Margin  \
Profit per Unit          1.000000          0.580837       0.240215
Discounted Sales         0.580837          1.000000       0.043816
Profit Margin            0.240215          0.043816       1.000000
Sales per Unit           0.538217          0.872183       0.007044
Quantity Discount       -0.194550         -0.024842      -0.702435
Sales Discount          -0.209161          0.401281      -0.130908
Profit                   0.912199          0.632732       0.223732

                  Sales per Unit  Quantity Discount  Sales Discount    Profit
Profit per Unit         0.538217          -0.194550       -0.209161  0.912199
Discounted Sales        0.872183          -0.024842        0.401281  0.632732
Profit Margin           0.007044          -0.702435       -0.130908  0.223732
Sales per Unit          1.000000          -0.028481        0.513050  0.468312
Quantity Discount      -0.028481           1.000000        0.220456 -0.210114
Sales Discount          0.513050           0.220456        1.000000 -0.259087
Profit                  0.468312          -0.210114       -0.259087  1.000000
```

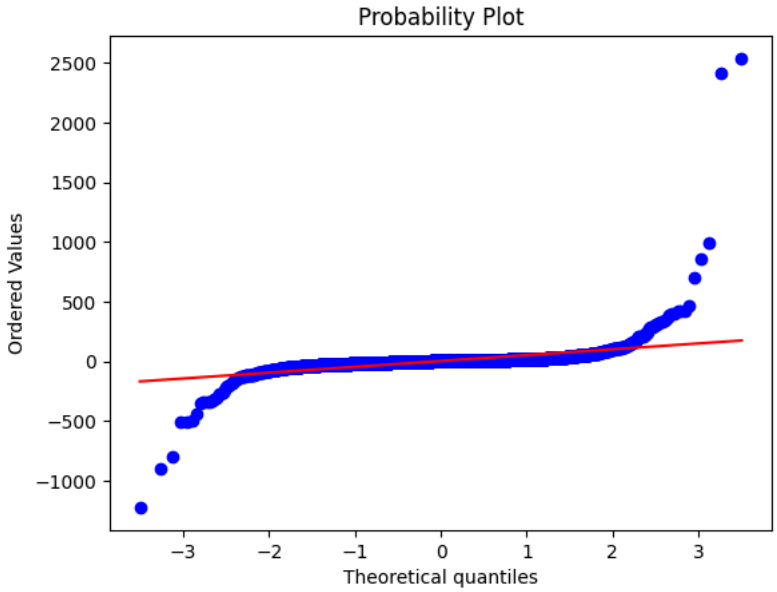



```
# Supervised ml regression algorithm
X1 = df[['Discount', 'Profit Margin','Profit per Unit', 'Discounted Sales', 'Sales per Unit', 'Region', 'Category', 'Sub-Categc
y = df['Profit']

# Train and Test the dataset
X_train, X_test, y_train, y_test = train_test_split(X1, y, test_size = 0.3, random_state= 42)

# Preprocessing
numeric_features = ['Discount', 'Profit per Unit', 'Discounted Sales', 'Sales per Unit', 'Profit Margin', 'Quantity Discount',
categorical_features = ['Region', 'Category', 'Sub-Category']

# Scale numbers
numeric_transformer = Pipeline(steps=[('scaler', StandardScaler())])

# Encode categories using Onehotencoder
categorical_transformer = Pipeline(steps=[('encoder', OneHotEncoder(handle_unknown='ignore'))])

preprocessor = ColumnTransformer(transformers=[('num', numeric_transformer, numeric_features), ('cat', categorical_transformer,
```

```
# Linear Regression Model
model1 = Pipeline(steps=[('preprocessor', preprocessor), ('regressor', LinearRegression())])
#
```

```
R^2: 0.8443175271089369
MAE: 26.413264405112542
RMSE: 91.27656275310502
```

```
residuals = y_test - y_pred1
import scipy.stats as stats
stats.probplot(residuals, dist="norm", plot=plt)
plt.show()
print("MAE:", mean_absolute_error(y_test, y_pred1))
```




```
# Random Forest Regression Model

model2 = Pipeline(steps =[('preprocessor', preprocessor), ('regressor', RandomForestRegressor(random_state = 42))] )
param_grid = {'regressor__max_depth': range(1,21)}
cv = KFold(n_splits=5, shuffle = True, random_state= 42)

grid = GridSearchCV(model2, param_grid, cv = cv, scoring= 'neg_mean_squared_error', n_jobs = -1)

grid.fit(X_train, y_train)
print("Best max_depth:", grid.best_params_['regressor__max_depth'])
```

```
Best max_depth: 16
```

```
best_max_depth = grid.best_params_['regressor__max_depth']
model2 = Pipeline(steps =[('preprocessor', preprocessor), ('regressor', RandomForestRegressor(random_state = 42, max_depth=best_
# Train model
model2.fit(X_train, y_train)
# Predict
y_pred2 = model2.predict(X_test)

# Evaluation
print("R^2:", r2_score(y_test, y_pred2))
print("MAE:", mean_absolute_error(y_test, y_pred2))
print("RMSE:", np.sqrt(mean_squared_error(y_test, y_pred2)))
```

```
R^2: 0.761671566218642
MAE: 9.714458282032473
RMSE: 112.93443783102603
```

```
residuals = y_test -  y_pred2
stats.probplot(residuals, dist="norm", plot=plt)
plt.show()
```