mazon-sales-data-perform-by-vikram

April 18, 2024

```python
[97]: import numpy as np
      import pandas as pd
      import seaborn as sn
      import matplotlib.pyplot as mp
```

```python
[98]: data = pd.read_csv('/content/Amazon Sales data.csv')
      sales_data = data.copy()
      sales_data.head()
```

```
[98]:                              Region                 Country       Item Type  \
      0             Australia and Oceania                  Tuvalu       Baby Food
      1   Central America and the Caribbean                Grenada          Cereal
      2                            Europe                  Russia  Office Supplies
      3               Sub-Saharan Africa  Sao Tome and Principe          Fruits
      4               Sub-Saharan Africa                  Rwanda  Office Supplies

        Sales Channel Order Priority Order Date   Order ID  Ship Date  Units Sold  \
      0       Offline              H  5/28/2010  669165933  6/27/2010        9925
      1        Online              C  8/22/2012  963881480  9/15/2012        2804
      2       Offline              L   5/2/2014  341417157   5/8/2014        1779
      3        Online              C  6/20/2014  514321792   7/5/2014        8102
      4       Offline              L   2/1/2013  115456712   2/6/2013        5062

        Unit Price  Unit Cost  Total Revenue  Total Cost  Total Profit
      0     255.28     159.42     2533654.00  1582243.50     951410.50
      1     205.70     117.11      576782.80   328376.44     248406.36
      2     651.21     524.96     1158502.59   933903.84     224598.75
      3       9.33       6.92       75591.66    56065.84      19525.82
      4     651.21     524.96     3296425.02  2657347.52     639077.50
```

```python
[99]: sales_data = sales_data[["Order ID","Order Date","Order Priority","Ship␣
      ↪Date","Item Type","Region","Country","Sales Channel","Units Sold","Unit␣
      ↪Price","Unit Cost","Total Revenue","Total Cost","Total Profit"]]
```

```python
[100]: sales_data.head()
```

```
[100]:      Order ID Order Date Order Priority  Ship Date        Item Type  \
     0  669165933  5/28/2010              H  6/27/2010        Baby Food
     1  963881480  8/22/2012              C  9/15/2012           Cereal
     2  341417157   5/2/2014              L   5/8/2014  Office Supplies
     3  514321792  6/20/2014              C   7/5/2014           Fruits
     4  115456712   2/1/2013              L   2/6/2013  Office Supplies

                               Region                Country Sales Channel  \
     0            Australia and Oceania                 Tuvalu       Offline
     1  Central America and the Caribbean              Grenada        Online
     2                           Europe                 Russia       Offline
     3               Sub-Saharan Africa  Sao Tome and Principe        Online
     4               Sub-Saharan Africa                 Rwanda       Offline

        Units Sold  Unit Price  Unit Cost  Total Revenue  Total Cost  Total Profit
     0        9925      255.28     159.42     2533654.00  1582243.50     951410.50
     1        2804      205.70     117.11      576782.80   328376.44     248406.36
     2        1779      651.21     524.96     1158502.59   933903.84     224598.75
     3        8102        9.33       6.92       75591.66    56065.84      19525.82
     4        5062      651.21     524.96     3296425.02  2657347.52     639077.50
```

```
[101]: sales_data.columns
```

```
[101]: Index(['Order ID', 'Order Date', 'Order Priority', 'Ship Date', 'Item Type',
              'Region', 'Country', 'Sales Channel', 'Units Sold', 'Unit Price',
              'Unit Cost', 'Total Revenue', 'Total Cost', 'Total Profit'],
             dtype='object')
```

```
[102]: sales_data.axes[0]
```

```
[102]: RangeIndex(start=0, stop=100, step=1)
```

```
[103]: sales_data.axes[1]
```

```
[103]: Index(['Order ID', 'Order Date', 'Order Priority', 'Ship Date', 'Item Type',
              'Region', 'Country', 'Sales Channel', 'Units Sold', 'Unit Price',
              'Unit Cost', 'Total Revenue', 'Total Cost', 'Total Profit'],
             dtype='object')
```

```
[104]: sales_data.dtypes
```

```
[104]: Order ID            int64
       Order Date         object
       Order Priority     object
       Ship Date          object
       Item Type          object
       Region             object
```

```
Country           object
Sales Channel     object
Units Sold         int64
Unit Price       float64
Unit Cost        float64
Total Revenue    float64
Total Cost       float64
Total Profit     float64
dtype: object
```

[105]: `sales_data.index`

[105]: RangeIndex(start=0, stop=100, step=1)

[106]: `sales_data.head()`

[106]:
```
    Order ID Order Date Order Priority  Ship Date       Item Type  \
0  669165933  5/28/2010              H  6/27/2010       Baby Food
1  963881480  8/22/2012              C  9/15/2012          Cereal
2  341417157   5/2/2014              L   5/8/2014  Office Supplies
3  514321792  6/20/2014              C   7/5/2014          Fruits
4  115456712   2/1/2013              L   2/6/2013  Office Supplies


                            Region                Country Sales Channel  \
0            Australia and Oceania                 Tuvalu       Offline
1  Central America and the Caribbean              Grenada        Online
2                           Europe                 Russia       Offline
3              Sub-Saharan Africa  Sao Tome and Principe        Online
4              Sub-Saharan Africa                 Rwanda       Offline

   Units Sold  Unit Price  Unit Cost  Total Revenue  Total Cost  Total Profit
0        9925      255.28     159.42     2533654.00  1582243.50     951410.50
1        2804      205.70     117.11      576782.80   328376.44     248406.36
2        1779      651.21     524.96     1158502.59   933903.84     224598.75
3        8102        9.33       6.92       75591.66    56065.84      19525.82
4        5062      651.21     524.96     3296425.02  2657347.52     639077.50
```

[107]: `sales_data.columns.isnull()`

[107]: array([False, False, False, False, False, False, False, False, False,
              False, False, False, False, False])

[108]: `sales_data.loc[:,["Total Revenue","Total Profit"]].iloc[:]`

[108]:
```
   Total Revenue  Total Profit
0     2533654.00     951410.50
1      576782.80     248406.36
```

3

```
2        1158502.59        224598.75
3          75591.66         19525.82
4        3296425.02        639077.50
..              …                …
95         97040.64         65214.72
96         58471.11         15103.47
97        228779.10         93748.05
98        471336.91        144521.02
99       3586605.09        889472.91

[100 rows x 2 columns]
```

[109]: `sales_data.shape`

[109]: (100, 14)

[110]: `np.corrcoef(sales_data.loc[:,'Total Revenue'].iloc[:],sales_data.loc[:,'Total␣`
       `↪Profit'].iloc[:])`

[110]: array([[1.   , 0.89732687],
              [0.89732687, 1.   ]])

[111]: `sales_data.set_index('Order ID',inplace=True)`

[112]: `sales_data.head()`

[112]:
```
            Order Date Order Priority  Ship Date        Item Type  \
Order ID
669165933   5/28/2010              H  6/27/2010        Baby Food
963881480   8/22/2012              C  9/15/2012           Cereal
341417157    5/2/2014              L   5/8/2014  Office Supplies
514321792   6/20/2014              C   7/5/2014           Fruits
115456712    2/1/2013              L   2/6/2013  Office Supplies

                                   Region               Country  \
Order ID
669165933            Australia and Oceania                Tuvalu
963881480  Central America and the Caribbean             Grenada
341417157                           Europe                Russia
514321792               Sub-Saharan Africa  Sao Tome and Principe
115456712               Sub-Saharan Africa                Rwanda

           Sales Channel  Units Sold  Unit Price  Unit Cost  Total Revenue  \
Order ID
669165933        Offline        9925      255.28     159.42     2533654.00
963881480         Online        2804      205.70     117.11      576782.80
341417157        Offline        1779      651.21     524.96     1158502.59
```

```
514321792         Online         8102         9.33         6.92         75591.66
115456712         Offline        5062       651.21       524.96       3296425.02


           Total Cost  Total Profit
Order ID
669165933  1582243.50    951410.50
963881480   328376.44    248406.36
341417157   933903.84    224598.75
514321792    56065.84     19525.82
115456712  2657347.52    639077.50
```
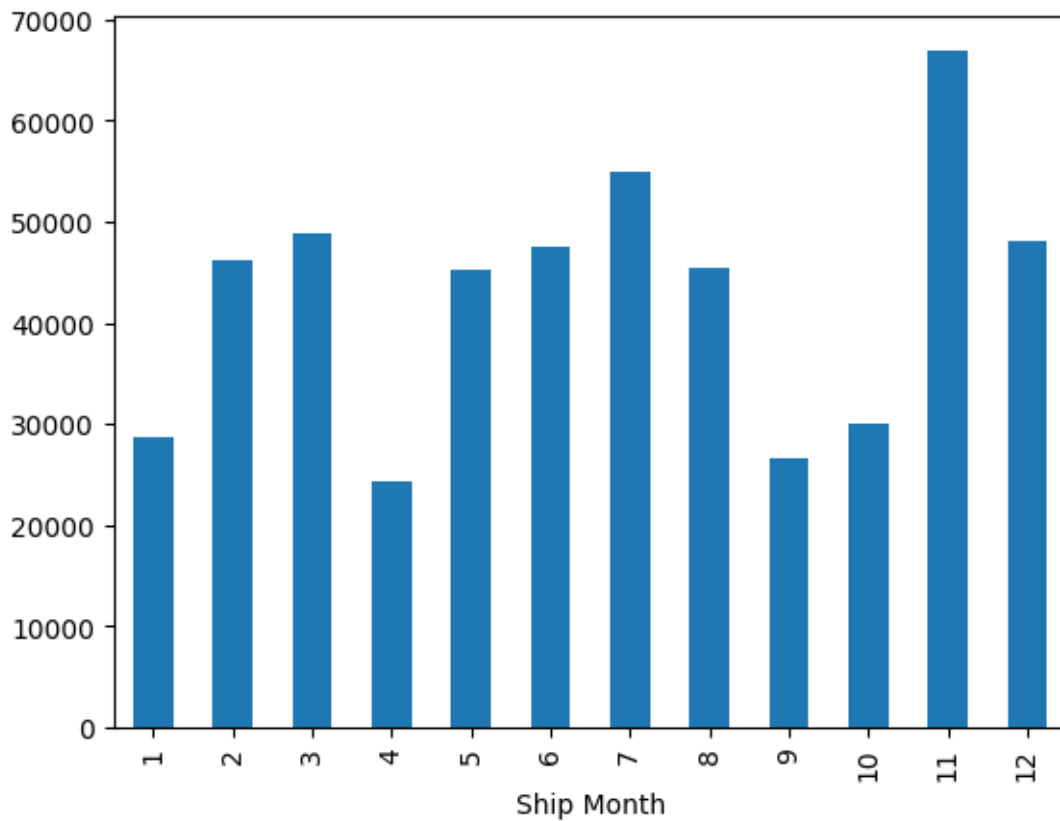
[113]: `sales_data['Ship Date'] = pd.to_datetime(sales_data['Ship Date'])`

[114]: `sales_data['Ship Month'] = sales_data['Ship Date'].dt.month`

[115]: `monthly_sales = sales_data.groupby('Ship Month')['Units Sold'].sum()`

[116]: `monthly_sales.plot(kind='bar')`

[116]: `<Axes: xlabel='Ship Month'>`

```
[117]:  np.average(sales_data['Total Profit'])
```

```
[117]:  441681.98399999994
```

At an average, the profit generated for a product is 441681.98.

```
[118]:  np.max(sales_data['Total Profit'])
```

```
[118]:  1719922.04
```

```
[119]:  np.min(sales_data['Total Profit'])
```

```
[119]:  1258.02
```

```
[120]:  np.var(sales_data['Total Profit'])
```

```
[120]:  190392340968.9648
```

