```
import pandas as pd

pip install kaggle
```

```
Requirement already satisfied: kaggle in c:\users\venky\anaconda3\lib\
site-packages (1.6.14)
Requirement already satisfied: six>=1.10 in c:\users\venky\anaconda3\
lib\site-packages (from kaggle) (1.16.0)
Requirement already satisfied: certifi>=2023.7.22 in c:\users\venky\
anaconda3\lib\site-packages (from kaggle) (2024.2.2)
Requirement already satisfied: python-dateutil in c:\users\venky\
anaconda3\lib\site-packages (from kaggle) (2.8.2)
Requirement already satisfied: requests in c:\users\venky\anaconda3\
lib\site-packages (from kaggle) (2.31.0)
Requirement already satisfied: tqdm in c:\users\venky\anaconda3\lib\
site-packages (from kaggle) (4.65.0)
Requirement already satisfied: python-slugify in c:\users\venky\
anaconda3\lib\site-packages (from kaggle) (5.0.2)
Requirement already satisfied: urllib3 in c:\users\venky\anaconda3\
lib\site-packages (from kaggle) (2.0.7)
Requirement already satisfied: bleach in c:\users\venky\anaconda3\lib\
site-packages (from kaggle) (4.1.0)
Requirement already satisfied: packaging in c:\users\venky\anaconda3\
lib\site-packages (from bleach->kaggle) (23.1)
Requirement already satisfied: webencodings in c:\users\venky\
anaconda3\lib\site-packages (from bleach->kaggle) (0.5.1)
Requirement already satisfied: text-unidecode>=1.3 in c:\users\venky\
anaconda3\lib\site-packages (from python-slugify->kaggle) (1.3)
Requirement already satisfied: charset-normalizer<4,>=2 in c:\users\
venky\anaconda3\lib\site-packages (from requests->kaggle) (2.0.4)
Requirement already satisfied: idna<4,>=2.5 in c:\users\venky\
anaconda3\lib\site-packages (from requests->kaggle) (3.4)
Requirement already satisfied: colorama in c:\users\venky\anaconda3\
lib\site-packages (from tqdm->kaggle) (0.4.6)
Note: you may need to restart the kernel to use updated packages.
```

Installed Kaggle Library

```
 ! kaggle datasets download -d akshatagrawal9431/shipment-orders-
dataset -f orders_data.csv
```

```
Dataset URL:
https://www.kaggle.com/datasets/akshatagrawal9431/shipment-orders-
dataset
License(s): CC0-1.0
Downloading orders_data.csv.zip to C:\Users\Venky\Meriskill project
```

```
  0%|          | 0.00/197k [00:00<?, ?B/s]
100%|##########| 197k/197k [00:00<00:00, 325kB/s]
100%|##########| 197k/197k [00:00<00:00, 325kB/s]
```

```
import zipfile
zip_ref = zipfile.ZipFile('orders_data.csv.zip')
zip_ref.extractall()
zip_ref.close()
```

Unzip the zip file downloaded

```
df=pd.read_csv('orders_data.csv')
df.head()
```

```
   Order Id  Order Date       Ship Mode    Segment        Country  \
0         1  01-03-2023    Second Class   Consumer  United States
1         2  15-08-2023    Second Class   Consumer  United States
2         3  10-01-2023    Second Class  Corporate  United States
3         4  18-06-2022  Standard Class   Consumer  United States
4         5  13-07-2022  Standard Class   Consumer  United States

              City       State  Postal Code Region         Category  \
0        Henderson    Kentucky        42420  South        Furniture
1        Henderson    Kentucky        42420  South        Furniture
2      Los Angeles  California        90036   West  Office Supplies
3  Fort Lauderdale     Florida        33311  South        Furniture
4  Fort Lauderdale     Florida        33311  South  Office Supplies

  Sub Category       Product Id  cost price  List Price  Quantity  \
0    Bookcases  FUR-BO-10001798         240         260         2
1       Chairs  FUR-CH-10000454         600         730         3
2       Labels  OFF-LA-10000240          10          10         2
3       Tables  FUR-TA-10000577         780         960         5
4      Storage  OFF-ST-10000760          20          20         2

   Discount Percent
0                 2
1                 3
2                 5
3                 2
4                 5
```

Reading the csv file ater unzipping from zip file

```
df.shape
```

```
(9994, 16)
```

```
df.describe()
```

```
          Order Id   Postal Code    cost price    List Price
Quantity  \
count  9994.000000   9994.000000   9994.000000   9994.000000
9994.000000
mean   4997.500000  55190.379428    201.189714    229.756854
3.789574
std    2885.163629  32063.693350    537.743203    623.245839
2.225110
min       1.000000   1040.000000      0.000000      0.000000
1.000000
25%    2499.250000  23223.000000     20.000000     20.000000
2.000000
50%    4997.500000  56430.500000     50.000000     50.000000
3.000000
75%    7495.750000  90008.000000    180.000000    210.000000
5.000000
max    9994.000000  99301.000000  18110.000000  22640.000000
14.000000

       Discount Percent
count       9994.000000
mean           3.484090
std            1.114211
min            2.000000
25%            2.000000
50%            3.000000
75%            4.000000
max            5.000000
```

Information of data frame

```
df.info
```

```
<bound method DataFrame.info of       Order Id  Order Date       Ship
Mode    Segment        Country  \
0            1  01-03-2023    Second Class   Consumer  United States

1            2  15-08-2023    Second Class   Consumer  United States

2            3  10-01-2023    Second Class  Corporate  United States

3            4  18-06-2022  Standard Class   Consumer  United States

4            5  13-07-2022  Standard Class   Consumer  United States

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

9989      9990  18-02-2023    Second Class   Consumer  United States
```

```
9990      9991  17-03-2023  Standard Class   Consumer  United States
9991      9992  07-08-2022  Standard Class   Consumer  United States
9992      9993  19-11-2022  Standard Class   Consumer  United States
9993      9994  17-07-2022    Second Class   Consumer  United States

                  City       State  Postal Code Region         Category
\
0            Henderson    Kentucky        42420  South        Furniture
1            Henderson    Kentucky        42420  South        Furniture
2          Los Angeles  California        90036   West  Office Supplies
3     Fort Lauderdale     Florida        33311  South        Furniture
4     Fort Lauderdale     Florida        33311  South  Office Supplies
...               ...         ...          ...    ...              ...
9989            Miami     Florida        33180  South        Furniture
9990       Costa Mesa  California        92627   West        Furniture
9991       Costa Mesa  California        92627   West       Technology
9992       Costa Mesa  California        92627   West  Office Supplies
9993      Westminster  California        92683   West  Office Supplies

     Sub Category       Product Id  cost price  List Price
Quantity  \
0       Bookcases  FUR-BO-10001798         240         260         2
1          Chairs  FUR-CH-10000454         600         730         3
2          Labels  OFF-LA-10000240          10          10         2
3          Tables  FUR-TA-10000577         780         960         5
4         Storage  OFF-ST-10000760          20          20         2
...           ...              ...         ...         ...       ...
9989  Furnishings  FUR-FU-10001889          30          30         3
9990  Furnishings  FUR-FU-10000747          70          90         2
```

```
9991       Phones  TEC-PH-10003645         220          260          2

9992        Paper  OFF-PA-10004041          30           30          4

9993   Appliances  OFF-AP-10002684         210          240          2


      Discount Percent
0                    2
1                    3
2                    5
3                    2
4                    5
...                ...
9989                 4
9990                 4
9991                 2
9992                 3
9993                 3

[9994 rows x 16 columns]>
```

```
df['Ship Mode'].unique()
```

```
array(['Second Class', 'Standard Class', 'Not Available', 'unknown',
       'First Class', nan, 'Same Day'], dtype=object)
```

```
df1 = pd.read_csv('orders_data.csv', na_values=['Not Available',
'unknown'])
df1['Ship Mode'].unique()
```

```
array(['Second Class', 'Standard Class', nan, 'First Class', 'Same
Day'],
      dtype=object)
```

```
df1.head()
```

```
   Order Id  Order Date       Ship Mode    Segment        Country  \
0         1  01-03-2023    Second Class   Consumer  United States
1         2  15-08-2023    Second Class   Consumer  United States
2         3  10-01-2023    Second Class  Corporate  United States
3         4  18-06-2022  Standard Class   Consumer  United States
4         5  13-07-2022  Standard Class   Consumer  United States

              City       State  Postal Code Region         Category  \
0        Henderson    Kentucky        42420  South        Furniture
1        Henderson    Kentucky        42420  South        Furniture
2      Los Angeles  California        90036   West  Office Supplies
3  Fort Lauderdale     Florida        33311  South        Furniture
4  Fort Lauderdale     Florida        33311  South  Office Supplies
```

```
  Sub Category       Product Id  cost price  List Price  Quantity  \
0    Bookcases  FUR-BO-10001798         240         260         2
1       Chairs  FUR-CH-10000454         600         730         3
2       Labels  OFF-LA-10000240          10          10         2
3       Tables  FUR-TA-10000577         780         960         5
4      Storage  OFF-ST-10000760          20          20         2

   Discount Percent
0                 2
1                 3
2                 5
3                 2
4                 5
```