Maximum and minimum profit generated are 1719922.04 and 1258.09 respectively.
Let's study total revenue

```
[121]:  np.max(sales_data['Total Revenue'])
```

```
[121]:  5997054.98
```

```
[122]:  np.min(sales_data['Total Revenue'])
```

```
[122]:  4870.26
```

```
[123]:  np.mean(sales_data['Total Revenue'])
```

```
[123]:  1373487.6831
```

```
[124]:  np.var(sales_data['Total Revenue'])
```

```
[124]:  2110366986501.2168
```

```
[125]:  np.std(sales_data['Total Revenue'])
```

```
[125]:  1452710.2211044077
```

```
[126]:  np.median(sales_data['Total Revenue'])
```

```
[126]:  752314.36
```

```
[127]:  np.percentile(sales_data['Total Revenue'],50,axis=0,overwrite_input=True)
```

```
[127]: 752314.36
```

```
[128]: sales_data['Total Revenue'].describe()
```

```
[128]: count    1.000000e+02
       mean     1.373488e+06
       std      1.460029e+06
       min      4.870260e+03
       25%      2.687212e+05
       50%      7.523144e+05
       75%      2.212045e+06
       max      5.997055e+06
       Name: Total Revenue, dtype: float64
```
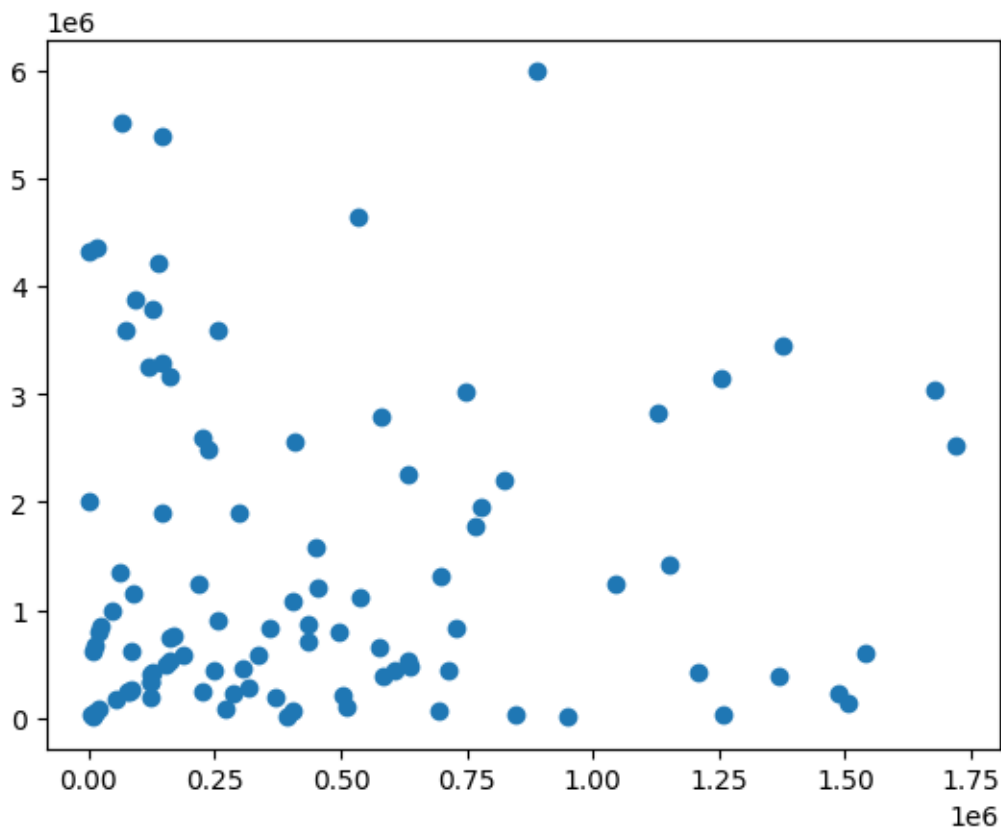
1 Maximum and minimum revenue generated by the product are 5997054.98 and 4870.26.

Revenue has very high variability in it's distribution. The median revenue generated is 752314.36

```
[129]: mp.scatter(sales_data['Total Profit'],sales_data['Total Revenue'])
```
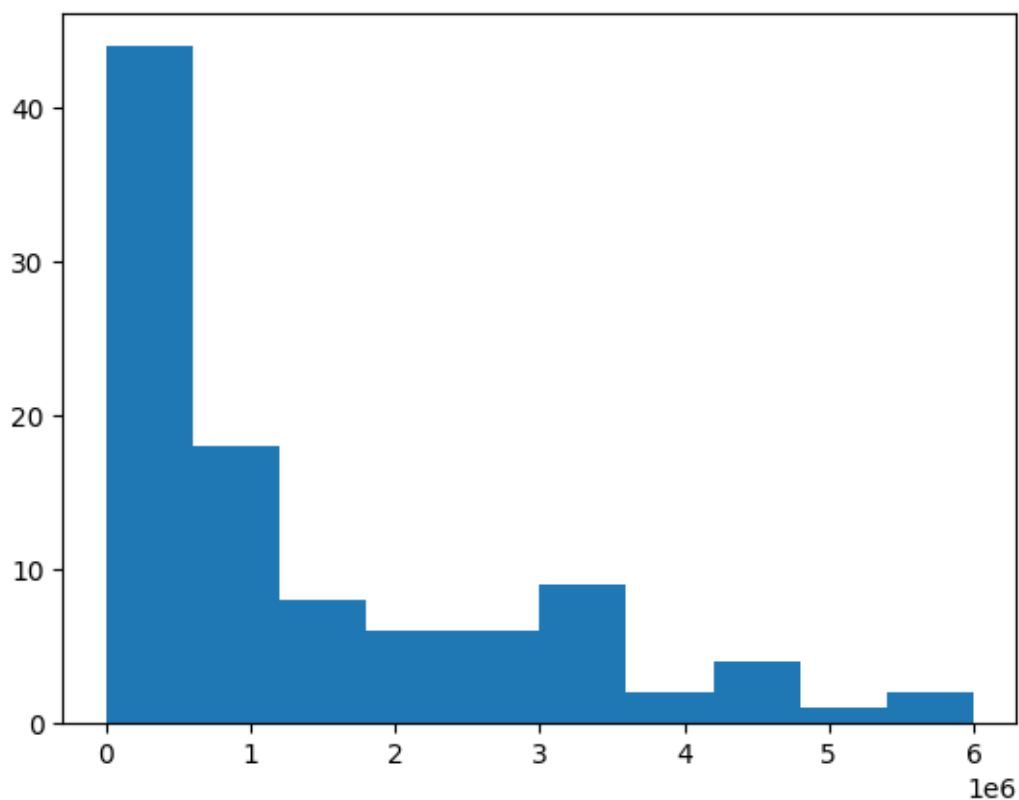
```
[129]: <matplotlib.collections.PathCollection at 0x78950c08d390>
```