Get new columns discount, sale price and profit

```
df1.head(2)
```

```
   Order Id  Order Date     Ship Mode   Segment        Country
City  \
0         1  01-03-2023  Second Class  Consumer  United States
Henderson
1         2  15-08-2023  Second Class  Consumer  United States
Henderson

      State  Postal Code Region   Category Sub Category       Product
Id  \
0  Kentucky        42420  South  Furniture    Bookcases  FUR-BO-
10001798
1  Kentucky        42420  South  Furniture       Chairs  FUR-CH-
10000454

   cost price  List Price  Quantity  Discount Percent
0         240         260         2                 2
1         600         730         3                 3
```

```
df1['Selling_Price'] = df1['List Price'] - df1['List
Price']*(df1['Discount Percent']/100)
df1['Selling_Price']
```

```
0       254.8
1       708.1
2         9.5
3       940.8
4        19.0
        ...
9989     28.8
9990     86.4
9991    254.8
```

```
9992     29.1
9993    232.8
Name: Selling_Price, Length: 9994, dtype: float64

df1

      Order Id  Order Date       Ship Mode    Segment
Country  \
0            1  01-03-2023    Second Class   Consumer  United States

1            2  15-08-2023    Second Class   Consumer  United States

2            3  10-01-2023    Second Class  Corporate  United States

3            4  18-06-2022  Standard Class   Consumer  United States

4            5  13-07-2022  Standard Class   Consumer  United States

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

9989      9990  18-02-2023    Second Class   Consumer  United States

9990      9991  17-03-2023  Standard Class   Consumer  United States

9991      9992  07-08-2022  Standard Class   Consumer  United States

9992      9993  19-11-2022  Standard Class   Consumer  United States

9993      9994  17-07-2022    Second Class   Consumer  United States


                  City       State  Postal Code Region         Category
\
0           Henderson    Kentucky        42420  South        Furniture

1           Henderson    Kentucky        42420  South        Furniture

2         Los Angeles  California        90036   West  Office Supplies

3     Fort Lauderdale     Florida        33311  South        Furniture

4     Fort Lauderdale     Florida        33311  South  Office Supplies

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

9989            Miami     Florida        33180  South        Furniture

9990       Costa Mesa  California        92627   West        Furniture

9991       Costa Mesa  California        92627   West       Technology

9992       Costa Mesa  California        92627   West  Office Supplies
```

```
9993      Westminster  California        92683   West  Office Supplies


     Sub Category       Product Id  cost price  List Price
Quantity  \
0       Bookcases  FUR-BO-10001798         240         260          2

1         Chairs  FUR-CH-10000454         600         730          3

2         Labels  OFF-LA-10000240          10          10          2

3         Tables  FUR-TA-10000577         780         960          5

4        Storage  OFF-ST-10000760          20          20          2

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

9989  Furnishings  FUR-FU-10001889          30          30          3

9990  Furnishings  FUR-FU-10000747          70          90          2

9991       Phones  TEC-PH-10003645         220         260          2

9992        Paper  OFF-PA-10004041          30          30          4

9993   Appliances  OFF-AP-10002684         210         240          2


      Discount Percent  Selling_Price
0                    2          254.8
1                    3          708.1
2                    5            9.5
3                    2          940.8
4                    5           19.0
...                ...            ...
9989                 4           28.8
9990                 4           86.4
9991                 2          254.8
9992                 3           29.1
9993                 3          232.8

[9994 rows x 17 columns]

df1['Profit']=(df1['Selling_Price'] - df1['cost price'])
df1['Profit']

0        14.8
1       108.1
2        -0.5
3       160.8
```

```
4        -1.0
        ...
9989     -1.2
9990     16.4
9991     34.8
9992     -0.9
9993     22.8
Name: Profit, Length: 9994, dtype: float64
```

```
df1.head(3)
```

```
   Order Id  Order Date     Ship Mode    Segment        Country
City  \
0         1  01-03-2023  Second Class   Consumer  United States
Henderson
1         2  15-08-2023  Second Class   Consumer  United States
Henderson
2         3  10-01-2023  Second Class  Corporate  United States  Los
Angeles

        State  Postal Code Region         Category Sub Category  \
0    Kentucky        42420  South        Furniture    Bookcases
1    Kentucky        42420  South        Furniture       Chairs
2  California        90036   West  Office Supplies       Labels

        Product Id  cost price  List Price  Quantity  Discount Percent
\
0  FUR-BO-10001798         240         260         2                 2

1  FUR-CH-10000454         600         730         3                 3

2  OFF-LA-10000240          10          10         2                 5


   Selling_Price  Profit
0          254.8    14.8
1          708.1   108.1
2            9.5    -0.5
```

```
df1.rename(columns={'Profit': 'Unit_Profit', 'Selling_Price':
'Unit_Selling_Price'}, inplace=True )
df1.head(2)
```

```
   Order Id  Order Date     Ship Mode   Segment        Country
City  \
0         1  01-03-2023  Second Class  Consumer  United States
Henderson
1         2  15-08-2023  Second Class  Consumer  United States
Henderson

      State  Postal Code Region   Category Sub Category       Product
```

```
Id  \
0  Kentucky        42420  South  Furniture    Bookcases  FUR-BO-
10001798
1  Kentucky        42420  South  Furniture       Chairs  FUR-CH-
10000454

   cost price  List Price  Quantity  Discount Percent
Unit_Selling_Price  \
0         240         260         2                 2
254.8
1         600         730         3                 3
708.1

   Unit_Profit
0         14.8
1        108.1
```

```
df1.dtypes
```

```
Order Id                int64
Order Date             object
Ship Mode              object
Segment                object
Country                object
City                   object
State                  object
Postal Code             int64
Region                 object
Category               object
Sub Category           object
Product Id             object
cost price              int64
List Price              int64
Quantity                int64
Discount Percent        int64
Unit_Selling_Price    float64
Unit_Profit           float64
dtype: object
```

Changing datatypes of order date column from object to datetime

```
df1.head(3)
```

```
   Order Id  Order Date     Ship Mode    Segment        Country
City  \
0         1  01-03-2023  Second Class   Consumer  United States
Henderson
1         2  15-08-2023  Second Class   Consumer  United States
Henderson
2         3  10-01-2023  Second Class  Corporate  United States  Los
```

```
Angeles
```

```
        State  Postal Code Region         Category Sub Category  \
0    Kentucky        42420  South        Furniture    Bookcases
1    Kentucky        42420  South        Furniture       Chairs
2  California        90036   West  Office Supplies       Labels

        Product Id  cost price  List Price  Quantity  Discount Percent
\
0  FUR-BO-10001798         240         260         2                 2

1  FUR-CH-10000454         600         730         3                 3

2  OFF-LA-10000240          10          10         2                 5


   Unit_Selling_Price  Unit_Profit
0               254.8         14.8
1               708.1        108.1
2                 9.5         -0.5
```

```
df1.dtypes
```

```
Order Id                int64
Order Date             object
Ship Mode              object
Segment                object
Country                object
City                   object
State                  object
Postal Code             int64
Region                 object
Category               object
Sub Category           object
Product Id             object
cost price              int64
List Price              int64
Quantity                int64
Discount Percent        int64
Unit_Selling_Price    float64
Unit_Profit           float64
dtype: object
```

```
df1['Order Date'] = pd.to_datetime(df1['Order Date'], format= '%d-%m-
%Y')
```

```
df1.dtypes
```

```
Order Id                       int64
Order Date            datetime64[ns]
Ship Mode                     object
```

```
Segment                        object
Country                        object
City                           object
State                          object
Postal Code                     int64
Region                         object
Category                       object
Sub Category                   object
Product Id                     object
cost price                      int64
List Price                      int64
Quantity                        int64
Discount Percent                int64
Unit_Selling_Price            float64
Unit_Profit                   float64
dtype: object
```

```
df1.head()
```

```
   Order Id Order Date       Ship Mode    Segment        Country  \
0         1 2023-03-01    Second Class   Consumer  United States
1         2 2023-08-15    Second Class   Consumer  United States
2         3 2023-01-10    Second Class  Corporate  United States
3         4 2022-06-18  Standard Class   Consumer  United States
4         5 2022-07-13  Standard Class   Consumer  United States

              City       State  Postal Code Region         Category  \
0        Henderson    Kentucky        42420  South        Furniture
1        Henderson    Kentucky        42420  South        Furniture
2      Los Angeles  California        90036   West  Office Supplies
3  Fort Lauderdale     Florida        33311  South        Furniture
4  Fort Lauderdale     Florida        33311  South  Office Supplies

  Sub Category       Product Id  cost price  List Price  Quantity  \
0    Bookcases  FUR-BO-10001798         240         260         2
1       Chairs  FUR-CH-10000454         600         730         3
2       Labels  OFF-LA-10000240          10          10         2
3       Tables  FUR-TA-10000577         780         960         5
4      Storage  OFF-ST-10000760          20          20         2

   Discount Percent  Unit_Selling_Price  Unit_Profit
0                 2               254.8         14.8
1                 3               708.1        108.1
2                 5                 9.5         -0.5
3                 2               940.8        160.8
4                 5                19.0         -1.0
```

```
df1.head(5)
```

```
   Order Id Order Date       Ship Mode    Segment        Country  \
0         1 2023-03-01    Second Class   Consumer  United States
1         2 2023-08-15    Second Class   Consumer  United States
2         3 2023-01-10    Second Class  Corporate  United States
3         4 2022-06-18  Standard Class   Consumer  United States
4         5 2022-07-13  Standard Class   Consumer  United States

               City       State  Postal Code Region         Category  \
0        Henderson    Kentucky        42420  South        Furniture
1        Henderson    Kentucky        42420  South        Furniture
2      Los Angeles  California        90036   West  Office Supplies
3  Fort Lauderdale     Florida        33311  South        Furniture
4  Fort Lauderdale     Florida        33311  South  Office Supplies

  Sub Category       Product Id  Quantity  Unit_Selling_Price
Unit_Profit
0    Bookcases  FUR-BO-10001798         2               254.8
14.8
1       Chairs  FUR-CH-10000454         3               708.1
108.1
2       Labels  OFF-LA-10000240         2                 9.5
-0.5
3       Tables  FUR-TA-10000577         5               940.8
160.8
4      Storage  OFF-ST-10000760         2                19.0
-1.0
```

```
#Caluclating total profit in each Order
df1['Total Profit'] = df1['Quantity']* df1 ['Unit_Profit']

df1.head()
```

```
   Order Id Order Date       Ship Mode    Segment        Country  \
0         1 2023-03-01    Second Class   Consumer  United States
1         2 2023-08-15    Second Class   Consumer  United States
2         3 2023-01-10    Second Class  Corporate  United States
3         4 2022-06-18  Standard Class   Consumer  United States
4         5 2022-07-13  Standard Class   Consumer  United States

               City       State  Postal Code Region         Category  \
0        Henderson    Kentucky        42420  South        Furniture
1        Henderson    Kentucky        42420  South        Furniture
2      Los Angeles  California        90036   West  Office Supplies
3  Fort Lauderdale     Florida        33311  South        Furniture
4  Fort Lauderdale     Florida        33311  South  Office Supplies

  Sub Category       Product Id  Quantity  Unit_Selling_Price
Unit_Profit  \
0    Bookcases  FUR-BO-10001798         2               254.8
14.8
```

```
1       Chairs  FUR-CH-10000454         3               708.1
108.1
2       Labels  OFF-LA-10000240         2                 9.5
-0.5
3       Tables  FUR-TA-10000577         5               940.8
160.8
4      Storage  OFF-ST-10000760         2                19.0
-1.0

   Total Profit
0          29.6
1         324.3
2          -1.0
3         804.0
4          -2.0
```

```
df.head(5)
```

```
   Order Id  Order Date       Ship Mode    Segment        Country  \
0         1  01-03-2023    Second Class   Consumer  United States
1         2  15-08-2023    Second Class   Consumer  United States
2         3  10-01-2023    Second Class  Corporate  United States
3         4  18-06-2022  Standard Class   Consumer  United States
4         5  13-07-2022  Standard Class   Consumer  United States

               City       State  Postal Code Region         Category  \
0        Henderson    Kentucky        42420  South        Furniture
1        Henderson    Kentucky        42420  South        Furniture
2      Los Angeles  California        90036   West  Office Supplies
3  Fort Lauderdale     Florida        33311  South        Furniture
4  Fort Lauderdale     Florida        33311  South  Office Supplies

  Sub Category       Product Id  cost price  List Price  Quantity  \
0    Bookcases  FUR-BO-10001798         240         260         2
1       Chairs  FUR-CH-10000454         600         730         3
2       Labels  OFF-LA-10000240          10          10         2
3       Tables  FUR-TA-10000577         780         960         5
4      Storage  OFF-ST-10000760          20          20         2

   Discount Percent
0                 2
1                 3
2                 5
3                 2
4                 5
```