The scatter plot also suggests that total profit and total revenue are directly proportional to each other.

```
[130]: mp.hist(sales_data['Total Revenue'])
```

```
[130]: (array([44., 18.,  8.,  6.,  6.,  9.,  2.,  4.,  1.,  2.]),
        array([4.87026000e+03, 6.04088732e+05, 1.20330720e+06, 1.80252568e+06,
               2.40174415e+06, 3.00096262e+06, 3.60018109e+06, 4.19939956e+06,
               4.79861804e+06, 5.39783651e+06, 5.99705498e+06]),
        <BarContainer object of 10 artists>)
```



```
[131]: np.correlate(sales_data['Total Revenue'],sales_data['Total Profit'])
```

```
[131]: array([6.25562019e+13])
```

```
[132]: sales_data.head()
```

```
[132]:            Order Date Order Priority  Ship Date       Item Type  \
        Order ID
```

```
669165933  5/28/2010              H 2010-06-27        Baby Food
963881480  8/22/2012              C 2012-09-15           Cereal
341417157   5/2/2014              L 2014-05-08  Office Supplies
514321792  6/20/2014              C 2014-07-05           Fruits
115456712   2/1/2013              L 2013-02-06  Office Supplies

                                 Region                  Country  \
Order ID
669165933              Australia and Oceania                Tuvalu
963881480  Central America and the Caribbean              Grenada
341417157                            Europe                Russia
514321792                Sub-Saharan Africa  Sao Tome and Principe
115456712                Sub-Saharan Africa                Rwanda

           Sales Channel  Units Sold  Unit Price  Unit Cost  Total Revenue  \
Order ID
669165933        Offline        9925      255.28     159.42        4870.26
963881480         Online        2804      205.70     117.11      435466.90
341417157        Offline        1779      651.21     524.96      247956.32
514321792         Online        8102        9.33       6.92       75591.66
115456712        Offline        5062      651.21     524.96      471336.91

            Total Cost  Total Profit  Ship Month
Order ID
669165933  1582243.50     951410.50           6
963881480   328376.44     248406.36           9
341417157   933903.84     224598.75           5
514321792    56065.84      19525.82           7
115456712  2657347.52     639077.50           2
```
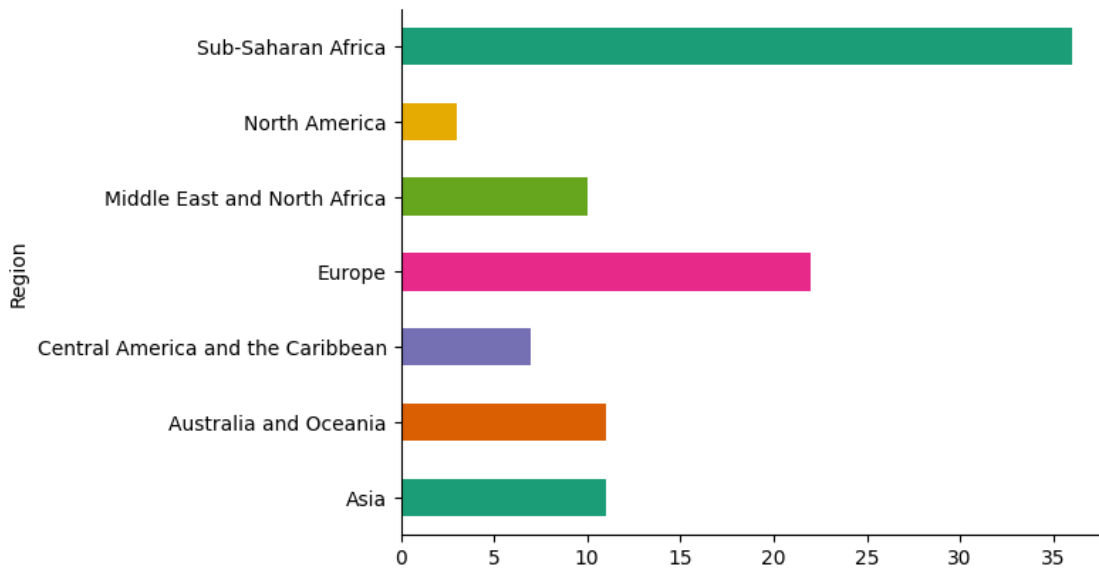
[133]:
```python
# @title Region

from matplotlib import pyplot as plt
import seaborn as sns
sales_data.groupby('Region').size().plot(kind='barh', color=sns.palettes.
  ↪mpl_palette('Dark2'))
plt.gca().spines[['top', 'right',]].set_visible(False)
```

```
[134]: np.histogram(sales_data['Total Cost'],bins=10)
```

```
[134]: (array([53, 13,  6,  5,  9,  8,  1,  2,  0,  3]),
        array([3.61224000e+03, 4.54230412e+05, 9.04848584e+05, 1.35546676e+06,
               1.80608493e+06, 2.25670310e+06, 2.70732127e+06, 3.15793944e+06,
               3.60855762e+06, 4.05917579e+06, 4.50979396e+06]))
```

```
[135]: mp.hist(sales_data['Total Cost'],bins=20)
```

```
[135]: (array([33., 20.,  7.,  6.,  3.,  3.,  2.,  3.,  5.,  4.,  3.,  5.,  1.,
               0.,  0.,  2.,  0.,  0.,  1.,  2.]),
        array([3.61224000e+03, 2.28921326e+05, 4.54230412e+05, 6.79539498e+05,
               9.04848584e+05, 1.13015767e+06, 1.35546676e+06, 1.58077584e+06,
               1.80608493e+06, 2.03139401e+06, 2.25670310e+06, 2.48201219e+06,
               2.70732127e+06, 2.93263036e+06, 3.15793944e+06, 3.38324853e+06,
               3.60855762e+06, 3.83386670e+06, 4.05917579e+06, 4.28448487e+06,
               4.50979396e+06]),
        <BarContainer object of 20 artists>)
```
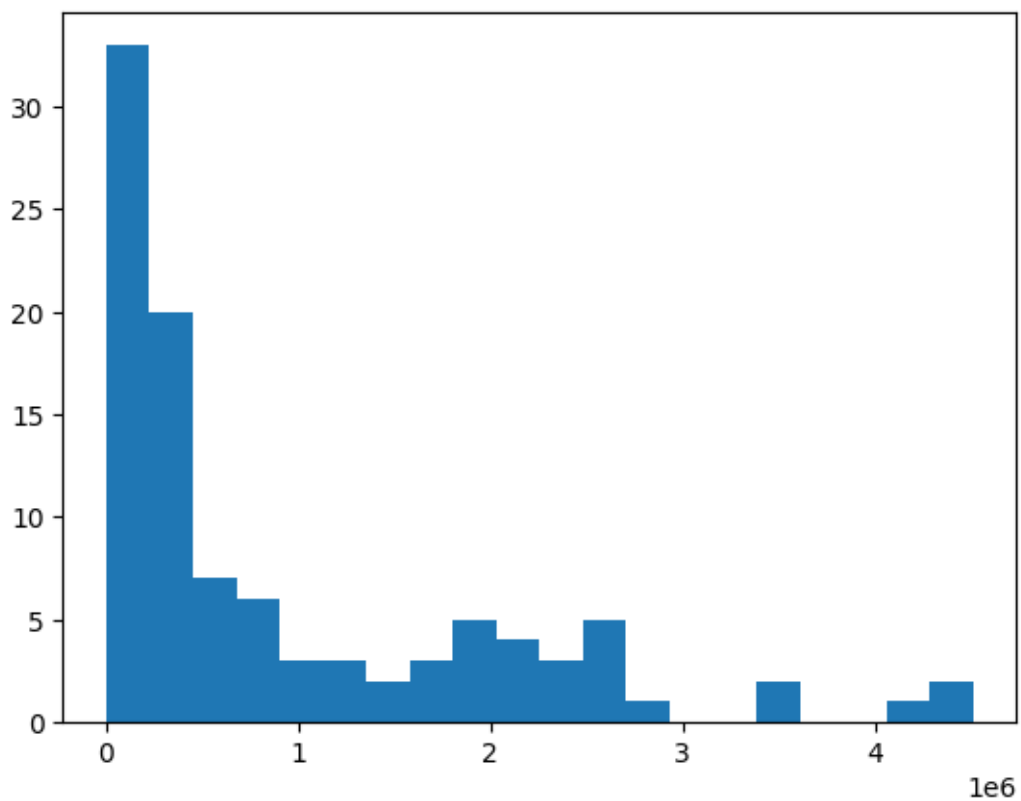
```
[136]: mp.scatter(sales_data['Units Sold'],sales_data['Unit Cost'])
```
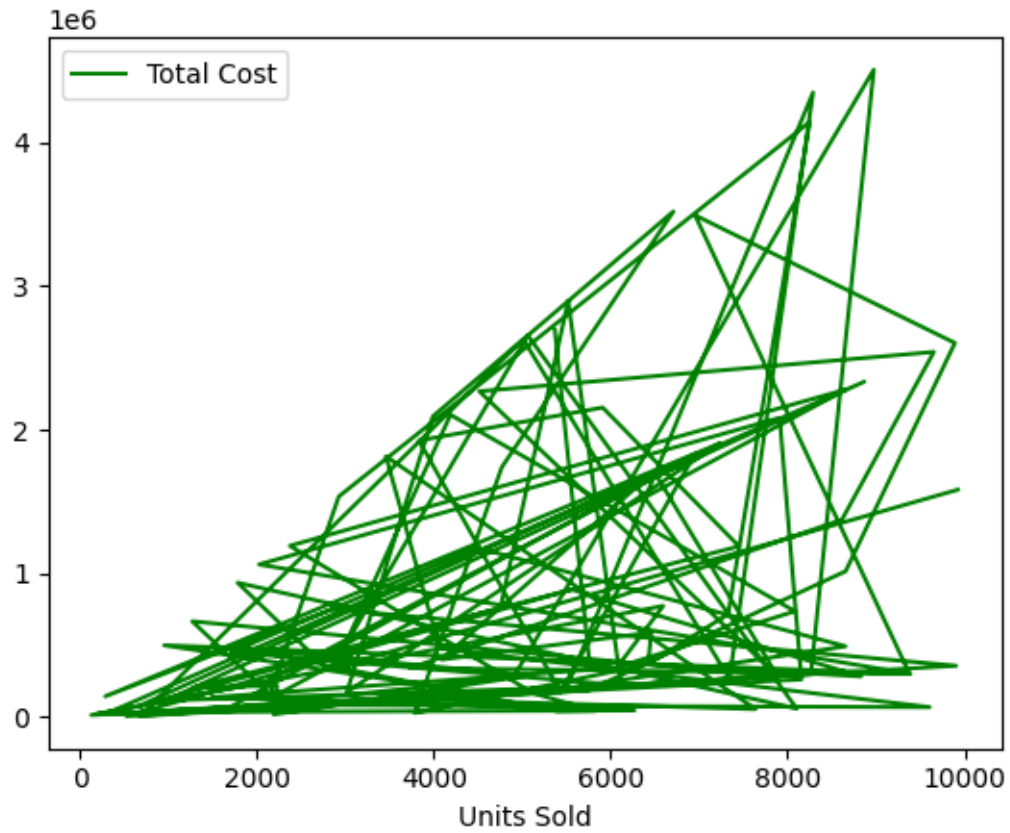
```
[136]: <matplotlib.collections.PathCollection at 0x78950a12ba30>
```
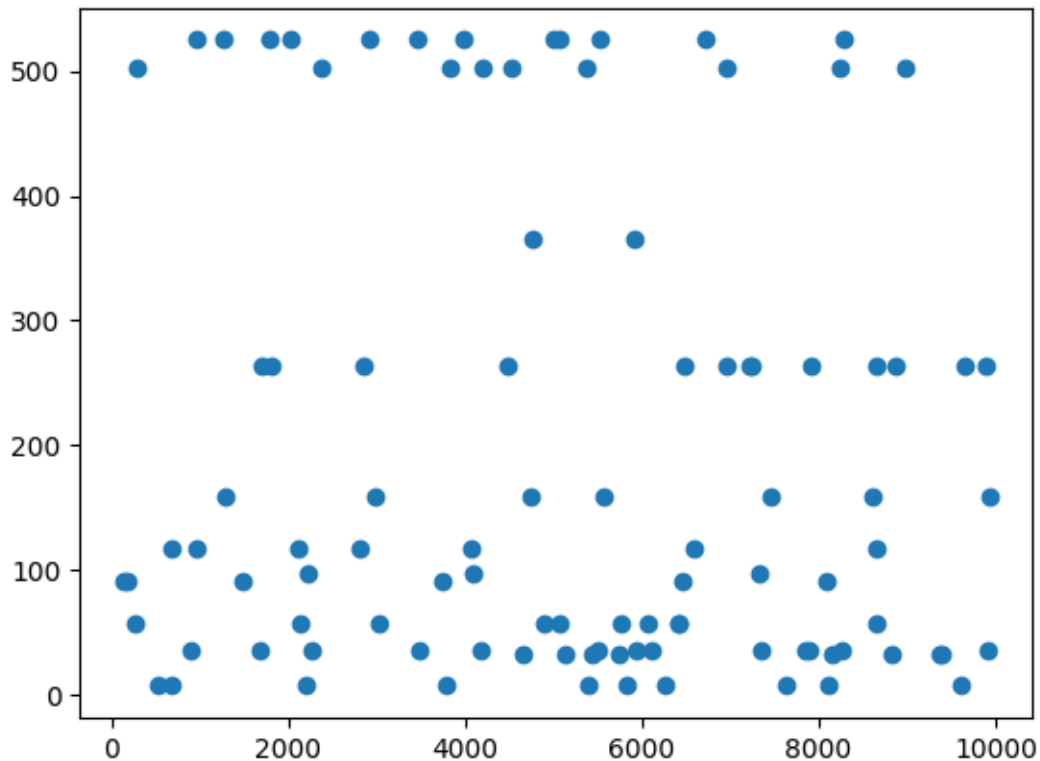
2 The above scatter plot implies that the two variables 'Units Sold' and 'Unit Cost' are inversely proportional to each other to some extent. When more units of a product are sold, the unit cost of that product becomes lesser and vice versa.
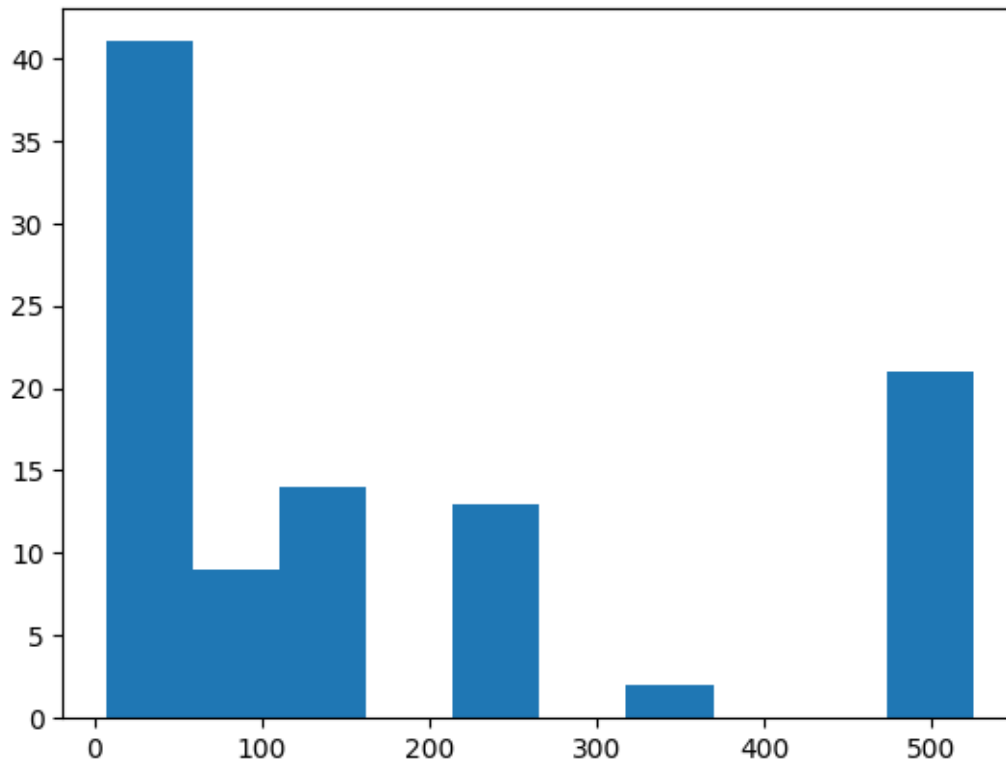
```
[137]: sales_data.plot.line(x='Units Sold',y='Total Cost',subplots=True,color={'Total␣
       ↪Cost': 'green'})
```

```
[137]: array([<Axes: xlabel='Units Sold'>], dtype=object)
```

```
[138]: mp.hist(sales_data['Unit Cost'])
```

```
[138]: (array([41.,  9., 14.,  0., 13.,  0.,  2.,  0.,  0., 21.]),
        array([  6.92 ,  58.724, 110.528, 162.332, 214.136, 265.94 , 317.744,
               369.548, 421.352, 473.156, 524.96 ]),
        <BarContainer object of 10 artists>)
```

[139]: `np.min(sales_data['Unit Cost'])`

[139]: 6.92

[140]: `np.max(sales_data['Unit Cost'])`

[140]: 524.96

[141]: `np.mean(sales_data['Unit Cost'])`

[141]: 191.048

[142]: `np.std(sales_data['Unit Cost'])`

[142]: 187.26477590299785

[143]: `np.var(sales_data['Unit Cost'])`

[143]: 35068.096294

[144]: `np.median(sales_data['Unit Cost'])`

107.275

Maximum and minimum unit costs are 6.92 and 524.96 respectively. Average unit cost of a product is 191.05. The Unit Cost variable varies considerably throughout it's distribution. The median cost of a unit stands at 107.28

```
area_plot = sales_data.plot.area(x='Unit Price',y='Total␣
 ↪Profit',color='violet',stacked=True,legend=None)
mp.ylabel('Total Profit')
```

Text(0, 0.5, 'Total Profit')
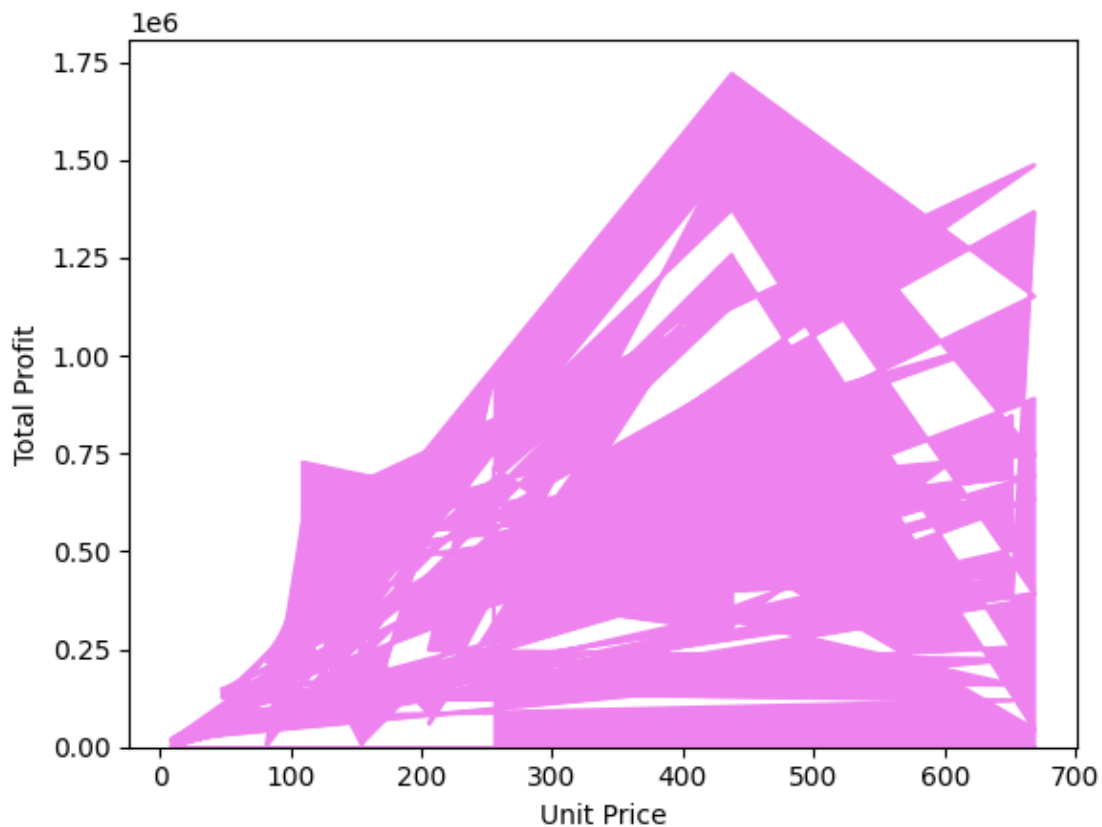
Maximum profit has been generated in the unit price range of 400- 500

```
sales_data.plot.area(x='Units Sold',y='Total Profit',color='yellow',legend=None)
mp.ylabel('Total Profit')
```