Questions starts here

```
df1.head()
```

```
   Order Id Order Date       Ship Mode    Segment        Country  \
0         1 2023-03-01    Second Class   Consumer  United States
1         2 2023-08-15    Second Class   Consumer  United States
2         3 2023-01-10    Second Class  Corporate  United States
3         4 2022-06-18  Standard Class   Consumer  United States
4         5 2022-07-13  Standard Class   Consumer  United States

              City       State  Postal Code Region         Category  \
0        Henderson    Kentucky        42420  South        Furniture
1        Henderson    Kentucky        42420  South        Furniture
2      Los Angeles  California        90036   West  Office Supplies
3  Fort Lauderdale     Florida        33311  South        Furniture
4  Fort Lauderdale     Florida        33311  South  Office Supplies

  Sub Category       Product Id  Quantity  Unit_Selling_Price
Unit_Profit  \
0    Bookcases  FUR-BO-10001798         2               254.8
14.8
1       Chairs  FUR-CH-10000454         3               708.1
108.1
2       Labels  OFF-LA-10000240         2                 9.5
-0.5
3       Tables  FUR-TA-10000577         5               940.8
160.8
4      Storage  OFF-ST-10000760         2                19.0
-1.0

   Total Profit
0          29.6
1         324.3
2          -1.0
3         804.0
4          -2.0
```

Questions :

Q1. Find top 10 highest profit generating products

```
df1.groupby('Product Id')
```

```
<pandas.core.groupby.generic.DataFrameGroupBy object at
0x000002A63BE4D950>
```

```
df1.groupby('Product Id').sum(['Total Profit'])
```

```
                  Order Id  Postal Code  Quantity
Unit_Selling_Price  \
Product Id

FUR-BO-10000112      3513        60653         9                805.1

FUR-BO-10000330     13658       256733        10               1036.4

FUR-BO-10000362     18886       184360        14               2089.4

FUR-BO-10000468     35749       357215        21                706.1

FUR-BO-10000711      9310       120635        12                830.9

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

TEC-PH-10004912     21739       166396        11                493.8

TEC-PH-10004922     31470       296048        19               1037.3

TEC-PH-10004924     13937       105326         8                 48.2

TEC-PH-10004959      5191        83105         4                328.8

TEC-PH-10004977     49957       473928        32               5099.0


                  Unit_Profit  Total Profit
Product Id
FUR-BO-10000112         45.1         405.9
FUR-BO-10000330        146.4         560.0
FUR-BO-10000362        189.4         638.6
FUR-BO-10000468         86.1         380.6
FUR-BO-10000711         40.9         345.9
...                      ...           ...
TEC-PH-10004912         53.8         203.8
TEC-PH-10004922         57.3         259.2
TEC-PH-10004924         -1.8          -6.0
TEC-PH-10004959         28.8          74.4
TEC-PH-10004977        399.0        1593.2

[1862 rows x 6 columns]

df1.groupby('Product Id').sum(['Total Profit']).sort_values(by =
['Total Profit'], ascending=False)

                  Order Id  Postal Code  Quantity
Unit_Selling_Price  \
Product Id

TEC-CO-10004722     28222       194875        20              59514.0
```

```
TEC-MA-10002412      2698        32216         6              21734.4

OFF-BI-10000545     57937       302599        48              18249.0

TEC-CO-10001449     41444       428939        38              18151.2

FUR-CH-10002024     44560       448062        39              21096.2

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

OFF-BI-10003712     46354       562879        45                124.7

OFF-AR-10003631     20973       277819        30                106.4

OFF-PA-10000143     38761       389204        29                153.8

OFF-BI-10002799     32515       368281        40                114.2

OFF-FA-10002280     54151       608473        47                251.8
```

```
                 Unit_Profit  Total Profit
Product Id
TEC-CO-10004722       5644.0       24816.0
TEC-MA-10002412       3624.4       21746.4
OFF-BI-10000545       1959.0       17867.7
TEC-CO-10001449       2631.2       15948.0
FUR-CH-10002024       2246.2       13930.7
...                      ...          ...
OFF-BI-10003712         -5.3         -35.7
OFF-AR-10003631         -3.6         -37.0
OFF-PA-10000143         -6.2         -39.5
OFF-BI-10002799         -5.8         -44.5
OFF-FA-10002280         -8.2         -48.2

[1862 rows x 6 columns]
```

```
df1.groupby('Product Id').sum(['Total Profit']).sort_values(by =
['Total Profit'], ascending=False).reset_index()[['Product Id','Total
Profit']].head(10)
```

```
        Product Id  Total Profit
0  TEC-CO-10004722       24816.0
1  TEC-MA-10002412       21746.4
2  OFF-BI-10000545       17867.7
3  TEC-CO-10001449       15948.0
4  FUR-CH-10002024       13930.7
5  OFF-BI-10003527       12792.9
6  TEC-PH-10001459       11481.9
7  TEC-MA-10000822       10102.3
```

```
8  FUR-TA-10000198       10015.1
9  TEC-MA-10001047        9989.0
```

```
df1.head(4)
```

```
   Order Id Order Date       Ship Mode    Segment        Country  \
0         1 2023-03-01    Second Class   Consumer  United States
1         2 2023-08-15    Second Class   Consumer  United States
2         3 2023-01-10    Second Class  Corporate  United States
3         4 2022-06-18  Standard Class   Consumer  United States

              City       State  Postal Code Region         Category  \
0        Henderson    Kentucky        42420  South        Furniture
1        Henderson    Kentucky        42420  South        Furniture
2      Los Angeles  California        90036   West  Office Supplies
3  Fort Lauderdale     Florida        33311  South        Furniture

  Sub Category       Product Id  Quantity  Unit_Selling_Price
Unit_Profit  \
0    Bookcases  FUR-BO-10001798         2               254.8
14.8
1       Chairs  FUR-CH-10000454         3               708.1
108.1
2       Labels  OFF-LA-10000240         2                 9.5
-0.5
3       Tables  FUR-TA-10000577         5               940.8
160.8

   Total Profit
0          29.6
1         324.3
2          -1.0
3         804.0
```

Q2. Create a column for Total sales Value of each other

```
df1['Quantity']
```

```
0       2
1       3
2       2
3       5
4       2
       ..
9989    3
9990    2
9991    2
9992    4
9993    2
Name: Quantity, Length: 9994, dtype: int64
```

```
df1['Unit_Selling_Price']
```

```
0       254.8
1       708.1
2         9.5
3       940.8
4        19.0
        ...
9989     28.8
9990     86.4
9991    254.8
9992     29.1
9993    232.8
Name: Unit_Selling_Price, Length: 9994, dtype: float64
```

```
df1['Total_sale'] = df1['Unit_Selling_Price']*df1['Quantity']
df1.head(4)
```

```
   Order Id Order Date       Ship Mode    Segment        Country  \
0         1 2023-03-01    Second Class   Consumer  United States
1         2 2023-08-15    Second Class   Consumer  United States
2         3 2023-01-10    Second Class  Corporate  United States
3         4 2022-06-18  Standard Class   Consumer  United States

              City       State  Postal Code Region         Category  \
0        Henderson    Kentucky        42420  South        Furniture
1        Henderson    Kentucky        42420  South        Furniture
2      Los Angeles  California        90036   West  Office Supplies
3  Fort Lauderdale     Florida        33311  South        Furniture

  Sub Category       Product Id  Quantity  Unit_Selling_Price
Unit_Profit  \
0    Bookcases  FUR-BO-10001798         2               254.8
14.8
1       Chairs  FUR-CH-10000454         3               708.1
108.1
2       Labels  OFF-LA-10000240         2                 9.5
-0.5
3       Tables  FUR-TA-10000577         5               940.8
160.8

   Total Profit  Total_sale
0          29.6       509.6
1         324.3      2124.3
2          -1.0        19.0
3         804.0      4704.0
```