Text(0, 0.5, 'Total Profit')
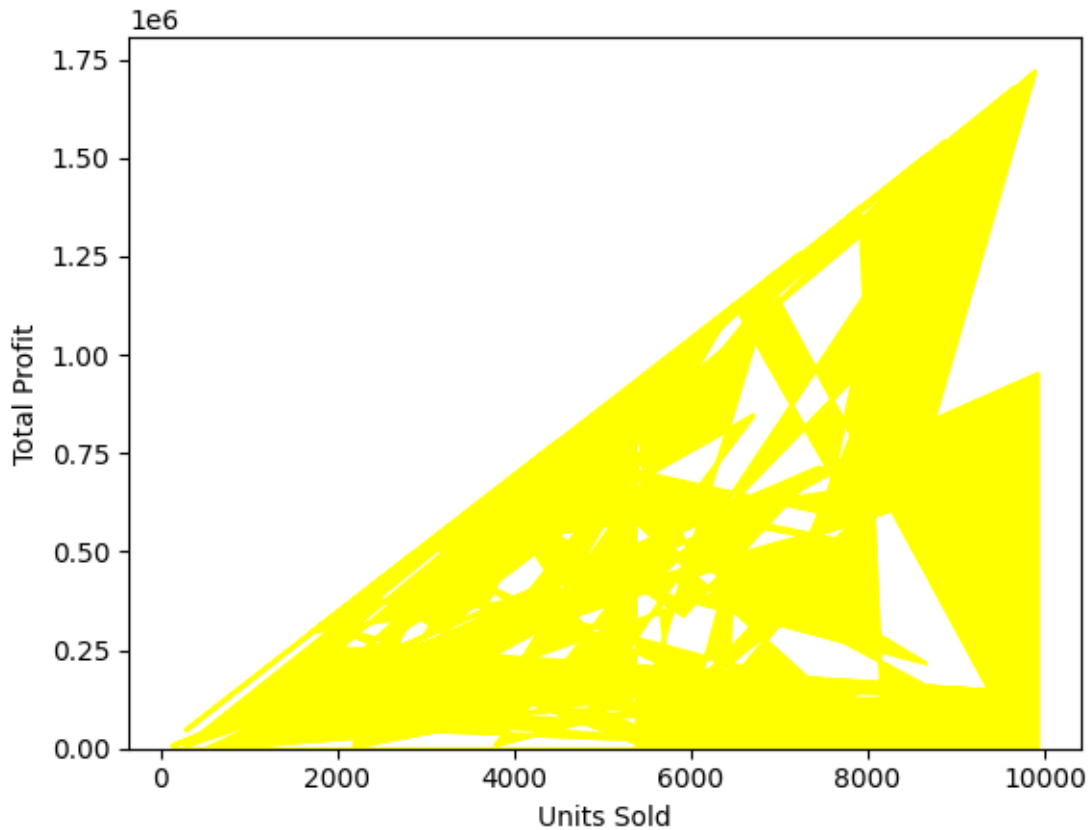
2.0.1 Maximum profit has been generated when the number of units sold were between 8000 and 10000 i.e. more the number of units sold, more will be the profit generated.

```
[147]: sales_data.plot.area(x='Units Sold',y='Total Cost',color='aqua',legend=None)
       mp.ylabel('Total Cost')
```

```
[147]: Text(0, 0.5, 'Total Cost')
```
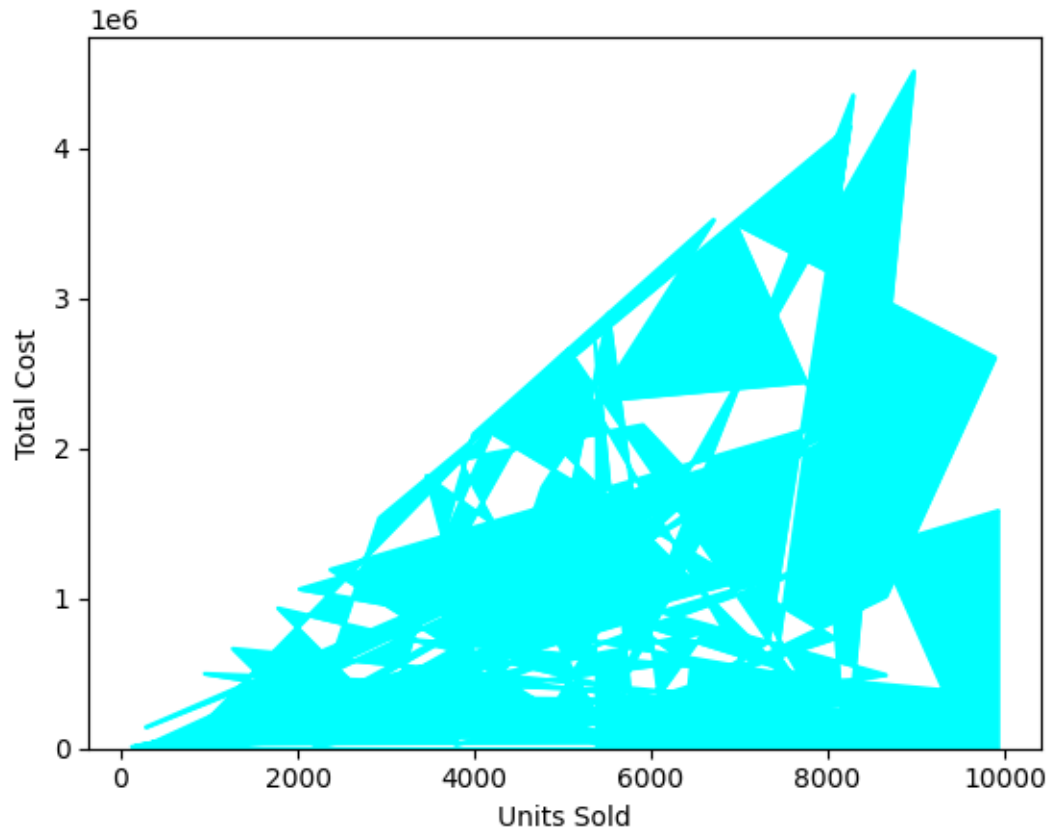
Maximum cost has been generated when 8000-9000 units were sold

```
[148]: sales_data.plot.area(x='Units Sold',y='Total Revenue',color='grey',legend=None)
       mp.ylabel('Total Revenue')
```

[148]: Text(0, 0.5, 'Total Revenue')
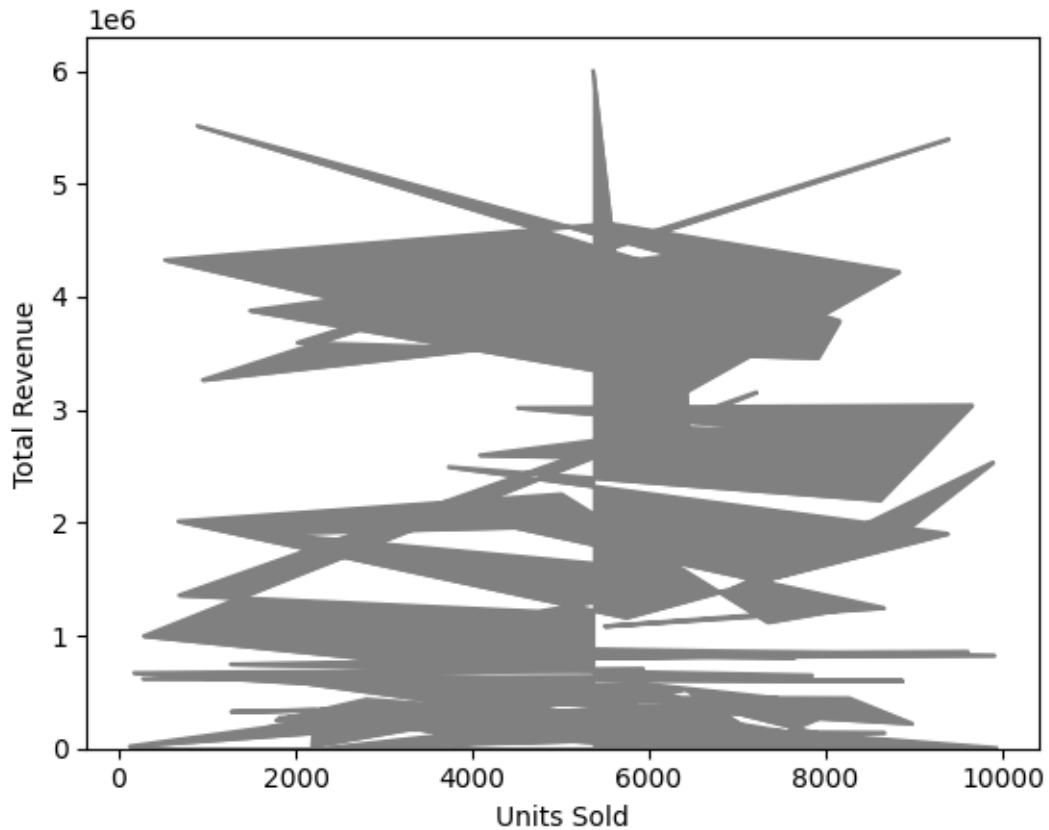
2.0.2 Maximum revenue has been generated when 5000-6500 units of a product were sold.

```
[149]: bar_plot = sales_data.plot.bar(x='Units Sold',y=['Total Revenue','Total␣
       ↪Cost','Total Profit'],color=['orange','red','blue'],stacked=True,rot=True)
       mp.xticks(rotation=90)
       mp.locator_params(nbins=38)
       mp.tick_params(axis='y', which='major', labelsize=7)
```
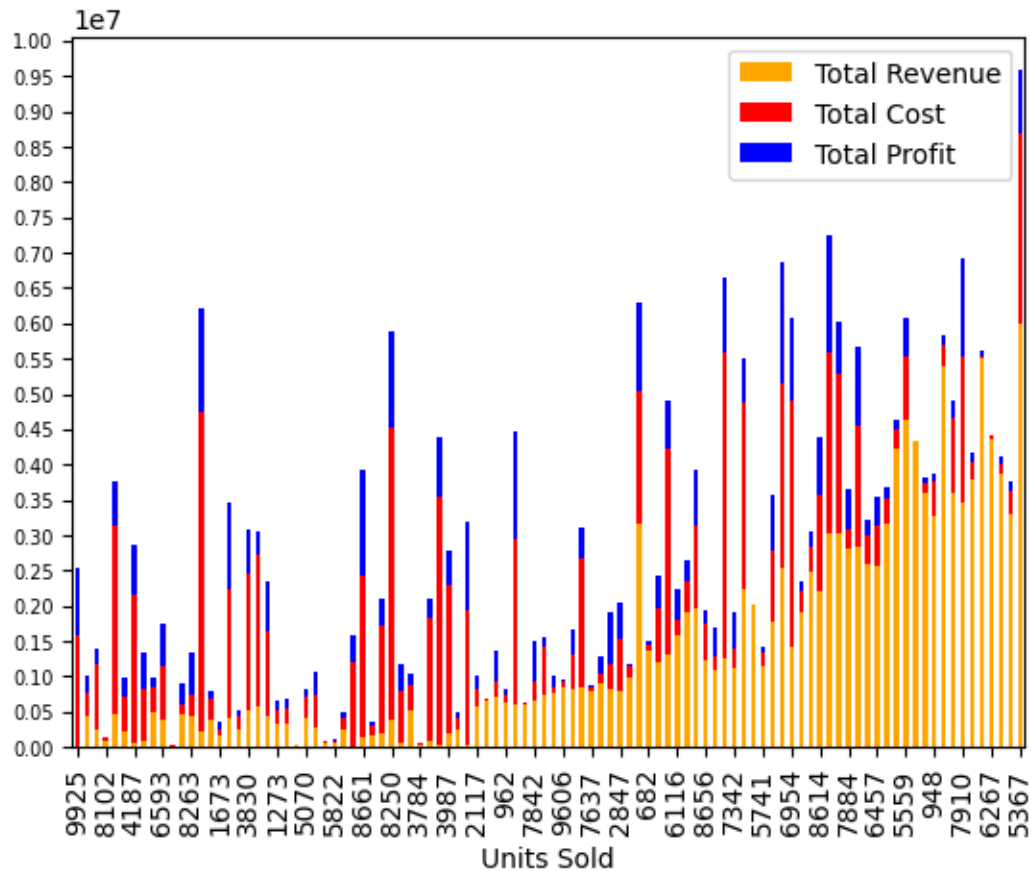
```
[150]: sales_data.head()
```

```
[150]:            Order Date Order Priority  Ship Date       Item Type  \
       Order ID
       669165933  5/28/2010              H  2010-06-27        Baby Food
       963881480  8/22/2012              C  2012-09-15           Cereal
       341417157   5/2/2014              L  2014-05-08  Office Supplies
       514321792  6/20/2014              C  2014-07-05           Fruits
       115456712   2/1/2013              L  2013-02-06  Office Supplies


                                         Region            Country  \
       Order ID
       669165933            Australia and Oceania              Tuvalu
       963881480  Central America and the Caribbean            Grenada
       341417157                           Europe             Russia
       514321792               Sub-Saharan Africa  Sao Tome and Principe
       115456712               Sub-Saharan Africa             Rwanda


                 Sales Channel  Units Sold  Unit Price  Unit Cost  Total Revenue  \
```

```
          Order ID
          669165933          Offline          9925          255.28          159.42          4870.26
          963881480          Online           2804          205.70          117.11        435466.90
          341417157          Offline          1779          651.21          524.96        247956.32
          514321792          Online           8102            9.33            6.92         75591.66
          115456712          Offline          5062          651.21          524.96        471336.91

                      Total Cost  Total Profit  Ship Month
          Order ID
          669165933  1582243.50      951410.50           6
          963881480   328376.44      248406.36           9
          341417157   933903.84      224598.75           5
          514321792    56065.84       19525.82           7
          115456712  2657347.52      639077.50           2
```