Q3. Write a Python Code to list all distinct cities where orders have been shipped

```
df1.head(2)
```

```
   Order Id Order Date     Ship Mode   Segment        Country
City  \
0         1 2023-03-01  Second Class  Consumer  United States
Henderson
1         2 2023-08-15  Second Class  Consumer  United States
Henderson

      State  Postal Code Region   Category Sub Category       Product
Id  \
0  Kentucky        42420  South  Furniture    Bookcases  FUR-BO-
10001798
1  Kentucky        42420  South  Furniture       Chairs  FUR-CH-
10000454

   Quantity  Unit_Selling_Price  Unit_Profit  Total Profit  Total_sale

0         2               254.8         14.8          29.6       509.6

1         3               708.1        108.1         324.3      2124.3
```

```
df['City'].unique()
#df['City'].nunique()
```

```
array(['Henderson', 'Los Angeles', 'Fort Lauderdale', 'Concord',
       'Seattle', 'Fort Worth', 'Madison', 'West Jordan', 'San
Francisco',
       'Fremont', 'Philadelphia', 'Orem', 'Houston', 'Richardson',
       'Naperville', 'Melbourne', 'Eagan', 'Westland', 'Dover',
       'New Albany', 'New York City', 'Troy', 'Chicago', 'Gilbert',
       'Springfield', 'Jackson', 'Memphis', 'Decatur', 'Durham',
       'Columbia', 'Rochester', 'Minneapolis', 'Portland', 'Saint
Paul',
       'Aurora', 'Charlotte', 'Orland Park', 'Urbandale', 'Columbus',
       'Bristol', 'Wilmington', 'Bloomington', 'Phoenix', 'Roseville',
       'Independence', 'Pasadena', 'Newark', 'Franklin', 'Scottsdale',
       'San Jose', 'Edmond', 'Carlsbad', 'San Antonio', 'Monroe',
       'Fairfield', 'Grand Prairie', 'Redlands', 'Hamilton',
'Westfield',
       'Akron', 'Denver', 'Dallas', 'Whittier', 'Saginaw', 'Medina',
       'Dublin', 'Detroit', 'Tampa', 'Santa Clara', 'Lakeville',
       'San Diego', 'Brentwood', 'Chapel Hill', 'Morristown',
       'Cincinnati', 'Inglewood', 'Tamarac', 'Colorado Springs',
       'Belleville', 'Taylor', 'Lakewood', 'Arlington', 'Arvada',
       'Hackensack', 'Saint Petersburg', 'Long Beach', 'Hesperia',
       'Murfreesboro', 'Layton', 'Austin', 'Lowell', 'Manchester',
```

```
       'Harlingen', 'Tucson', 'Quincy', 'Pembroke Pines', 'Des
Moines',
       'Peoria', 'Las Vegas', 'Warwick', 'Miami', 'Huntington Beach',
       'Richmond', 'Louisville', 'Lawrence', 'Canton', 'New Rochelle',
       'Gastonia', 'Jacksonville', 'Auburn', 'Norman', 'Park Ridge',
       'Amarillo', 'Lindenhurst', 'Huntsville', 'Fayetteville',
       'Costa Mesa', 'Parker', 'Atlanta', 'Gladstone', 'Great Falls',
       'Lakeland', 'Montgomery', 'Mesa', 'Green Bay', 'Anaheim',
       'Marysville', 'Salem', 'Laredo', 'Grove City', 'Dearborn',
       'Warner Robins', 'Vallejo', 'Mission Viejo', 'Rochester Hills',
       'Plainfield', 'Sierra Vista', 'Vancouver', 'Cleveland',
'Tyler',
       'Burlington', 'Waynesboro', 'Chester', 'Cary', 'Palm Coast',
       'Mount Vernon', 'Hialeah', 'Oceanside', 'Evanston', 'Trenton',
       'Cottage Grove', 'Bossier City', 'Lancaster', 'Asheville',
       'Lake Elsinore', 'Omaha', 'Edmonds', 'Santa Ana', 'Milwaukee',
       'Florence', 'Lorain', 'Linden', 'Salinas', 'New Brunswick',
       'Garland', 'Norwich', 'Alexandria', 'Toledo', 'Farmington',
       'Riverside', 'Torrance', 'Round Rock', 'Boca Raton',
       'Virginia Beach', 'Murrieta', 'Olympia', 'Washington',
       'Jefferson City', 'Saint Peters', 'Rockford', 'Brownsville',
       'Yonkers', 'Oakland', 'Clinton', 'Encinitas', 'Roswell',
       'Jonesboro', 'Antioch', 'Homestead', 'La Porte', 'Lansing',
       'Cuyahoga Falls', 'Reno', 'Harrisonburg', 'Escondido', 'Royal
Oak',
       'Rockville', 'Coral Springs', 'Buffalo', 'Boynton Beach',
       'Gulfport', 'Fresno', 'Greenville', 'Macon', 'Cedar Rapids',
       'Providence', 'Pueblo', 'Deltona', 'Murray', 'Middletown',
       'Freeport', 'Pico Rivera', 'Provo', 'Pleasant Grove', 'Smyrna',
       'Parma', 'Mobile', 'New Bedford', 'Irving', 'Vineland',
'Glendale',
       'Niagara Falls', 'Thomasville', 'Westminster', 'Coppell',
'Pomona',
       'North Las Vegas', 'Allentown', 'Tempe', 'Laguna Niguel',
       'Bridgeton', 'Everett', 'Watertown', 'Appleton', 'Bellevue',
       'Allen', 'El Paso', 'Grapevine', 'Carrollton', 'Kent',
'Lafayette',
       'Tigard', 'Skokie', 'Plano', 'Suffolk', 'Indianapolis',
'Bayonne',
       'Greensboro', 'Baltimore', 'Kenosha', 'Olathe', 'Tulsa',
'Redmond',
       'Raleigh', 'Muskogee', 'Meriden', 'Bowling Green', 'South
Bend',
       'Spokane', 'Keller', 'Port Orange', 'Medford',
'Charlottesville',
       'Missoula', 'Apopka', 'Reading', 'Broomfield', 'Paterson',
       'Oklahoma City', 'Chesapeake', 'Lubbock', 'Johnson City',
       'San Bernardino', 'Leominster', 'Bozeman', 'Perth Amboy',
       'Ontario', 'Rancho Cucamonga', 'Moorhead', 'Mesquite',
```

```
'Stockton',
       'Ormond Beach', 'Sunnyvale', 'York', 'College Station',
       'Saint Louis', 'Manteca', 'San Angelo', 'Salt Lake City',
       'Knoxville', 'Little Rock', 'Lincoln Park', 'Marion',
'Littleton',
       'Bangor', 'Southaven', 'New Castle', 'Midland', 'Sioux Falls',
       'Fort Collins', 'Clarksville', 'Sacramento', 'Thousand Oaks',
       'Malden', 'Holyoke', 'Albuquerque', 'Sparks', 'Coachella',
       'Elmhurst', 'Passaic', 'North Charleston', 'Newport News',
       'Jamestown', 'Mishawaka', 'La Quinta', 'Tallahassee',
'Nashville',
       'Bellingham', 'Woodstock', 'Haltom City', 'Wheeling',
       'Summerville', 'Hot Springs', 'Englewood', 'Las Cruces',
'Hoover',
       'Frisco', 'Vacaville', 'Waukesha', 'Bakersfield', 'Pompano
Beach',
       'Corpus Christi', 'Redondo Beach', 'Orlando', 'Orange',
       'Lake Charles', 'Highland Park', 'Hempstead', 'Noblesville',
       'Apple Valley', 'Mount Pleasant', 'Sterling Heights', 'Eau
Claire',
       'Pharr', 'Billings', 'Gresham', 'Chattanooga', 'Meridian',
       'Bolingbrook', 'Maple Grove', 'Woodland', 'Missouri City',
       'Pearland', 'San Mateo', 'Grand Rapids', 'Visalia',
       'Overland Park', 'Temecula', 'Yucaipa', 'Revere', 'Conroe',
       'Tinley Park', 'Dubuque', 'Dearborn Heights', 'Santa Fe',
       'Hickory', 'Carol Stream', 'Saint Cloud', 'North Miami',
       'Plantation', 'Port Saint Lucie', 'Rock Hill', 'Odessa',
       'West Allis', 'Chula Vista', 'Manhattan', 'Altoona',
'Thornton',
       'Champaign', 'Texarkana', 'Edinburg', 'Baytown', 'Greenwood',
       'Woonsocket', 'Superior', 'Bedford', 'Covington', 'Broken
Arrow',
       'Miramar', 'Hollywood', 'Deer Park', 'Wichita', 'Mcallen',
       'Iowa City', 'Boise', 'Cranston', 'Port Arthur', 'Citrus
Heights',
       'The Colony', 'Daytona Beach', 'Bullhead City', 'Portage',
'Fargo',
       'Elkhart', 'San Gabriel', 'Margate', 'Sandy Springs', 'Mentor',
       'Lawton', 'Hampton', 'Rome', 'La Crosse', 'Lewiston',
       'Hattiesburg', 'Danville', 'Logan', 'Waterbury', 'Athens',
       'Avondale', 'Marietta', 'Yuma', 'Wausau', 'Pasco', 'Oak Park',
       'Pensacola', 'League City', 'Gaithersburg', 'Lehi',
'Tuscaloosa',
       'Moreno Valley', 'Georgetown', 'Loveland', 'Chandler',
'Helena',
       'Kirkwood', 'Waco', 'Frankfort', 'Bethlehem', 'Grand Island',
       'Woodbury', 'Rogers', 'Clovis', 'Jupiter', 'Santa Barbara',
       'Cedar Hill', 'Norfolk', 'Draper', 'Ann Arbor', 'La Mesa',
       'Pocatello', 'Holland', 'Milford', 'Buffalo Grove', 'Lake
```

```
Forest',
       'Redding', 'Chico', 'Utica', 'Conway', 'Cheyenne', 'Owensboro',
       'Caldwell', 'Kenner', 'Nashua', 'Bartlett', 'Redwood City',
       'Lebanon', 'Santa Maria', 'Des Plaines', 'Longview',
       'Hendersonville', 'Waterloo', 'Cambridge', 'Palatine',
'Beverly',
       'Eugene', 'Oxnard', 'Renton', 'Glenview', 'Delray Beach',
       'Commerce City', 'Texas City', 'Wilson', 'Rio Rancho',
'Goldsboro',
       'Montebello', 'El Cajon', 'Beaumont', 'West Palm Beach',
'Abilene',
       'Normal', 'Saint Charles', 'Camarillo', 'Hillsboro', 'Burbank',
       'Modesto', 'Garden City', 'Atlantic City', 'Longmont', 'Davis',
       'Morgan Hill', 'Clifton', 'Sheboygan', 'East Point', 'Rapid
City',
       'Andover', 'Kissimmee', 'Shelton', 'Danbury', 'Sanford',
       'San Marcos', 'Greeley', 'Mansfield', 'Elyria', 'Twin Falls',
       'Coral Gables', 'Romeoville', 'Marlborough', 'Laurel', 'Bryan',
       'Pine Bluff', 'Aberdeen', 'Hagerstown', 'East Orange',
       'Arlington Heights', 'Oswego', 'Coon Rapids', 'San Clemente',
       'San Luis Obispo', 'Springdale', 'Lodi', 'Mason'],
dtype=object)
```

Q4. Caluclate the total selling price and profits for all orders

```
df1['Order Value'] = df1['Unit_Selling_Price'] * df1['Quantity']
df1.head(2)
```

```
   Order Id Order Date     Ship Mode   Segment        Country
City  \
0         1 2023-03-01  Second Class  Consumer  United States
Henderson
1         2 2023-08-15  Second Class  Consumer  United States
Henderson

      State  Postal Code Region   Category Sub Category       Product
Id  \
0  Kentucky        42420  South  Furniture    Bookcases  FUR-BO-
10001798
1  Kentucky        42420  South  Furniture       Chairs  FUR-CH-
10000454

   Quantity  Unit_Selling_Price  Unit_Profit  Total Profit  Total_sale
\
0         2               254.8         14.8          29.6       509.6

1         3               708.1        108.1         324.3      2124.3


   Order Value
```

```
0        509.6
1       2124.3
```

```
df1.groupby('Order Id')
```

```
<pandas.core.groupby.generic.DataFrameGroupBy object at
0x000002A63D028950>
```

```
df1.groupby('Order Id').sum(['quantity'])
```

```
          Postal Code  Quantity  Unit_Selling_Price  Unit_Profit  \
Order Id
1               42420         2               254.8         14.8
2               42420         3               708.1        108.1
3               90036         2                 9.5         -0.5
4               33311         5               940.8        160.8
5               33311         2                19.0         -1.0
...               ...       ...                 ...          ...
9990            33180         3                28.8         -1.2
9991            92627         2                86.4         16.4
9992            92627         2               254.8         34.8
9993            92627         4                29.1         -0.9
9994            92683         2               232.8         22.8

          Total Profit  Total_sale  Order Value
Order Id
1                 29.6       509.6        509.6
2                324.3      2124.3       2124.3
3                 -1.0        19.0         19.0
4                804.0      4704.0       4704.0
5                 -2.0        38.0         38.0
...                ...         ...          ...
9990              -3.6        86.4         86.4
9991              32.8       172.8        172.8
9992              69.6       509.6        509.6
9993              -3.6       116.4        116.4
9994              45.6       465.6        465.6

[9994 rows x 7 columns]
```

```
df1.groupby('Order Id').sum(['Quantity'])[['Order Value','Total
Profit']].reset_index()
```

```
      Order Id  Order Value  Total Profit
0            1        509.6          29.6
1            2       2124.3         324.3
2            3         19.0          -1.0
3            4       4704.0         804.0
4            5         38.0          -2.0
...        ...          ...           ...
9989      9990         86.4          -3.6
```

```
9990      9991        172.8          32.8
9991      9992        509.6          69.6
9992      9993        116.4          -3.6
9993      9994        465.6          45.6

[9994 rows x 3 columns]
```

```
df1.groupby('Order Id').sum(['Quantity'])[['Order Value', 'Total
Profit']]
```

```
          Order Value  Total Profit
Order Id
1               509.6          29.6
2              2124.3         324.3
3                19.0          -1.0
4              4704.0         804.0
5                38.0          -2.0
...               ...           ...
9990             86.4          -3.6
9991            172.8          32.8
9992            509.6          69.6
9993            116.4          -3.6
9994            465.6          45.6

[9994 rows x 2 columns]
```

Q5. Find all orders from the 'Technology' category that were shipped using 'second class' ship mode , ordered by order date

```
df1.head(4)
```

```
   Order Id Order Date       Ship Mode    Segment        Country  \
0         1 2023-03-01    Second Class   Consumer  United States
1         2 2023-08-15    Second Class   Consumer  United States
2         3 2023-01-10    Second Class  Corporate  United States
3         4 2022-06-18  Standard Class   Consumer  United States

              City       State  Postal Code Region         Category  \
0        Henderson    Kentucky        42420  South        Furniture
1        Henderson    Kentucky        42420  South        Furniture
2      Los Angeles  California        90036   West  Office Supplies
3  Fort Lauderdale     Florida        33311  South        Furniture

  Sub Category       Product Id  Quantity  Unit_Selling_Price
Unit_Profit  \
0    Bookcases  FUR-BO-10001798         2               254.8
14.8
1       Chairs  FUR-CH-10000454         3               708.1
108.1
2       Labels  OFF-LA-10000240         2                 9.5
```

```
-0.5
3       Tables  FUR-TA-10000577         5               940.8
160.8

   Total Profit  Total_sale  Order Value
0          29.6       509.6        509.6
1         324.3      2124.3       2124.3
2          -1.0        19.0         19.0
3         804.0      4704.0       4704.0
```

```
df1[(df1['Category'] == 'Technology') & (df1['Ship Mode'] == 'Second
Class')]
```

```
      Order Id Order Date     Ship Mode      Segment        Country  \
19          20 2022-01-11  Second Class     Consumer  United States
26          27 2022-02-08  Second Class     Consumer  United States
161        162 2023-11-27  Second Class     Consumer  United States
181        182 2022-06-25  Second Class    Corporate  United States
182        183 2022-01-10  Second Class  Home Office  United States
...        ...        ...           ...          ...            ...
9891      9892 2023-05-21  Second Class    Corporate  United States
9892      9893 2023-01-02  Second Class    Corporate  United States
9949      9950 2022-01-07  Second Class    Corporate  United States
9950      9951 2022-03-19  Second Class    Corporate  United States
9960      9961 2023-02-26  Second Class  Home Office  United States

               City         State  Postal Code   Region    Category  \
19    San Francisco    California        94109     West  Technology
26      Los Angeles    California        90049     West  Technology
161    Philadelphia  Pennsylvania        19134     East  Technology
181         Decatur      Illinois        62521  Central  Technology
182          Monroe     Louisiana        71203    South  Technology
...             ...           ...          ...      ...         ...
9891      Milwaukee     Wisconsin        53209  Central  Technology
9892      Milwaukee     Wisconsin        53209  Central  Technology
9949   Indianapolis       Indiana        46203  Central  Technology
9950   Indianapolis       Indiana        46203  Central  Technology
9960       Florence      Kentucky        41042    South  Technology

     Sub Category       Product Id  Quantity  Unit_Selling_Price
Unit_Profit  \
19         Phones  TEC-PH-10001949         3               203.7
33.7
26    Accessories  TEC-AC-10003027         3                87.3
7.3
161   Accessories  TEC-AC-10003657         2                47.5
7.5
181   Accessories  TEC-AC-10004659         7               397.7
17.7
182        Phones  TEC-PH-10003273         4               475.0
```

```
5.0
...           ...               ...       ...                  ...
...
9891  Accessories  TEC-AC-10003116         7               116.4
6.4
9892       Phones  TEC-PH-10002262         5               291.0
31.0
9949  Accessories  TEC-AC-10004568         3                77.6
7.6
9950  Accessories  TEC-AC-10001714         1                38.8
-1.2
9960  Accessories  TEC-AC-10002305         1                19.0
-1.0

      Total Profit  Total_sale  Order Value
19           101.1       611.1        611.1
26            21.9       261.9        261.9
161           15.0        95.0         95.0
181          123.9      2783.9       2783.9
182           20.0      1900.0       1900.0
...            ...         ...          ...
9891          44.8       814.8        814.8
9892         155.0      1455.0       1455.0
9949          22.8       232.8        232.8
9950          -1.2        38.8         38.8
9960          -1.0        19.0         19.0

[366 rows x 18 columns]
```

Q6. Find the average order value

```
df1.head(4)
```

```
   Order Id Order Date       Ship Mode    Segment        Country  \
0         1 2023-03-01    Second Class   Consumer  United States
1         2 2023-08-15    Second Class   Consumer  United States
2         3 2023-01-10    Second Class  Corporate  United States
3         4 2022-06-18  Standard Class   Consumer  United States

              City       State  Postal Code Region         Category  \
0        Henderson    Kentucky        42420  South        Furniture
1        Henderson    Kentucky        42420  South        Furniture
2      Los Angeles  California        90036   West  Office Supplies
3  Fort Lauderdale     Florida        33311  South        Furniture

  Sub Category       Product Id  Quantity  Unit_Selling_Price
Unit_Profit  \
0    Bookcases  FUR-BO-10001798         2               254.8
14.8
1       Chairs  FUR-CH-10000454         3               708.1
```

```
108.1
2       Labels  OFF-LA-10000240         2                 9.5
-0.5
3       Tables  FUR-TA-10000577         5               940.8
160.8

   Total Profit  Total_sale  Order Value
0          29.6       509.6        509.6
1         324.3      2124.3       2124.3
2          -1.0        19.0         19.0
3         804.0      4704.0       4704.0
```

```
import numpy as np
np.mean(df1['Total_sale'])
```

```
1108.5979787872725
```

Q7. Find the city with the highest total quantity of products ordered .

```
df1.head(4)
```

```
   Order Id Order Date       Ship Mode    Segment        Country  \
0         1 2023-03-01    Second Class   Consumer  United States
1         2 2023-08-15    Second Class   Consumer  United States
2         3 2023-01-10    Second Class  Corporate  United States
3         4 2022-06-18  Standard Class   Consumer  United States

               City       State  Postal Code Region         Category  \
0        Henderson    Kentucky        42420  South        Furniture
1        Henderson    Kentucky        42420  South        Furniture
2      Los Angeles  California        90036   West  Office Supplies
3  Fort Lauderdale     Florida        33311  South        Furniture

  Sub Category       Product Id  Quantity  Unit_Selling_Price
Unit_Profit  \
0    Bookcases  FUR-BO-10001798         2               254.8
14.8
1       Chairs  FUR-CH-10000454         3               708.1
108.1
2       Labels  OFF-LA-10000240         2                 9.5
-0.5
3       Tables  FUR-TA-10000577         5               940.8
160.8

   Total Profit  Total_sale  Order Value
0          29.6       509.6        509.6
1         324.3      2124.3       2124.3
2          -1.0        19.0         19.0
3         804.0      4704.0       4704.0
```

```
df1.groupby('City').sum('Quantity')
['Quantity'].reset_index().sort_values
```

```
<bound method DataFrame.sort_values of             City  Quantity
0       Aberdeen         3
1        Abilene         2
2          Akron        65
3     Albuquerque        65
4      Alexandria        84
..            ...       ...
526    Woonsocket        15
527       Yonkers        57
528          York        19
529       Yucaipa         5
530          Yuma        22

[531 rows x 2 columns]>
```

```
df1.groupby('City').sum('Qunatity')
['Quantity'].reset_index().sort_values(by = ['Quantity'], ascending =
False)
```

```
             City  Quantity
329  New York City      3417
266    Los Angeles      2879
374   Philadelphia      1981
438  San Francisco      1935
452        Seattle      1590
..             ...       ...
213      Iowa City         1
386    Port Orange         1
257    Lindenhurst         1
259      Littleton         1
140         Elyria         1

[531 rows x 2 columns]
```

Q8. Rank orders in each region by quatity in descending order

```
df1[['Order Id', 'Region', 'Quantity']].sort_values(by = ['Region',
'Quantity'], ascending = [True, False])
```

```
      Order Id   Region  Quantity
660        661  Central        14
1045      1046  Central        14
7387      7388  Central        14
8074      8075  Central        14
9515      9516  Central        14
...        ...      ...       ...
9752      9753     West         1
```

```
9838      9839     West          1
9902      9903     West          1
9953      9954     West          1
9986      9987     West          1

[9994 rows x 3 columns]
```