[151]:
```python
sales_data.plot.barh(x='Item Type',y='Total Revenue',color='green')
mp.locator_params(nbins=28)
mp.xlabel('Total Revenue')
```

[151]: Text(0.5, 0, 'Total Revenue')
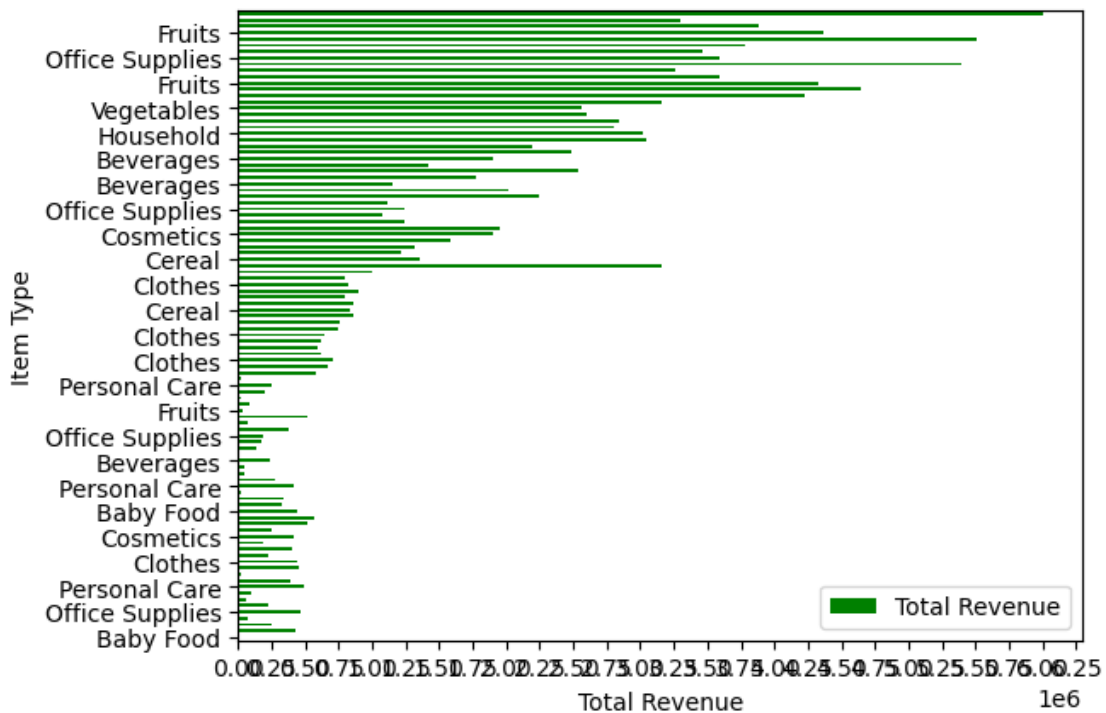
[152]:
```python
sales_data['Item Type'].unique()
```

20

```
[152]: array(['Baby Food', 'Cereal', 'Office Supplies', 'Fruits', 'Household',
              'Vegetables', 'Personal Care', 'Clothes', 'Cosmetics', 'Beverages',
              'Meat', 'Snacks'], dtype=object)
```

```
[153]: items = ['Baby Food', 'Cereal', 'Office Supplies', 'Fruits', 'Household',
                'Vegetables', 'Personal Care', 'Clothes', 'Cosmetics', 'Beverages',
                'Meat', 'Snacks']
```

```
[154]: sales_data['Item Type'] = pd.Categorical(sales_data['Item␣
        ↪Type'],categories=items,ordered=True)
```

```
[155]: sales_data.groupby('Item Type')['Total Revenue'].count().
        ↪plot(kind='bar',color=['green','orange','blue','red','purple','pink','yellow','brown','aqua
```

```
[155]: <Axes: xlabel='Item Type'>
```
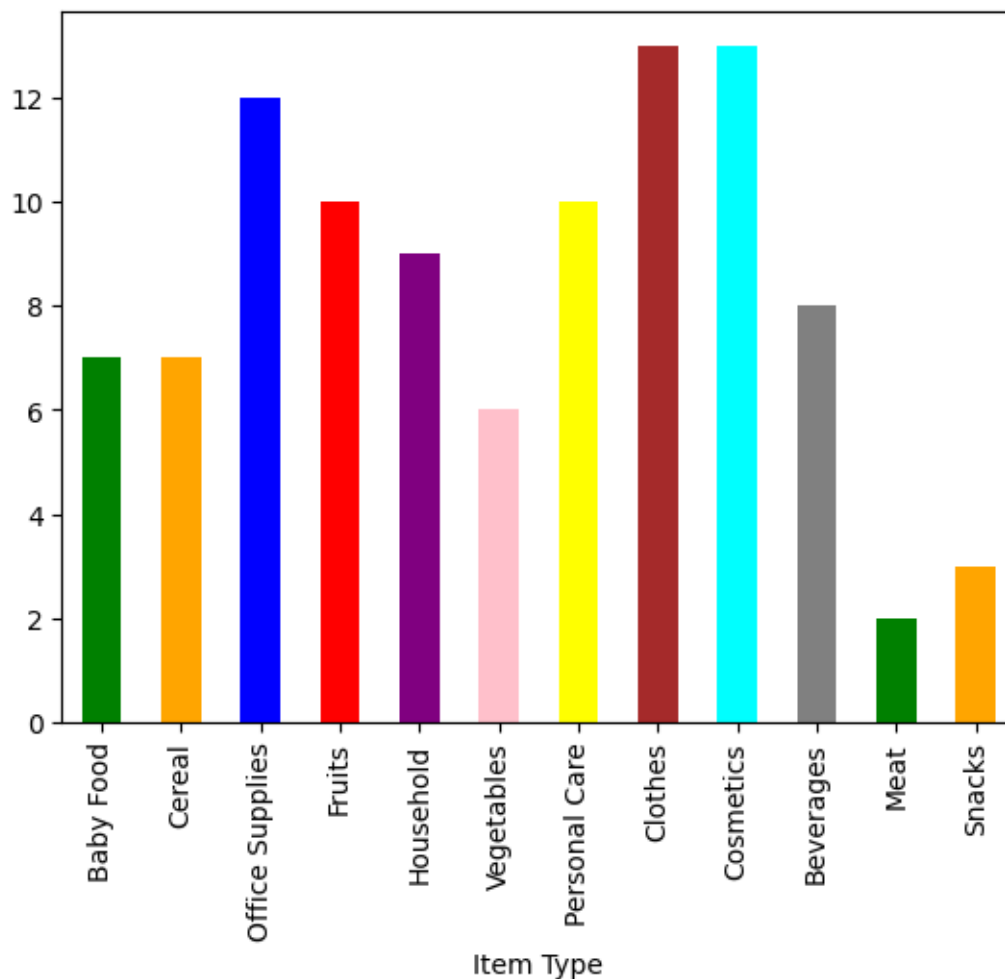
```
[156]: pd.pivot_table(sales_data,values='Total Revenue',index='Item␣
       ↪Type',aggfunc='count').plot(kind='bar')
```

[156]: `<Axes: xlabel='Item Type'>`
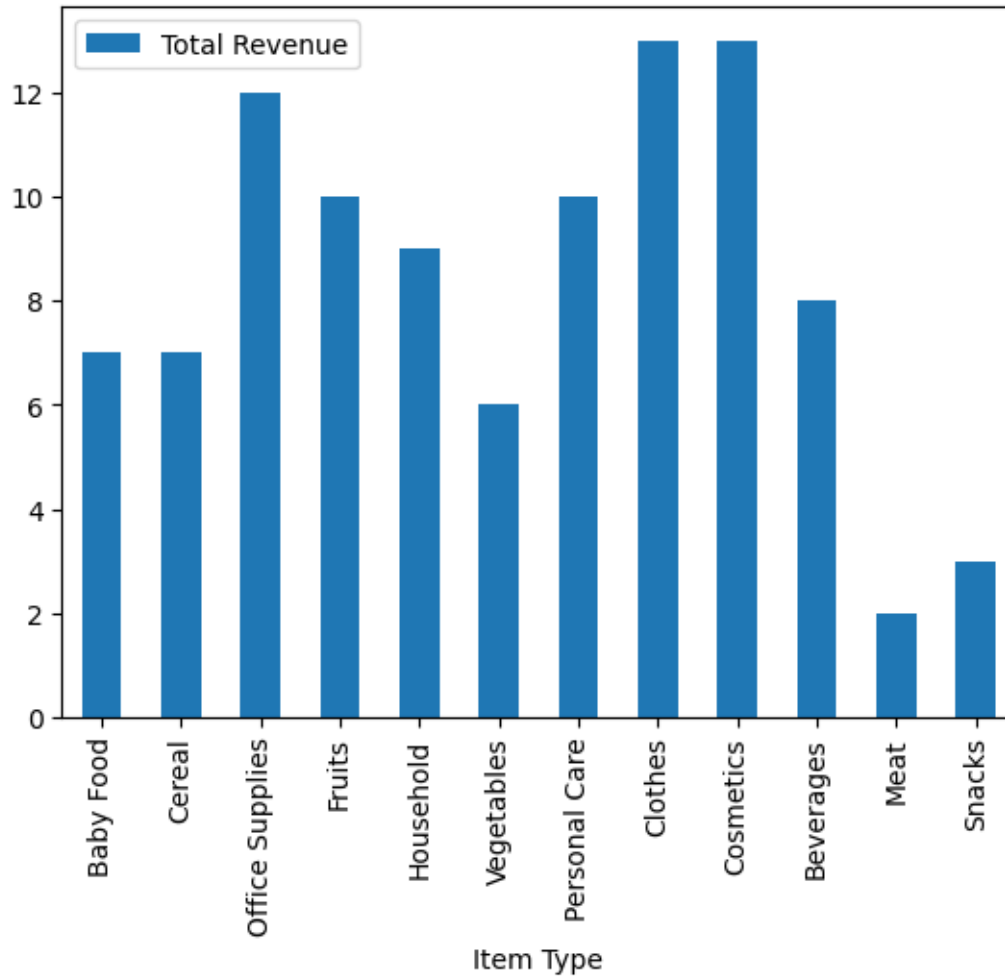
Maximum revenue has been generated from the items 'Clothes' and 'Cosmetics' closely followed by
'Office Supplies

```
[157]: #pd.pivot_table(sales_data,values=['Total Revenue','Total Cost','Total␣
       ↪Profit'],index='Units Sold',aggfunc='count').
       ↪plot(kind='kde',color=['green','orange','red'],stacked=True)
       pd.pivot_table(sales_data,index='Units Sold',aggfunc='count').
       ↪plot(kind='kde',stacked=True,legend=None)
       mp.xlabel('Units Sold')
```

[157]: `Text(0.5, 0, 'Units Sold')`
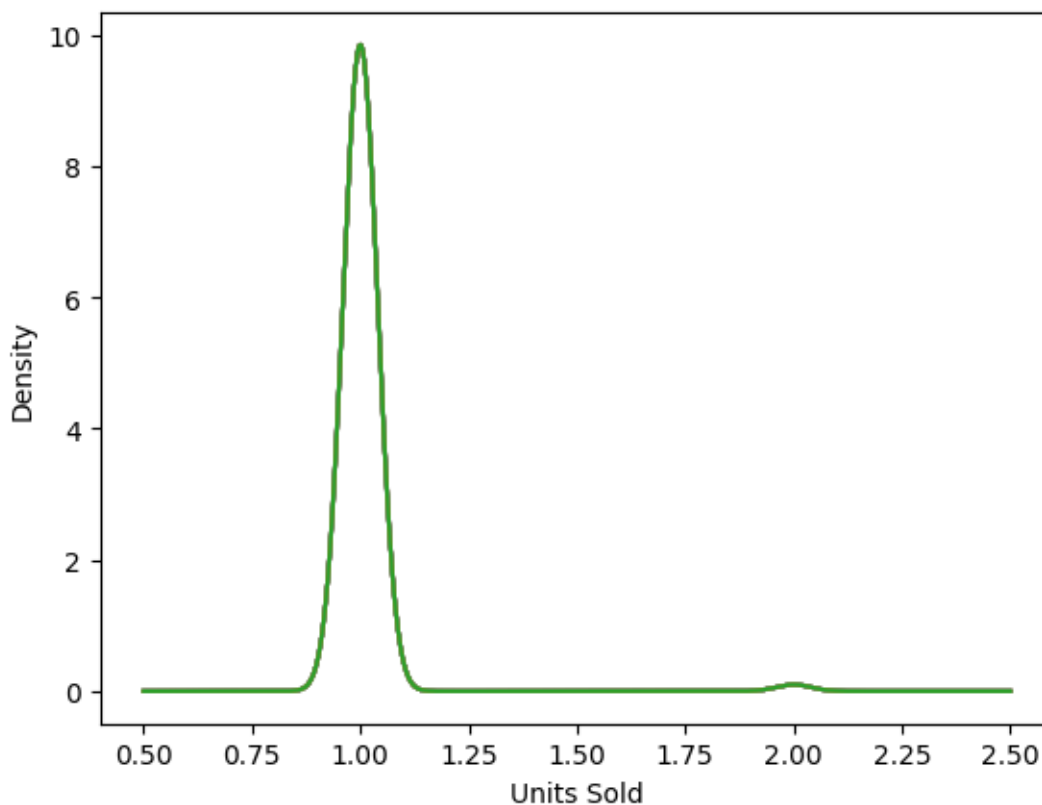
```
[158]: sales_data.head()
```

```
[158]:            Order Date Order Priority  Ship Date      Item Type  \
       Order ID
       669165933  5/28/2010              H 2010-06-27       Baby Food
       963881480  8/22/2012              C 2012-09-15          Cereal
       341417157   5/2/2014              L 2014-05-08  Office Supplies
       514321792  6/20/2014              C 2014-07-05          Fruits
       115456712   2/1/2013              L 2013-02-06  Office Supplies


                                          Region              Country  \
       Order ID
       669165933              Australia and Oceania               Tuvalu
       963881480  Central America and the Caribbean              Grenada
       341417157                            Europe               Russia
       514321792                Sub-Saharan Africa  Sao Tome and Principe
       115456712                Sub-Saharan Africa               Rwanda


                 Sales Channel  Units Sold  Unit Price  Unit Cost  Total Revenue  \
       Order ID
       669165933       Offline        9925      255.28     159.42        4870.26
```

```
             963881480        Online        2804       205.70       117.11       435466.90
             341417157       Offline        1779       651.21       524.96       247956.32
             514321792        Online        8102         9.33         6.92        75591.66
             115456712       Offline        5062       651.21       524.96       471336.91


                        Total Cost  Total Profit  Ship Month
            Order ID
            669165933  1582243.50      951410.50           6
            963881480   328376.44      248406.36           9
            341417157   933903.84      224598.75           5
            514321792    56065.84       19525.82           7
            115456712  2657347.52      639077.50           2
```

[159]: `sales_data.info()`

```
<class 'pandas.core.frame.DataFrame'>
Index: 100 entries, 669165933 to 665095412
Data columns (total 14 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   Order Date      100 non-null    object
 1   Order Priority  100 non-null    object
 2   Ship Date       100 non-null    datetime64[ns]
 3   Item Type       100 non-null    category
 4   Region          100 non-null    object
 5   Country         100 non-null    object
 6   Sales Channel   100 non-null    object
 7   Units Sold      100 non-null    int64
 8   Unit Price      100 non-null    float64
 9   Unit Cost       100 non-null    float64
 10  Total Revenue   100 non-null    float64
 11  Total Cost      100 non-null    float64
 12  Total Profit    100 non-null    float64
 13  Ship Month      100 non-null    int32
dtypes: category(1), datetime64[ns](1), float64(5), int32(1), int64(1),
object(5)
memory usage: 11.0+ KB
```

[163]: `sales_data['Region'].unique()`

[163]: array(['Australia and Oceania', 'Central America and the Caribbean',
        'Europe', 'Sub-Saharan Africa', 'Asia',
        'Middle East and North Africa', 'North America'], dtype=object)

[164]:
```
regions = ['Australia and Oceania', 'Central America and the Caribbean',
        'Europe', 'Sub-Saharan Africa', 'Asia',
        'Middle East and North Africa', 'North America']
```

```
[165]: sales_data['Region'] = pd.Categorical(sales_data['Region'],categories =
        ↪regions,ordered=True)
```

```
[166]: sales_data.head()
```

```
[166]:            Order Date Order Priority  Ship Date       Item Type  \
       Order ID
       669165933  5/28/2010              H 2010-06-27       Baby Food
       963881480  8/22/2012              C 2012-09-15          Cereal
       341417157   5/2/2014              L 2014-05-08  Office Supplies
       514321792  6/20/2014              C 2014-07-05          Fruits
       115456712   2/1/2013              L 2013-02-06  Office Supplies


                                       Region              Country  \
       Order ID
       669165933              Australia and Oceania               Tuvalu
       963881480  Central America and the Caribbean              Grenada
       341417157                             Europe               Russia
       514321792                 Sub-Saharan Africa  Sao Tome and Principe
       115456712                 Sub-Saharan Africa               Rwanda


                 Sales Channel  Units Sold  Unit Price  Unit Cost  Total Revenue  \
       Order ID
       669165933       Offline        9925      255.28     159.42        4870.26
       963881480        Online        2804      205.70     117.11       435466.90
       341417157       Offline        1779      651.21     524.96       247956.32
       514321792        Online        8102        9.33       6.92        75591.66
       115456712       Offline        5062      651.21     524.96       471336.91


                   Total Cost  Total Profit  Ship Month
       Order ID
       669165933  1582243.50     951410.50           6
       963881480   328376.44     248406.36           9
       341417157   933903.84     224598.75           5
       514321792    56065.84      19525.82           7
       115456712  2657347.52     639077.50           2
```

```
[167]: mp.scatter(sales_data['Units Sold'],sales_data['Total Cost'])
       mp.xlabel('Units Sold')
       mp.ylabel('Total Cost')
```

```
[167]: Text(0, 0.5, 'Total Cost')
```
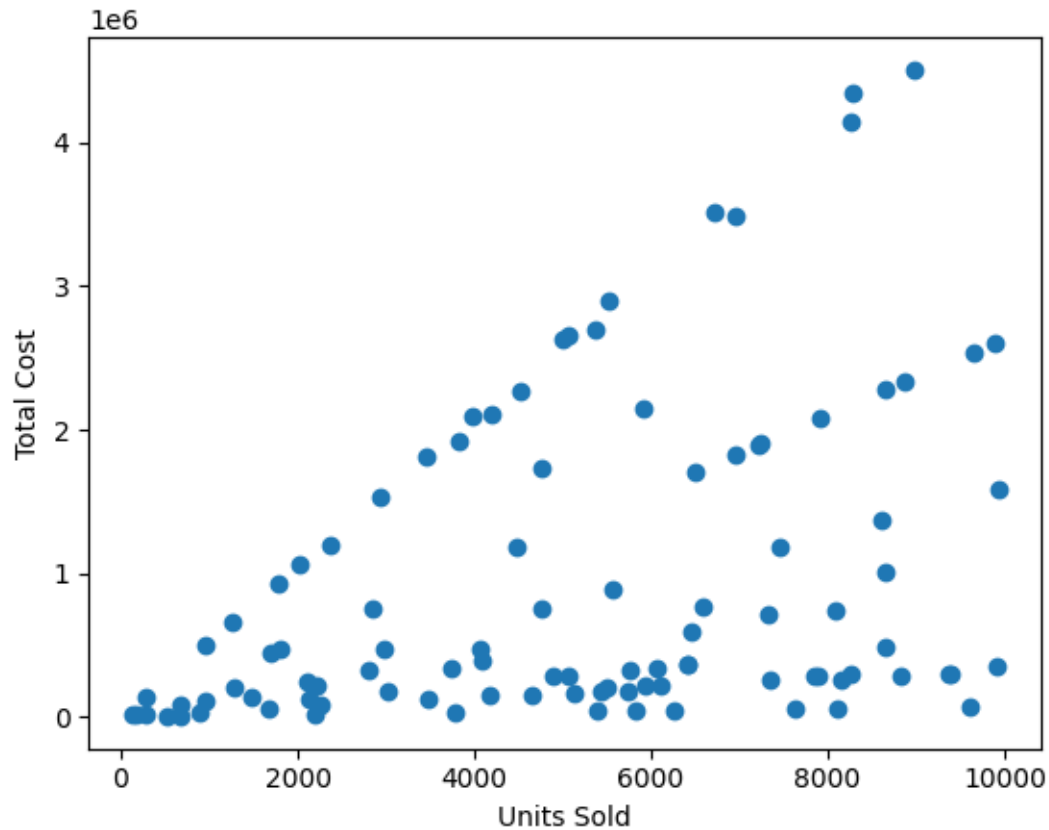
More the number of units sold of a product, more will be the total cost associated with it

```
[168]:  sn.barplot(x='Region',y='Total Cost',data=sales_data)
        mp.xticks(rotation=90)
```

```
[168]:  ([0, 1, 2, 3, 4, 5, 6],
         [Text(0, 0, 'Australia and Oceania'),
          Text(1, 0, 'Central America and the Caribbean'),
          Text(2, 0, 'Europe'),
          Text(3, 0, 'Sub-Saharan Africa'),
          Text(4, 0, 'Asia'),
          Text(5, 0, 'Middle East and North Africa'),
          Text(6, 0, 'North America')])
```
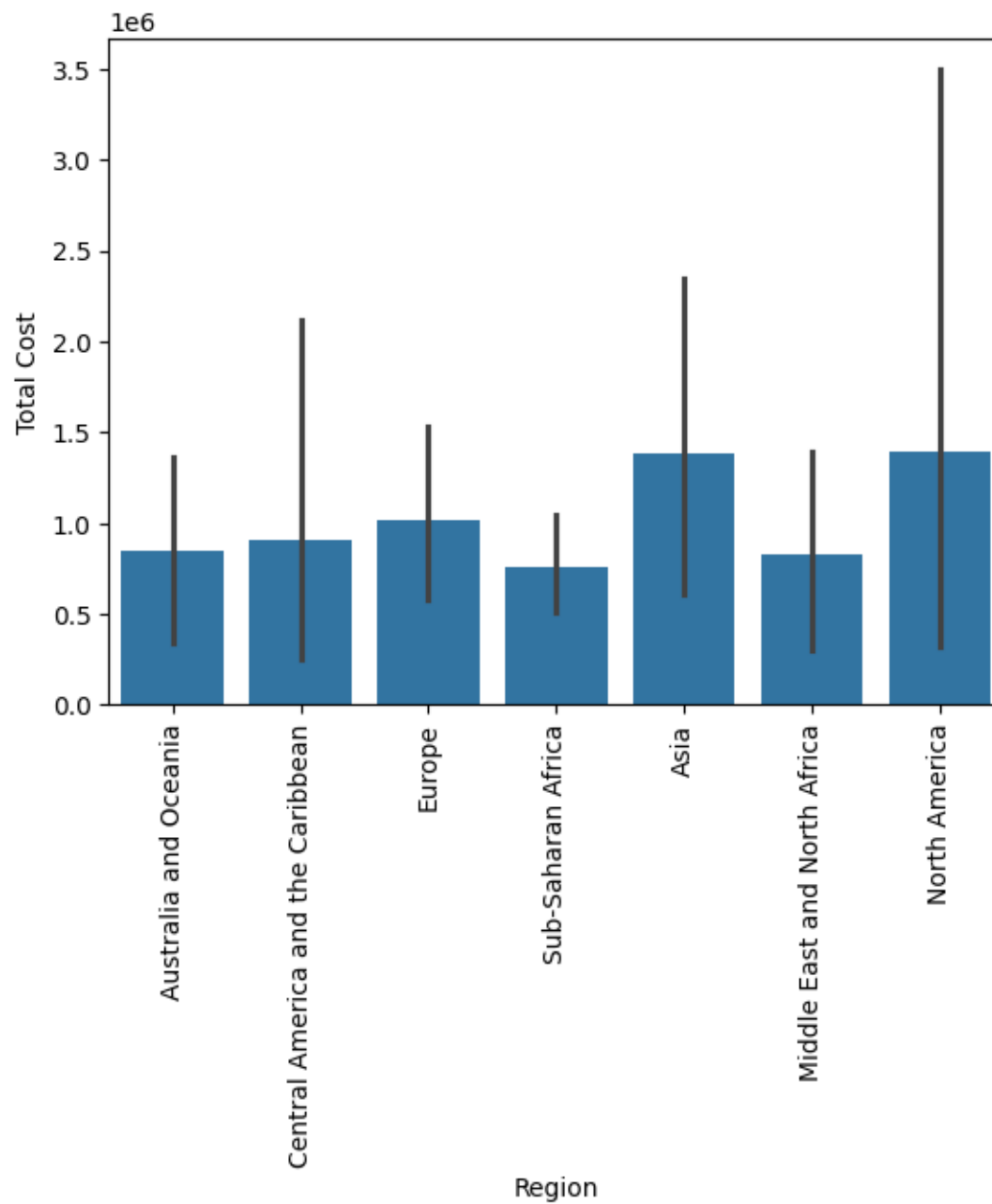
2.0.3 Cost of items is maximum in Asia and North America, and minimum in Sub-Saharan Africa

```
[169]: sn.pairplot(sales_data)
```

```
[169]: <seaborn.axisgrid.PairGrid at 0x789509849000>
```
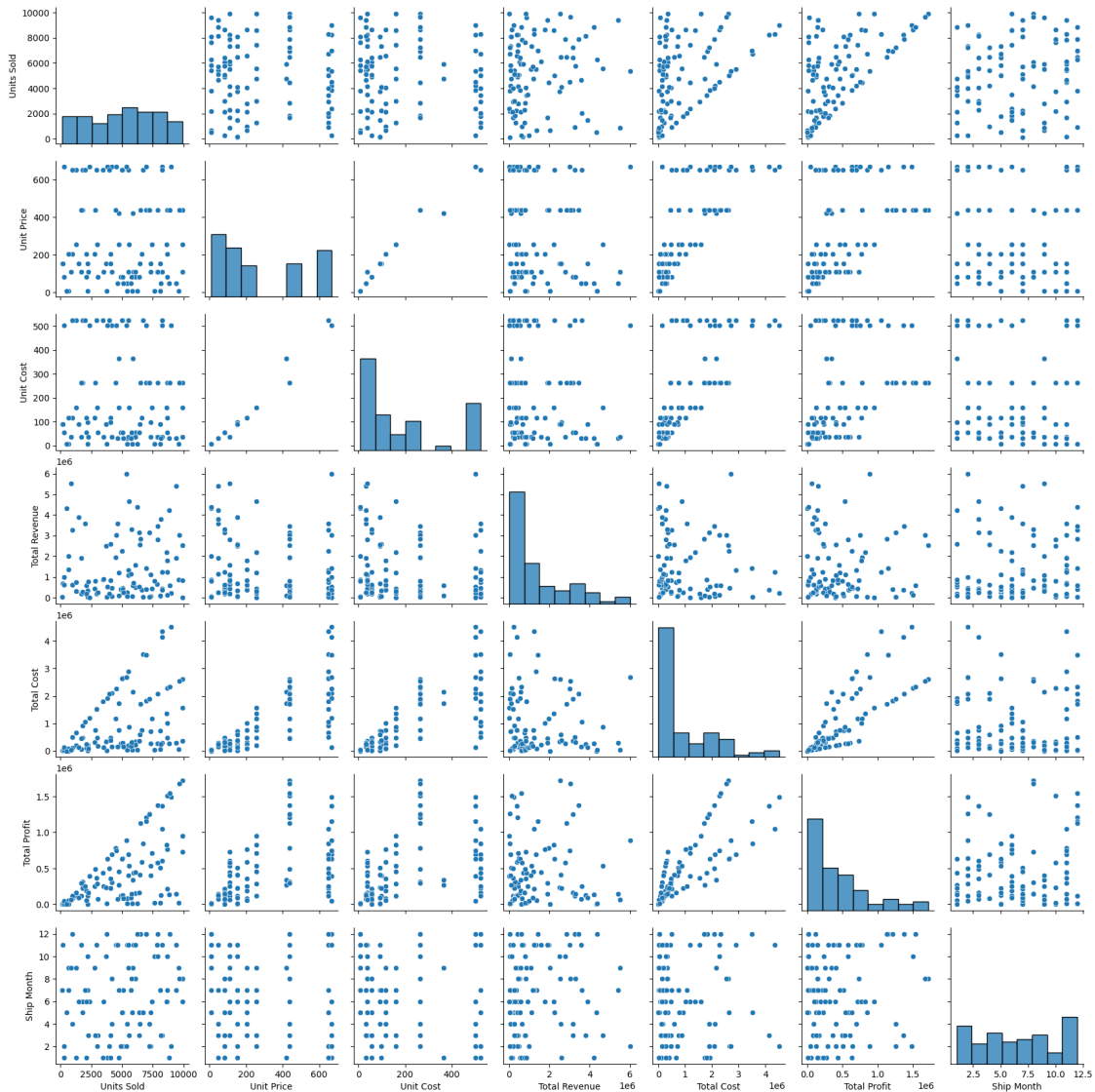
```
[173]: mp.figure(figsize=(10,4))
       sn.lineplot(x='Region',y='Total Cost',data=sales_data,ci=None)
       mp.xticks(rotation=45)
```