Q9. List all orders in the first quarter of any year (Jan to Mar) , including the total cost for these orders

```
df1.head(4)
```

```
   Order Id Order Date       Ship Mode    Segment        Country  \
0         1 2023-03-01    Second Class   Consumer  United States
1         2 2023-08-15    Second Class   Consumer  United States
2         3 2023-01-10    Second Class  Corporate  United States
3         4 2022-06-18  Standard Class   Consumer  United States

              City       State  Postal Code Region         Category  \
0        Henderson    Kentucky        42420  South        Furniture
1        Henderson    Kentucky        42420  South        Furniture
2      Los Angeles  California        90036   West  Office Supplies
3  Fort Lauderdale     Florida        33311  South        Furniture

  Sub Category       Product Id  Quantity  Unit_Selling_Price
Unit_Profit  \
0    Bookcases  FUR-BO-10001798         2               254.8
14.8
1       Chairs  FUR-CH-10000454         3               708.1
108.1
2       Labels  OFF-LA-10000240         2                 9.5
-0.5
3       Tables  FUR-TA-10000577         5               940.8
160.8

   Total Profit  Total_sale  Order Value
0          29.6       509.6        509.6
1         324.3      2124.3       2124.3
2          -1.0        19.0         19.0
3         804.0      4704.0       4704.0
```

```
df1['Order Date'].dt.month
```

```
0        3
1        8
2        1
3        6
4        7
        ..
9989     2
```

```
9990     3
9991     8
9992    11
9993     7
Name: Order Date, Length: 9994, dtype: int32
```

```
df1['Month'] = df1['Order Date'].dt.month
df1.head(4)
```

```
   Order Id Order Date       Ship Mode    Segment        Country  \
0         1 2023-03-01    Second Class   Consumer  United States
1         2 2023-08-15    Second Class   Consumer  United States
2         3 2023-01-10    Second Class  Corporate  United States
3         4 2022-06-18  Standard Class   Consumer  United States

               City       State  Postal Code Region         Category  \
0        Henderson    Kentucky        42420  South        Furniture
1        Henderson    Kentucky        42420  South        Furniture
2      Los Angeles  California        90036   West  Office Supplies
3  Fort Lauderdale     Florida        33311  South        Furniture

  Sub Category       Product Id  Quantity  Unit_Selling_Price
Unit_Profit  \
0    Bookcases  FUR-BO-10001798         2               254.8
14.8
1       Chairs  FUR-CH-10000454         3               708.1
108.1
2       Labels  OFF-LA-10000240         2                 9.5
-0.5
3       Tables  FUR-TA-10000577         5               940.8
160.8

   Total Profit  Total_sale  Order Value  Month
0          29.6       509.6        509.6      3
1         324.3      2124.3       2124.3      8
2          -1.0        19.0         19.0      1
3         804.0      4704.0       4704.0      6
```

```
df1[df1['Month'].isin([1,2,3])]
```

```
      Order Id Order Date       Ship Mode      Segment        Country
\
0            1 2023-03-01    Second Class     Consumer  United States

2            3 2023-01-10    Second Class    Corporate  United States

5            6 2022-03-13             NaN     Consumer  United States

7            8 2022-01-25  Standard Class     Consumer  United States

8            9 2023-03-23             NaN     Consumer  United States
```

```
...        ...        ...             ...          ...            ...
9963      9964 2022-01-17        Same Day     Consumer  United States
9966      9967 2022-02-21    Second Class    Corporate  United States
9973      9974 2023-03-07  Standard Class  Home Office  United States
9989      9990 2023-02-18    Second Class     Consumer  United States
9990      9991 2023-03-17  Standard Class     Consumer  United States

               City         State  Postal Code Region         Category
\
0         Henderson      Kentucky        42420  South        Furniture
2       Los Angeles    California        90036   West  Office Supplies
5       Los Angeles    California        90032   West        Furniture
7       Los Angeles    California        90032   West       Technology
8       Los Angeles    California        90032   West  Office Supplies
...             ...           ...          ...    ...              ...
9963  Philadelphia  Pennsylvania        19140   East  Office Supplies
9966        Newark      Delaware        19711   East  Office Supplies
9973   Los Angeles    California        90032   West       Technology
9989         Miami       Florida        33180  South        Furniture
9990    Costa Mesa    California        92627   West        Furniture

     Sub Category       Product Id  Quantity  Unit_Selling_Price
Unit_Profit  \
0       Bookcases  FUR-BO-10001798         2               254.8
14.8
2          Labels  OFF-LA-10000240         2                 9.5
-0.5
5     Furnishings  FUR-FU-10001487         7                48.5
-1.5
7          Phones  TEC-PH-10002275         6               864.5
4.5
8         Binders  OFF-BI-10003910         3                19.6
-0.4
```

```
...           ...              ...       ...                  ...
...
9963        Paper  OFF-PA-10003072         2                  9.7
-0.3
9966    Envelopes  OFF-EN-10004483         7                105.6
5.6
9973       Phones  TEC-PH-10004080         5                259.2
29.2
9989  Furnishings  FUR-FU-10001889         3                 28.8
-1.2
9990  Furnishings  FUR-FU-10000747         2                 86.4
16.4

      Total Profit  Total_sale  Order Value  Month
0             29.6       509.6        509.6      3
2             -1.0        19.0         19.0      1
5            -10.5       339.5        339.5      3
7             27.0      5187.0       5187.0      1
8             -1.2        58.8         58.8      3
...            ...         ...          ...    ...
9963          -0.6        19.4         19.4      1
9966          39.2       739.2        739.2      2
9973         146.0      1296.0       1296.0      3
9989          -3.6        86.4         86.4      2
9990          32.8       172.8        172.8      3

[2493 rows x 19 columns]

df1[df1['Order Date'].dt.quarter.isin([1])]

      Order Id Order Date       Ship Mode      Segment        Country
\
0            1 2023-03-01    Second Class     Consumer  United States

2            3 2023-01-10    Second Class    Corporate  United States

5            6 2022-03-13             NaN     Consumer  United States

7            8 2022-01-25  Standard Class     Consumer  United States

8            9 2023-03-23             NaN     Consumer  United States

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

9963      9964 2022-01-17        Same Day     Consumer  United States

9966      9967 2022-02-21    Second Class    Corporate  United States

9973      9974 2023-03-07  Standard Class  Home Office  United States

9989      9990 2023-02-18    Second Class     Consumer  United States
```

```
9990      9991 2023-03-17  Standard Class     Consumer  United States


               City         State  Postal Code Region         Category
\
0         Henderson      Kentucky        42420  South        Furniture

2       Los Angeles    California        90036   West  Office Supplies

5       Los Angeles    California        90032   West        Furniture

7       Los Angeles    California        90032   West       Technology

8       Los Angeles    California        90032   West  Office Supplies

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

9963  Philadelphia  Pennsylvania        19140   East  Office Supplies

9966        Newark      Delaware        19711   East  Office Supplies

9973   Los Angeles    California        90032   West       Technology

9989         Miami       Florida        33180  South        Furniture

9990    Costa Mesa    California        92627   West        Furniture


     Sub Category       Product Id  Quantity  Unit_Selling_Price
Unit_Profit  \
0       Bookcases  FUR-BO-10001798         2               254.8
14.8
2          Labels  OFF-LA-10000240         2                 9.5
-0.5
5     Furnishings  FUR-FU-10001487         7                48.5
-1.5
7          Phones  TEC-PH-10002275         6               864.5
4.5
8         Binders  OFF-BI-10003910         3                19.6
-0.4
...            ...              ...       ...                 ...
...
9963        Paper  OFF-PA-10003072         2                 9.7
-0.3
9966    Envelopes  OFF-EN-10004483         7               105.6
5.6
9973       Phones  TEC-PH-10004080         5               259.2
29.2
9989  Furnishings  FUR-FU-10001889         3                28.8
-1.2
```

```
9990  Furnishings  FUR-FU-10000747         2                86.4
16.4

      Total Profit  Total_sale  Order Value  Month
0             29.6       509.6        509.6      3
2             -1.0        19.0         19.0      1
5            -10.5       339.5        339.5      3
7             27.0      5187.0       5187.0      1
8             -1.2        58.8         58.8      3
...            ...         ...          ...    ...
9963          -0.6        19.4         19.4      1
9966          39.2       739.2        739.2      2
9973         146.0      1296.0       1296.0      3
9989          -3.6        86.4         86.4      2
9990          32.8       172.8        172.8      3

[2493 rows x 19 columns]
```

```
df1['year'] = df1['Order Date'].dt.year
df1.head(4)
```

```
   Order Id Order Date       Ship Mode    Segment        Country  \
0         1 2023-03-01    Second Class   Consumer  United States
1         2 2023-08-15    Second Class   Consumer  United States
2         3 2023-01-10    Second Class  Corporate  United States
3         4 2022-06-18  Standard Class   Consumer  United States

              City       State  Postal Code Region         Category  \
0        Henderson    Kentucky        42420  South        Furniture
1        Henderson    Kentucky        42420  South        Furniture
2      Los Angeles  California        90036   West  Office Supplies
3  Fort Lauderdale     Florida        33311  South        Furniture

  Sub Category       Product Id  Quantity  Unit_Selling_Price
Unit_Profit  \
0    Bookcases  FUR-BO-10001798         2               254.8
14.8
1       Chairs  FUR-CH-10000454         3               708.1
108.1
2       Labels  OFF-LA-10000240         2                 9.5
-0.5
3       Tables  FUR-TA-10000577         5               940.8
160.8

   Total Profit  Total_sale  Order Value  Month  year
0          29.6       509.6        509.6      3  2023
1         324.3      2124.3       2124.3      8  2023
2          -1.0        19.0         19.0      1  2023
3         804.0      4704.0       4704.0      6  2022
```

Q9. Find top 10 highest profit generating products

```
df1.head(4)
```

```
   Order Id Order Date       Ship Mode    Segment        Country  \
0         1 2023-03-01    Second Class   Consumer  United States
1         2 2023-08-15    Second Class   Consumer  United States
2         3 2023-01-10    Second Class  Corporate  United States
3         4 2022-06-18  Standard Class   Consumer  United States

               City       State  Postal Code Region         Category  \
0        Henderson    Kentucky        42420  South        Furniture
1        Henderson    Kentucky        42420  South        Furniture
2      Los Angeles  California        90036   West  Office Supplies
3  Fort Lauderdale     Florida        33311  South        Furniture

  Sub Category       Product Id  Quantity  Unit_Selling_Price
Unit_Profit  \
0    Bookcases  FUR-BO-10001798         2               254.8
14.8
1       Chairs  FUR-CH-10000454         3               708.1
108.1
2       Labels  OFF-LA-10000240         2                 9.5
-0.5
3       Tables  FUR-TA-10000577         5               940.8
160.8

   Total Profit  Total_sale  Order Value  Month  year
0          29.6       509.6        509.6      3  2023
1         324.3      2124.3       2124.3      8  2023
2          -1.0        19.0         19.0      1  2023
3         804.0      4704.0       4704.0      6  2022
```

```
df1.groupby('Product Id').sum(['Total Profit'])['Total
Profit'].reset_index().sort_values(by = 'Total Profit', ascending =
False).head(10)
```

```
          Product Id  Total Profit
1614  TEC-CO-10004722       24816.0
1642  TEC-MA-10002412       21746.4
657   OFF-BI-10000545       17867.7
1604  TEC-CO-10001449       15948.0
80    FUR-CH-10002024       13930.7
776   OFF-BI-10003527       12792.9
1724  TEC-PH-10001459       11481.9
1624  TEC-MA-10000822       10102.3
318   FUR-TA-10000198       10015.1
1630  TEC-MA-10001047        9989.0
```

```
df1.groupby('Product Id').sum(['Total Profit'])['Total
Profit'].reset_index().sort_values(by = 'Total Profit', ascending =
False).head(10)
```

	Product Id	Total Profit
1614	TEC-CO-10004722	24816.0
1642	TEC-MA-10002412	21746.4
657	OFF-BI-10000545	17867.7
1604	TEC-CO-10001449	15948.0
80	FUR-CH-10002024	13930.7
776	OFF-BI-10003527	12792.9
1724	TEC-PH-10001459	11481.9
1624	TEC-MA-10000822	10102.3
318	FUR-TA-10000198	10015.1
1630	TEC-MA-10001047	9989.0

Q10. Find top 3 highest selling products in each region according to quantity

```
df1.groupby(['Region', 'Product Id']).sum('Quantity')
['Quantity'].reset_index()
```

	Region	Product Id	Quantity
0	Central	FUR-BO-10000112	9
1	Central	FUR-BO-10000362	3
2	Central	FUR-BO-10000468	4
3	Central	FUR-BO-10000711	3
4	Central	FUR-BO-10000780	17
...	...	...	...
5293	West	TEC-PH-10004896	3
5294	West	TEC-PH-10004897	3
5295	West	TEC-PH-10004908	22
5296	West	TEC-PH-10004922	9
5297	West	TEC-PH-10004977	10

[5298 rows x 3 columns]

```
x = df1.groupby(['Region', 'Product Id']).sum('Quantity')
['Quantity'].reset_index().sort_values(by = ['Region', 'Quantity'],
ascending = [True, False])
x
```