<ipython-input-173-8ea0a5d66eef>:2: FutureWarning:

The `ci` parameter is deprecated. Use `errorbar=None` for the same effect.

  sn.lineplot(x='Region',y='Total Cost',data=sales_data,ci=None)

```
[173]: ([0, 1, 2, 3, 4, 5, 6],
        [Text(0, 0, 'Australia and Oceania'),
         Text(1, 0, 'Central America and the Caribbean'),
```

```
            Text(2, 0, 'Europe'),
            Text(3, 0, 'Sub-Saharan Africa'),
            Text(4, 0, 'Asia'),
            Text(5, 0, 'Middle East and North Africa'),
            Text(6, 0, 'North America')])
```
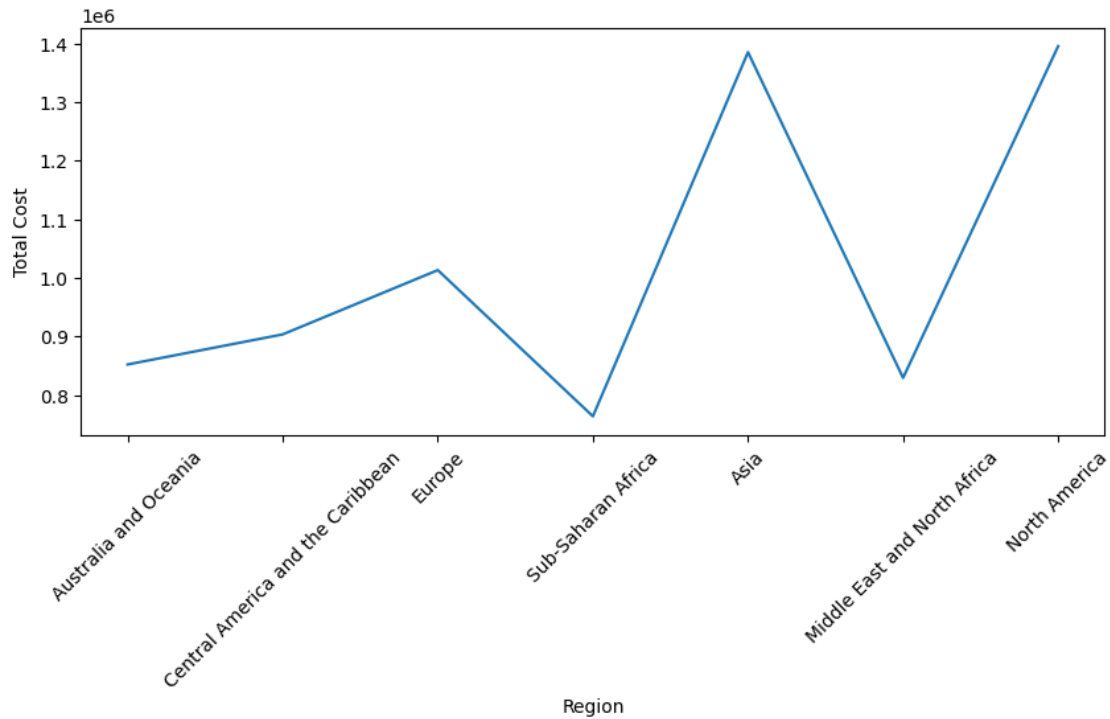

Products have been much more expensive in Asia and North America in comparison to other continents

```
[174]: sales_data.head()
```

```
[174]:             Order Date Order Priority  Ship Date        Item Type  \
        Order ID
        669165933 2010-05-28              H 2010-06-27        Baby Food
        963881480 2012-08-22              C 2012-09-15           Cereal
        341417157 2014-05-02              L 2014-05-08  Office Supplies
        514321792 2014-06-20              C 2014-07-05           Fruits
        115456712 2013-02-01              L 2013-02-06  Office Supplies


                                        Region          Country  \
        Order ID
        669165933              Australia and Oceania            Tuvalu
        963881480  Central America and the Caribbean           Grenada
        341417157                             Europe            Russia
```

```
514321792              Sub-Saharan Africa  Sao Tome and Principe
115456712              Sub-Saharan Africa                 Rwanda

          Sales Channel  Units Sold  Unit Price  Unit Cost  Total Revenue  \
Order ID
669165933       Offline        9925      255.28     159.42        4870.26
963881480        Online        2804      205.70     117.11      435466.90
341417157       Offline        1779      651.21     524.96      247956.32
514321792        Online        8102        9.33       6.92       75591.66
115456712       Offline        5062      651.21     524.96      471336.91

          Total Cost  Total Profit  Ship Month
Order ID
669165933  1582243.50     951410.50           6
963881480   328376.44     248406.36           9
341417157   933903.84     224598.75           5
514321792    56065.84      19525.82           7
115456712  2657347.52     639077.50           2
```

[175]:
```python
sales_data['Country'].unique()
```

[175]:
```
array(['Tuvalu', 'Grenada', 'Russia', 'Sao Tome and Principe', 'Rwanda',
       'Solomon Islands', 'Angola', 'Burkina Faso',
       'Republic of the Congo', 'Senegal', 'Kyrgyzstan', 'Cape Verde',
       'Bangladesh', 'Honduras', 'Mongolia', 'Bulgaria', 'Sri Lanka',
       'Cameroon', 'Turkmenistan', 'East Timor', 'Norway', 'Portugal',
       'New Zealand', 'Moldova ', 'France', 'Kiribati', 'Mali',
       'The Gambia', 'Switzerland', 'South Sudan', 'Australia', 'Myanmar',
       'Djibouti', 'Costa Rica', 'Syria', 'Brunei', 'Niger', 'Azerbaijan',
       'Slovakia', 'Comoros', 'Iceland', 'Macedonia', 'Mauritania',
       'Albania', 'Lesotho', 'Saudi Arabia', 'Sierra Leone',
       "Cote d'Ivoire", 'Fiji', 'Austria', 'United Kingdom', 'San Marino',
       'Libya', 'Haiti', 'Gabon', 'Belize', 'Lithuania', 'Madagascar',
       'Democratic Republic of the Congo', 'Pakistan', 'Mexico',
       'Federated States of Micronesia', 'Laos', 'Monaco', 'Samoa ',
       'Spain', 'Lebanon', 'Iran', 'Zambia', 'Kenya', 'Kuwait',
       'Slovenia', 'Romania', 'Nicaragua', 'Malaysia', 'Mozambique'],
      dtype=object)
```

[176]:
```python
countries = ['Tuvalu', 'Grenada', 'Russia', 'Sao Tome and Principe', 'Rwanda',
       'Solomon Islands', 'Angola', 'Burkina Faso',
       'Republic of the Congo', 'Senegal', 'Kyrgyzstan', 'Cape Verde',
       'Bangladesh', 'Honduras', 'Mongolia', 'Bulgaria', 'Sri Lanka',
       'Cameroon', 'Turkmenistan', 'East Timor', 'Norway', 'Portugal',
       'New Zealand', 'Moldova ', 'France', 'Kiribati', 'Mali',
       'The Gambia', 'Switzerland', 'South Sudan', 'Australia', 'Myanmar',
       'Djibouti', 'Costa Rica', 'Syria', 'Brunei', 'Niger', 'Azerbaijan',
```

```
              'Slovakia', 'Comoros', 'Iceland', 'Macedonia', 'Mauritania',
              'Albania', 'Lesotho', 'Saudi Arabia', 'Sierra Leone',
              "Cote d'Ivoire", 'Fiji', 'Austria', 'United Kingdom', 'San Marino',
              'Libya', 'Haiti', 'Gabon', 'Belize', 'Lithuania', 'Madagascar',
              'Democratic Republic of the Congo', 'Pakistan', 'Mexico',
              'Federated States of Micronesia', 'Laos', 'Monaco', 'Samoa ',
              'Spain', 'Lebanon', 'Iran', 'Zambia', 'Kenya', 'Kuwait',
              'Slovenia', 'Romania', 'Nicaragua', 'Malaysia', 'Mozambique']
```

[177]:
```python
sales_data['Country'] = pd.
 ↪Categorical(sales_data['Country'],categories=countries,ordered=True)
```

[178]:
```python
mp.figure(figsize=(15,6))
sn.barplot(x='Country', y='Total Revenue', data=sales_data, ci=None)
mp.xticks(rotation=90)
mp.tick_params(axis='x', which='major', labelsize=10)
```

<ipython-input-178-e54b80e7cacb>:2: FutureWarning:

The `ci` parameter is deprecated. Use `errorbar=None` for the same effect.