	Region	Product Id	Quantity
461	Central	OFF-BI-10000301	34
474	Central	OFF-BI-10000756	33
470	Central	OFF-BI-10000546	29
491	Central	OFF-BI-10001249	29
61	Central	FUR-CH-10002304	27
...	...	...	...
5219	West	TEC-PH-10002185	1

```
5227     West  TEC-PH-10002447         1
5229     West  TEC-PH-10002483         1
5256     West  TEC-PH-10003357         1
5276     West  TEC-PH-10004120         1

[5298 rows x 3 columns]
```

```
x.groupby('Region').head(3).reset_index()
```

```
    index   Region       Product Id  Quantity
0     461  Central  OFF-BI-10000301        34
1     474  Central  OFF-BI-10000756        33
2     470  Central  OFF-BI-10000546        29
3    2190     East  OFF-PA-10001970        33
4    1914     East  OFF-BI-10003656        32
5    1549     East  FUR-FU-10004848        31
6    3560    South  OFF-ST-10003716        26
7    2757    South  FUR-CH-10000513        24
8    3231    South  OFF-BI-10004728        24
9    5088     West  TEC-AC-10003832        45
10   4321     West  OFF-BI-10000174        32
11   4346     West  OFF-BI-10001036        31
```

Q11. Total sales in each region

```
df1.groupby('Region').sum(['Total_sale'])
['Total_sale'].reset_index().sort_values(by = ['Total_sale'],
ascending = False)
```

```
    Region  Total_sale
3     West   3467409.6
1     East   3257983.8
0  Central   2387881.2
2    South   1966053.6
```

```
df1.head(4)
```

```
   Order Id Order Date       Ship Mode    Segment        Country  \
0         1 2023-03-01    Second Class   Consumer  United States
1         2 2023-08-15    Second Class   Consumer  United States
2         3 2023-01-10    Second Class  Corporate  United States
3         4 2022-06-18  Standard Class   Consumer  United States

              City       State  Postal Code Region         Category  \
0        Henderson    Kentucky        42420  South        Furniture
1        Henderson    Kentucky        42420  South        Furniture
2      Los Angeles  California        90036   West  Office Supplies
3  Fort Lauderdale     Florida        33311  South        Furniture

  Sub Category       Product Id  Quantity  Unit_Selling_Price
Unit_Profit  \
```

```
0    Bookcases  FUR-BO-10001798         2               254.8
14.8
1       Chairs  FUR-CH-10000454         3               708.1
108.1
2       Labels  OFF-LA-10000240         2                 9.5
-0.5
3       Tables  FUR-TA-10000577         5               940.8
160.8

   Total Profit  Total_sale  Order Value  Month  year
0          29.6       509.6        509.6      3  2023
1         324.3      2124.3       2124.3      8  2023
2          -1.0        19.0         19.0      1  2023
3         804.0      4704.0       4704.0      6  2022
```

Q12. Find top 5 highest selling products in each region according to sales

```
df1.groupby(['Region', 'Product Id']).sum(['Total_Sale'])
#.sort_values(by = ['Total_Sale'],ascending=False)
```

```
                          Order Id  Postal Code  Quantity
Unit_Selling_Price  \
Region  Product Id

Central FUR-BO-10000112      3513        60653         9
805.1
        FUR-BO-10000362      1595        61604         3
349.2
        FUR-BO-10000468      3031        48146         4
180.5
        FUR-BO-10000711      5221        48234         3
203.7
        FUR-BO-10000780     16297       229596        17
2223.1
...                           ...          ...       ...
...
West    TEC-PH-10004896      1088        94601         3
68.6
        TEC-PH-10004897      4320        87105         3
19.2
        TEC-PH-10004908     38344       421143        22
1433.2
        TEC-PH-10004922     11121       180757         9
459.2
        TEC-PH-10004977     11216       262951        10
1500.9

                          Unit_Profit  Total Profit  Total_sale  Order
```

```
Value  \
Region  Product Id

Central FUR-BO-10000112         45.1          405.9       7245.9
7245.9
        FUR-BO-10000362         49.2          147.6       1047.6
1047.6
        FUR-BO-10000468         30.5          122.0        722.0
722.0
        FUR-BO-10000711          3.7           11.1        611.1
611.1
        FUR-BO-10000780        243.1         1503.7      14003.7
14003.7
...                              ...           ...          ...
...
West    TEC-PH-10004896          8.6           25.8        205.8
205.8
        TEC-PH-10004897         -0.8           -2.4         57.6
57.6
        TEC-PH-10004908        123.2          680.4       7300.4
7300.4
        TEC-PH-10004922         39.2          199.2       2299.2
2299.2
        TEC-PH-10004977         90.9          311.2       5101.2
5101.2

                           Month   year
Region  Product Id
Central FUR-BO-10000112        4   2023
        FUR-BO-10000362        1   2022
        FUR-BO-10000468        9   2022
        FUR-BO-10000711        7   2023
        FUR-BO-10000780       16   6068
...                          ...    ...
West    TEC-PH-10004896        1   2022
        TEC-PH-10004897       12   2023
        TEC-PH-10004908       25  10114
        TEC-PH-10004922       17   4044
        TEC-PH-10004977       19   6067

[5298 rows x 10 columns]
```

```
df1.groupby(['Region', 'Product
Id']).sum(['Total_Sale']).reset_index()
```

```
       Region       Product Id  Order Id  Postal Code  Quantity  \
0     Central  FUR-BO-10000112      3513        60653         9
1     Central  FUR-BO-10000362      1595        61604         3
2     Central  FUR-BO-10000468      3031        48146         4
3     Central  FUR-BO-10000711      5221        48234         3
```

```
4     Central  FUR-BO-10000780     16297       229596         17
...       ...              ...       ...          ...        ...
5293     West  TEC-PH-10004896      1088        94601          3
5294     West  TEC-PH-10004897      4320        87105          3
5295     West  TEC-PH-10004908     38344       421143         22
5296     West  TEC-PH-10004922     11121       180757          9
5297     West  TEC-PH-10004977     11216       262951         10

      Unit_Selling_Price  Unit_Profit  Total Profit  Total_sale  Order
Value  \
0                  805.1         45.1         405.9      7245.9
7245.9
1                  349.2         49.2         147.6      1047.6
1047.6
2                  180.5         30.5         122.0       722.0
722.0
3                  203.7          3.7          11.1       611.1
611.1
4                 2223.1        243.1        1503.7     14003.7
14003.7
...                  ...          ...          ...         ...
...
5293                68.6          8.6          25.8       205.8
205.8
5294                19.2         -0.8          -2.4        57.6
57.6
5295              1433.2        123.2         680.4      7300.4
7300.4
5296               459.2         39.2         199.2      2299.2
2299.2
5297              1500.9         90.9         311.2      5101.2
5101.2

      Month   year
0         4   2023
1         1   2022
2         9   2022
3         7   2023
4        16   6068
...     ...    ...
5293      1   2022
5294     12   2023
5295     25  10114
5296     17   4044
5297     19   6067

[5298 rows x 12 columns]
```

```
x = df1.groupby(['Region', 'Product Id']).sum(['Total_sale'])
['Total_sale'].reset_index()
```

```
y = x.sort_values(by = ['Region', 'Total_sale'], ascending=[True,
False])
y.groupby('Region').head(5).reset_index()
```

	index	Region	Product Id	Total_sale
0	469	Central	OFF-BI-10000545	125827.5
1	1166	Central	TEC-CO-10004722	84875.0
2	1168	Central	TEC-MA-10000822	77509.8
3	488	Central	OFF-BI-10001120	55282.5
4	617	Central	OFF-BI-10004995	42210.0
5	2556	East	TEC-CO-10004722	106421.0
6	2565	East	TEC-MA-10001047	81549.0
7	1342	East	FUR-BO-10004834	66364.2
8	2548	East	TEC-CO-10001449	60948.0
9	1369	East	FUR-CH-10002024	60189.6
10	3691	South	TEC-MA-10002412	130406.4
11	3719	South	TEC-PH-10001459	73932.1
12	2910	South	FUR-TA-10000198	68789.9
13	3687	South	TEC-MA-10001127	50499.2
14	3192	South	OFF-BI-10003527	32029.4
15	5088	West	TEC-AC-10003832	61170.8
16	5126	West	TEC-CO-10004722	53760.0
17	4964	West	OFF-SU-10000151	53337.9
18	4099	West	FUR-TA-10003473	52669.5
19	3890	West	FUR-CH-10003973	43596.4

```
x.sort_values(by = ['Region', 'Total_sale'], ascending=[True, False])
```

	Region	Product Id	Total_sale
469	Central	OFF-BI-10000545	125827.5
1166	Central	TEC-CO-10004722	84875.0
1168	Central	TEC-MA-10000822	77509.8
488	Central	OFF-BI-10001120	55282.5
617	Central	OFF-BI-10004995	42210.0
...	...	...	...
4571	West	OFF-FA-10003472	0.0
4581	West	OFF-LA-10000134	0.0
4622	West	OFF-LA-10003714	0.0
4817	West	OFF-PA-10004000	0.0
4854	West	OFF-PA-10004948	0.0

```
[5298 rows x 3 columns]
```

Q14. Find month over month growth comparision for 2022 and 2023 sales eg: jan 2022 Vs jan 2023

```
df1.groupby(['Month', 'year'])
```

```
<pandas.core.groupby.generic.DataFrameGroupBy object at
0x000002A63F7E58D0>
```

```
df1.groupby(['Month', 'year']).sum('Total_Sale')
```

```
            Order Id  Postal Code  Quantity  Unit_Selling_Price
Unit_Profit  \
Month year

1     2022   2203269     25254924      1650              94712.5
9092.5
      2023   2159265     22745269      1543              88632.6
8272.6
2     2022   1942952     21548462      1511              90091.0
9101.0
      2023   1963083     21939004      1574             128124.2
11554.2
3     2022   2001085     20660876      1447              80106.0
7076.0
      2023   2204143     23815029      1651              82512.3
7642.3
4     2022   2214787     25361800      1648              95451.6
8801.6
      2023   1941539     22098149      1501             111568.6
10028.6
5     2022   2076080     23908332      1639              79448.3
7088.3
      2023   2038723     23553002      1561              86447.9
8187.9
6     2022   1967618     21814012      1528              94170.5
8300.5
      2023   1864583     20450282      1422              68976.5
6186.5
7     2022   2157975     23743535      1579              78652.2
7422.2
      2023   2418956     26432714      1844              90563.8
8163.8
8     2022   2237092     24724255      1706             104808.0
11768.0
      2023   2155249     23170424      1642              87733.6
8353.6
9     2022   2091402     21205155      1471              79142.2
6892.2
      2023   1767956     18928012      1281              76658.6
6678.6
10    2022   2153052     24592596      1744             118912.7
9632.7
      2023   2027157     23090094      1627             121061.5
```

```
12621.5
11    2022   2126015     23371960      1568              84225.3
7765.3
      2023   1989561     21892615      1487              75432.8
6522.8
12    2022   1990905     22506776      1585              95869.9
8679.9
      2023   2252568     24765375      1664             102556.1
9336.1
```

Month	year	Total Profit	Total_sale	Order Value
1	2022	43521.3	437431.3	437431.3
	2023	42525.5	434765.5	434765.5
2	2022	47461.1	444011.1	444011.1
	2023	67118.8	731638.8	731638.8
3	2022	33915.2	394105.2	394105.2
	2023	37491.9	393051.9	393051.9
4	2022	42620.9	476400.9	476400.9
	2023	49811.5	543231.5	543231.5
5	2022	35745.5	413625.5	413625.5
	2023	40517.9	410707.9	410707.9
6	2022	41220.3	465300.3	465300.3
	2023	30029.0	328939.0	328939.0
7	2022	35938.4	375278.4	375278.4
	2023	35923.7	422533.7	422533.7
8	2022	61852.4	534562.4	534562.4
	2023	44980.3	465010.3	465010.3
9	2022	41067.0	433887.0	433887.0
	2023	35860.5	420620.5	420620.5
10	2022	50857.8	601707.8	601707.8
	2023	67028.3	626498.3	626498.3
11	2022	44009.6	451809.6	451809.6
	2023	28350.6	334940.6	334940.6
12	2022	38861.8	447421.8	447421.8
	2023	43218.9	491848.9	491848.9

```
df1.groupby(['Month', 'year']).sum('Total_sale')['Total_sale']
```

```
Month  year
1      2022    437431.3
       2023    434765.5
2      2022    444011.1
       2023    731638.8
3      2022    394105.2
       2023    393051.9
4      2022    476400.9
       2023    543231.5
5      2022    413625.5
       2023    410707.9
```

```
6       2022    465300.3
        2023    328939.0
7       2022    375278.4
        2023    422533.7
8       2022    534562.4
        2023    465010.3
9       2022    433887.0
        2023    420620.5
10      2022    601707.8
        2023    626498.3
11      2022    451809.6
        2023    334940.6
12      2022    447421.8
        2023    491848.9
Name: Total_sale, dtype: float64
```

```
a = df1.groupby(['Month', 'year']).sum('Total_sale')
['Total_sale'].reset_index()
a
```

```
    Month  year  Total_sale
0       1  2022    437431.3
1       1  2023    434765.5
2       2  2022    444011.1
3       2  2023    731638.8
4       3  2022    394105.2
5       3  2023    393051.9
6       4  2022    476400.9
7       4  2023    543231.5
8       5  2022    413625.5
9       5  2023    410707.9
10      6  2022    465300.3
11      6  2023    328939.0
12      7  2022    375278.4
13      7  2023    422533.7
14      8  2022    534562.4
15      8  2023    465010.3
16      9  2022    433887.0
17      9  2023    420620.5
18     10  2022    601707.8
19     10  2023    626498.3
20     11  2022    451809.6
21     11  2023    334940.6
22     12  2022    447421.8
23     12  2023    491848.9
```

```
pivot_table = a.pivot_table(index='Month', columns='year',
values='Total_sale', aggfunc='sum')
pivot_table
```

```
year       2022      2023
Month
1      437431.3  434765.5
2      444011.1  731638.8
3      394105.2  393051.9
4      476400.9  543231.5
5      413625.5  410707.9
6      465300.3  328939.0
7      375278.4  422533.7
8      534562.4  465010.3
9      433887.0  420620.5
10     601707.8  626498.3
11     451809.6  334940.6
12     447421.8  491848.9
```

```
pivot_table = pivot_table.reset_index()
pivot_table
```

```
year  Month      2022      2023
0         1  437431.3  434765.5
1         2  444011.1  731638.8
2         3  394105.2  393051.9
3         4  476400.9  543231.5
4         5  413625.5  410707.9
5         6  465300.3  328939.0
6         7  375278.4  422533.7
7         8  534562.4  465010.3
8         9  433887.0  420620.5
9        10  601707.8  626498.3
10       11  451809.6  334940.6
11       12  447421.8  491848.9
```

```
pivot_table['YoY_growth'] = pivot_table[2022] - pivot_table[2023]
pivot_table
```

```
year  Month      2022      2023  YoY_growth
0         1  437431.3  434765.5      2665.8
1         2  444011.1  731638.8   -287627.7
2         3  394105.2  393051.9      1053.3
3         4  476400.9  543231.5    -66830.6
4         5  413625.5  410707.9      2917.6
5         6  465300.3  328939.0    136361.3
6         7  375278.4  422533.7    -47255.3
7         8  534562.4  465010.3     69552.1
8         9  433887.0  420620.5     13266.5
9        10  601707.8  626498.3    -24790.5
10       11  451809.6  334940.6    116869.0
11       12  447421.8  491848.9    -44427.1
```

```
df1['Month-Year'] = df1['Order Date'].dt.strftime('%Y-%m')
df.head(4)
```

```
   Order Id  Order Date       Ship Mode    Segment        Country  \
0         1  01-03-2023    Second Class   Consumer  United States
1         2  15-08-2023    Second Class   Consumer  United States
2         3  10-01-2023    Second Class  Corporate  United States
3         4  18-06-2022  Standard Class   Consumer  United States

               City       State  Postal Code Region         Category  \
0        Henderson    Kentucky        42420  South        Furniture
1        Henderson    Kentucky        42420  South        Furniture
2      Los Angeles  California        90036   West  Office Supplies
3  Fort Lauderdale     Florida        33311  South        Furniture

  Sub Category       Product Id  cost price  List Price  Quantity  \
0    Bookcases  FUR-BO-10001798         240         260         2
1       Chairs  FUR-CH-10000454         600         730         3
2       Labels  OFF-LA-10000240          10          10         2
3       Tables  FUR-TA-10000577         780         960         5

   Discount Percent
0                 2
1                 3
2                 5
3                 2
```

Q15. For each category which month had highest sales

```
df1.head(2).groupby(['Month-Year', 'Category']).sum(['Total_sale'])
['Total_sale'].reset_index()
```

```
  Month-Year   Category  Total_sale
0    2023-03  Furniture       509.6
1    2023-08  Furniture      2124.3
```

```
b = df1.groupby(['Category', 'Month-Year']).sum(['Total_sale'])
['Total_sale'].reset_index()
b
```

```
      Category Month-Year  Total_sale
0    Furniture    2022-01    170244.5
1    Furniture    2022-02    192062.9
2    Furniture    2022-03    150128.1
3    Furniture    2022-04    162806.9
4    Furniture    2022-05    110578.8
..         ...        ...         ...
67  Technology    2023-08    117665.2
68  Technology    2023-09    137873.1
69  Technology    2023-10    295586.5
```

```
70  Technology    2023-11    121226.6
71  Technology    2023-12    156836.6

[72 rows x 3 columns]
```

```
c = b.sort_values(by = ['Category', 'Total_sale'], ascending=[True,
False])
c
```

```
      Category Month-Year  Total_sale
19   Furniture    2023-08    230523.5
9    Furniture    2022-10    229024.8
13   Furniture    2023-02    217851.0
10   Furniture    2022-11    203673.6
1    Furniture    2022-02    192062.9
..         ...        ...         ...
67  Technology    2023-08    117665.2
48  Technology    2022-01    117116.5
51  Technology    2022-04    114747.7
54  Technology    2022-07    102816.1
49  Technology    2022-02     93493.2

[72 rows x 3 columns]
```

```
c.groupby('Category').head(1)
```

```
           Category Month-Year  Total_sale
19        Furniture    2023-08    230523.5
37  Office Supplies    2023-02    287244.6
69       Technology    2023-10    295586.5
```

```
df1.head(2)
```

```
   Order Id Order Date     Ship Mode   Segment        Country
City  \
0         1 2023-03-01  Second Class  Consumer  United States
Henderson
1         2 2023-08-15  Second Class  Consumer  United States
Henderson

      State  Postal Code Region   Category  ...       Product Id
Quantity  \
0  Kentucky        42420  South  Furniture  ...  FUR-BO-10001798
2
1  Kentucky        42420  South  Furniture  ...  FUR-CH-10000454
3

   Unit_Selling_Price  Unit_Profit  Total Profit  Total_sale  Order
Value  \
0               254.8         14.8          29.6       509.6
509.6
```

```
1               708.1        108.1         324.3      2124.3
2124.3

   Month  year  Month-Year
0      3  2023     2023-03
1      8  2023     2023-08

[2 rows x 21 columns]
```

Q16. Which sub category had highest growth by profit in 2023 compare to 2022

```
m = df1.groupby(['Sub Category', 'year']).sum(['Total Profit'])['Total
Profit'].reset_index()
m
```

	Sub Category	year	Total Profit
0	Accessories	2022	38798.3
1	Accessories	2023	40613.7
2	Appliances	2022	37088.7
3	Appliances	2023	17493.9
4	Art	2022	5333.0
5	Art	2023	5079.2
6	Binders	2022	40812.5
7	Binders	2023	57990.3
8	Bookcases	2022	25494.7
9	Bookcases	2023	25776.1
10	Chairs	2022	78672.5
11	Chairs	2023	75233.2
12	Copiers	2022	42703.1
13	Copiers	2023	26561.8
14	Envelopes	2022	2607.3
15	Envelopes	2023	3502.4
16	Fasteners	2022	370.7
17	Fasteners	2023	87.9
18	Furnishings	2022	22655.3
19	Furnishings	2023	16428.4
20	Labels	2022	2007.2
21	Labels	2023	2181.6
22	Machines	2022	34605.5
23	Machines	2023	56939.8
24	Paper	2022	16150.7
25	Paper	2023	16211.9
26	Phones	2022	63060.3
27	Phones	2023	76438.2
28	Storage	2022	42927.8
29	Storage	2023	53177.5
30	Supplies	2022	5161.0
31	Supplies	2023	9241.5

```
32       Tables  2022       58622.7
33       Tables  2023       39899.5
```

```
n = m.pivot_table(index = 'Sub Category', columns='year', values =
'Total Profit').reset_index()
n
```

year	Sub Category	2022	2023
0	Accessories	38798.3	40613.7
1	Appliances	37088.7	17493.9
2	Art	5333.0	5079.2
3	Binders	40812.5	57990.3
4	Bookcases	25494.7	25776.1
5	Chairs	78672.5	75233.2
6	Copiers	42703.1	26561.8
7	Envelopes	2607.3	3502.4
8	Fasteners	370.7	87.9
9	Furnishings	22655.3	16428.4
10	Labels	2007.2	2181.6
11	Machines	34605.5	56939.8
12	Paper	16150.7	16211.9
13	Phones	63060.3	76438.2
14	Storage	42927.8	53177.5
15	Supplies	5161.0	9241.5
16	Tables	58622.7	39899.5

```
n['diff'] = n[2023] - n[2022]
```

```
n.sort_values(by = ['diff'], ascending=False).head(1)
```

year	Sub Category	2022	2023	diff
11	Machines	34605.5	56939.8	22334.3