  sn.barplot(x='Country', y='Total Revenue', data=sales_data, ci=None)
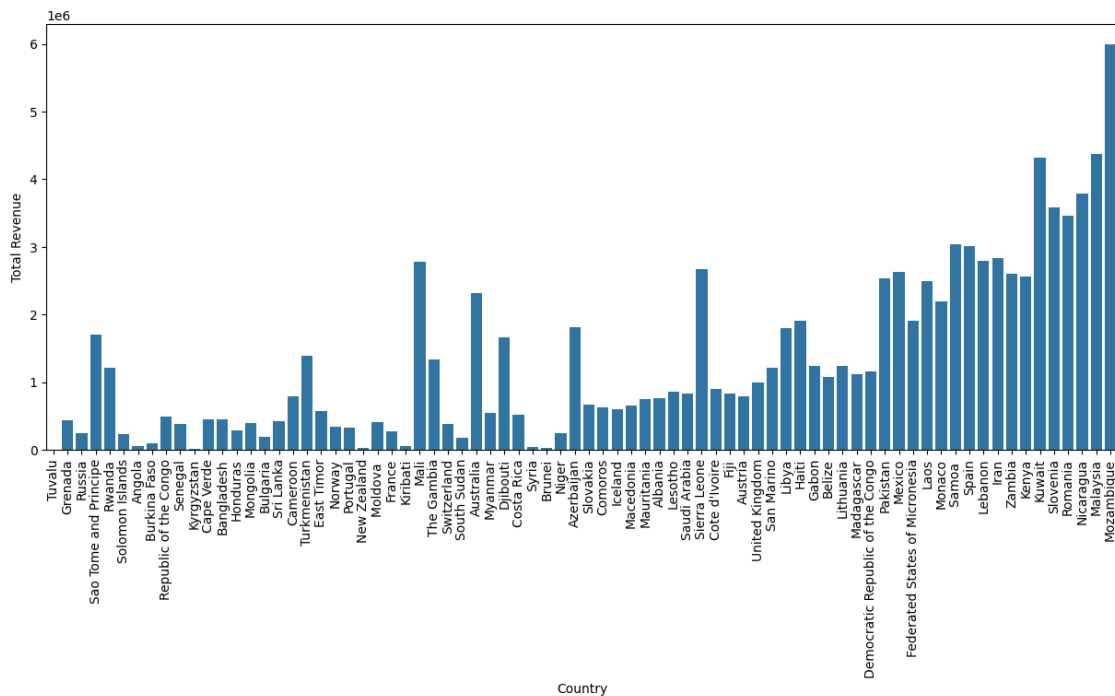
Mozambique is the country where maximum revenue has been generated followed by Kenya.

```
[179]: sales_data.head()
```

```
[179]:             Order Date Order Priority  Ship Date        Item Type  \
        Order ID
        669165933 2010-05-28              H 2010-06-27        Baby Food
        963881480 2012-08-22              C 2012-09-15           Cereal
        341417157 2014-05-02              L 2014-05-08  Office Supplies
        514321792 2014-06-20              C 2014-07-05           Fruits
        115456712 2013-02-01              L 2013-02-06  Office Supplies


                                       Region                  Country  \
        Order ID
        669165933           Australia and Oceania                 Tuvalu
        963881480  Central America and the Caribbean             Grenada
        341417157                          Europe                 Russia
        514321792              Sub-Saharan Africa  Sao Tome and Principe
        115456712              Sub-Saharan Africa                 Rwanda


                  Sales Channel  Units Sold  Unit Price  Unit Cost  Total Revenue  \
        Order ID
        669165933       Offline        9925      255.28     159.42        4870.26
        963881480        Online        2804      205.70     117.11      435466.90
        341417157       Offline        1779      651.21     524.96      247956.32
        514321792        Online        8102        9.33       6.92       75591.66
        115456712       Offline        5062      651.21     524.96      471336.91


                   Total Cost  Total Profit  Ship Month
        Order ID
        669165933  1582243.50     951410.50           6
        963881480   328376.44     248406.36           9
        341417157   933903.84     224598.75           5
        514321792    56065.84      19525.82           7
        115456712  2657347.52     639077.50           2
```

```
[180]: sn.lmplot(x='Unit Price',y='Total Cost',data=sales_data,col='Sales␣
        ↪Channel',hue='Unit Cost',aspect=1.5,height=8)
```

```
[180]: <seaborn.axisgrid.FacetGrid at 0x789502657a60>
```
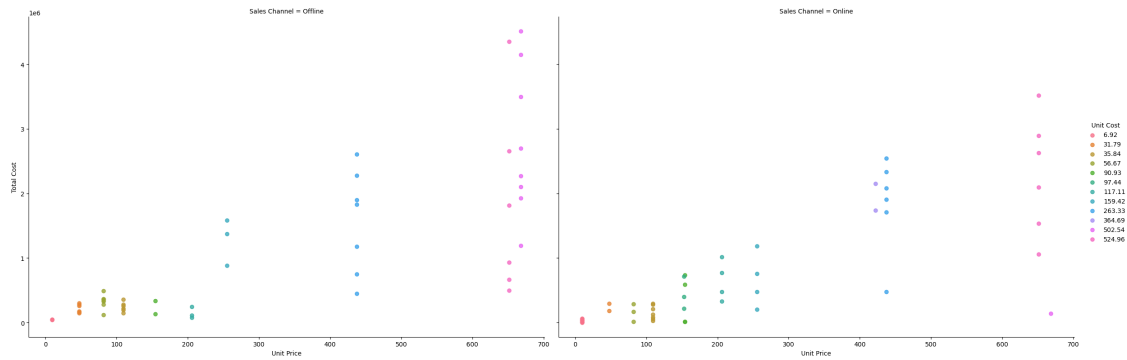
```
[181]: sales_data.sort_values(by='Unit Price')
```

```
[181]:            Order Date Order Priority  Ship Date    Item Type  \
       Order ID
       142278373 2014-09-08              H 2014-10-04       Fruits
       508980977 2013-09-17              H 2013-10-24       Fruits
       162052476 2011-11-22              L 2011-12-03       Fruits
       514321792 2014-06-20              C 2014-07-05       Fruits
       810711038 2011-11-11              L 2011-12-28       Fruits
       ...              ...            ...        ...          ...
       886494815 2012-05-26              L 2012-06-09    Household
       213487374 2012-10-21              L 2012-11-30    Household
       955357205 2012-01-05              L 2012-02-14    Household
       441619336 2010-12-30              L 2011-01-20    Household
       665095412 2012-02-10              L 2012-02-15    Household

                                   Region                Country Sales Channel  \
       Order ID
       142278373       Australia and Oceania            New Zealand        Online
       508980977          Sub-Saharan Africa  Sao Tome and Principe       Offline
       162052476  Middle East and North Africa                 Syria        Online
       514321792          Sub-Saharan Africa  Sao Tome and Principe        Online
       810711038                        Asia               Malaysia       Offline
       ...                               ...                    ...           ...
       886494815          Sub-Saharan Africa             The Gambia       Offline
       213487374                      Europe                  Spain       Offline
       955357205                      Europe         United Kingdom        Online
       441619336                        Asia           Turkmenistan       Offline
       665095412          Sub-Saharan Africa             Mozambique       Offline

                  Units Sold  Unit Price  Unit Cost  Total Revenue  Total Cost  \
       Order ID
       142278373        2187        9.33       6.92       20404.71    15134.04
       508980977        7637        9.33       6.92      802333.76    52848.04
```

33

```
162052476        3784      9.33      6.92      35304.72     26185.28
514321792        8102      9.33      6.92      75591.66     56065.84
810711038        6267      9.33      6.92   4368316.68     43367.64
...               ...       ...       ...          ...          ...
886494815        2370    668.27    502.54       6279.09  1191019.80
213487374        4513    668.27    502.54    3015902.51  2267963.02
955357205         282    668.27    502.54     994765.42   141716.28
441619336        3830    668.27    502.54     524870.06  1924728.20
665095412        5367    668.27    502.54    5997054.98  2697132.18


              Total Profit  Ship Month
Order ID
142278373          5270.67          10
508980977         18405.17          10
162052476          9119.44          12
514321792         19525.82           7
810711038         15103.47          12
...                    ...         ...
886494815        392780.10           6
213487374        747939.49          11
955357205         46735.86           2
441619336        634745.90           1
665095412        889472.91           2

[100 rows x 14 columns]
```

```
[182]: sales_data.plot.bar(x='Unit Price',y='Total␣
        ↪Cost',legend=None,figsize=(15,10),rot=0)
       mp.ylabel('Total Cost')
       mp.xticks(rotation=90)
       mp.locator_params(nbins=90)
```
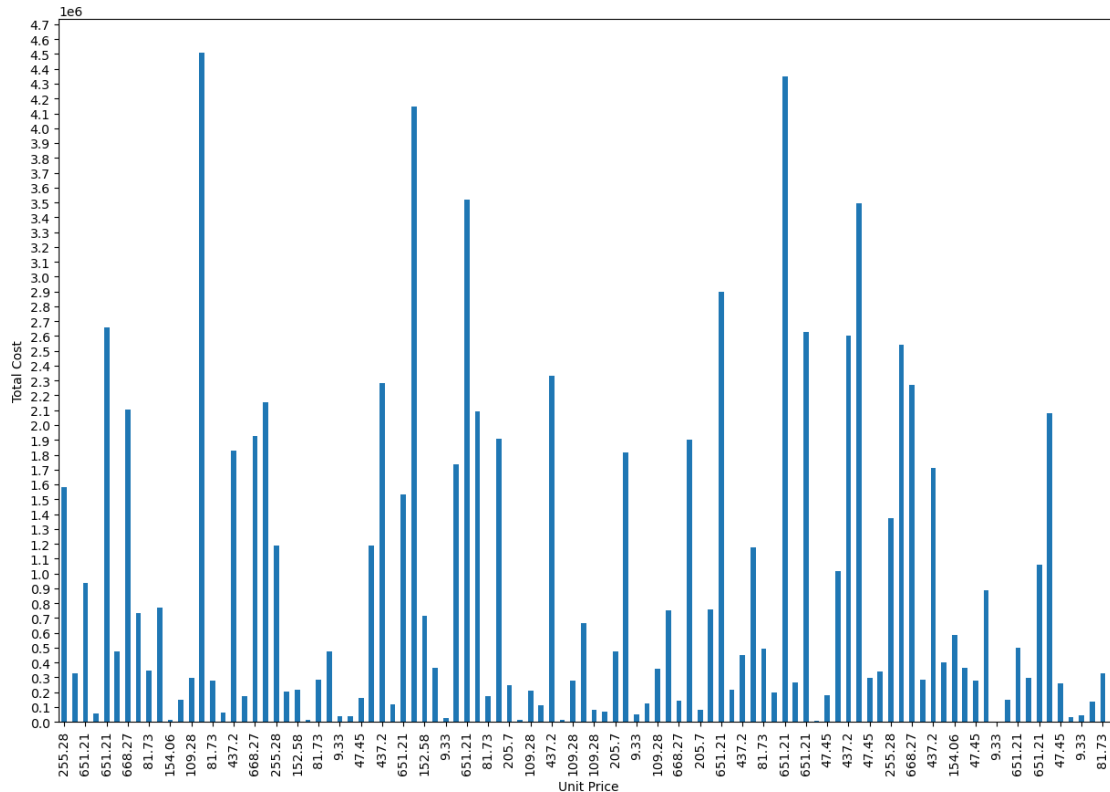
2.0.4 The above bar plot suggests that higher the value of unit price of a product, more will be the total cost of it.

```
[183]: np.cov(sales_data['Unit Price'],sales_data['Total Cost'])
```

```
[183]: array([[5.55037038e+04, 2.01205393e+08],
              [2.01205393e+08, 1.17492213e+12]])
```

```
[184]: np.corrcoef(sales_data['Unit Price'],sales_data['Total Cost'])
```

```
[184]: array([[1.        , 0.78790543],
              [0.78790543, 1.        ]])
```

The high value of degree of correlation between 'Unit Price' and 'Total Cost' variables indicates that they are almost directly proportional to each other and highly dependent on each other

```
[185]: sales_data.head()
```

```
[185]:            Order Date Order Priority  Ship Date     Item Type  \
       Order ID
       669165933 2010-05-28              H 2010-06-27     Baby Food
       963881480 2012-08-22              C 2012-09-15        Cereal
```

35

```
341417157 2014-05-02              L 2014-05-08  Office Supplies
514321792 2014-06-20              C 2014-07-05           Fruits
115456712 2013-02-01              L 2013-02-06  Office Supplies


                                   Region              Country  \
Order ID
669165933            Australia and Oceania               Tuvalu
963881480  Central America and the Caribbean            Grenada
341417157                          Europe               Russia
514321792              Sub-Saharan Africa  Sao Tome and Principe
115456712              Sub-Saharan Africa               Rwanda


          Sales Channel  Units Sold  Unit Price  Unit Cost  Total Revenue  \
Order ID
669165933       Offline        9925      255.28     159.42        4870.26
963881480        Online        2804      205.70     117.11      435466.90
341417157       Offline        1779      651.21     524.96      247956.32
514321792        Online        8102        9.33       6.92       75591.66
115456712       Offline        5062      651.21     524.96      471336.91


           Total Cost  Total Profit  Ship Month
Order ID
669165933  1582243.50     951410.50           6
963881480   328376.44     248406.36           9
341417157   933903.84     224598.75           5
514321792    56065.84      19525.82           7
115456712  2657347.52     639077.50           2
```
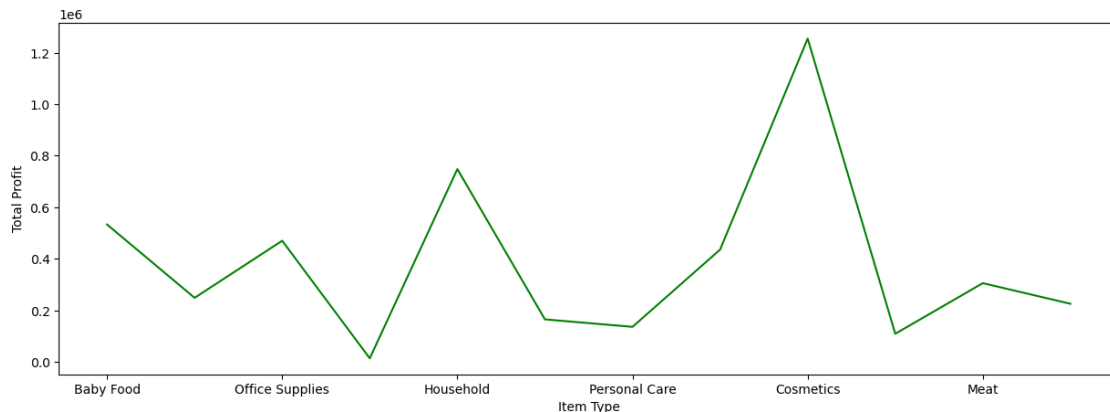
[186]:
```python
pd.pivot_table(sales_data,index='Item Type',values='Total Profit',aggfunc=np.
 ↪median).plot(kind='line',color='green',figsize=(15,5),legend=None)
mp.ylabel('Total Profit')
```

[186]: Text(0, 0.5, 'Total Profit')

```
[187]: sales_data.groupby('Region')['Total Profit'].count().
       ↪plot(kind='area',color=['purple','brown','blue','green'])
       mp.xticks(rotation=90)
       mp.ylabel('Total Profit')
```

[187]: Text(0, 0.5, 'Total Profit')
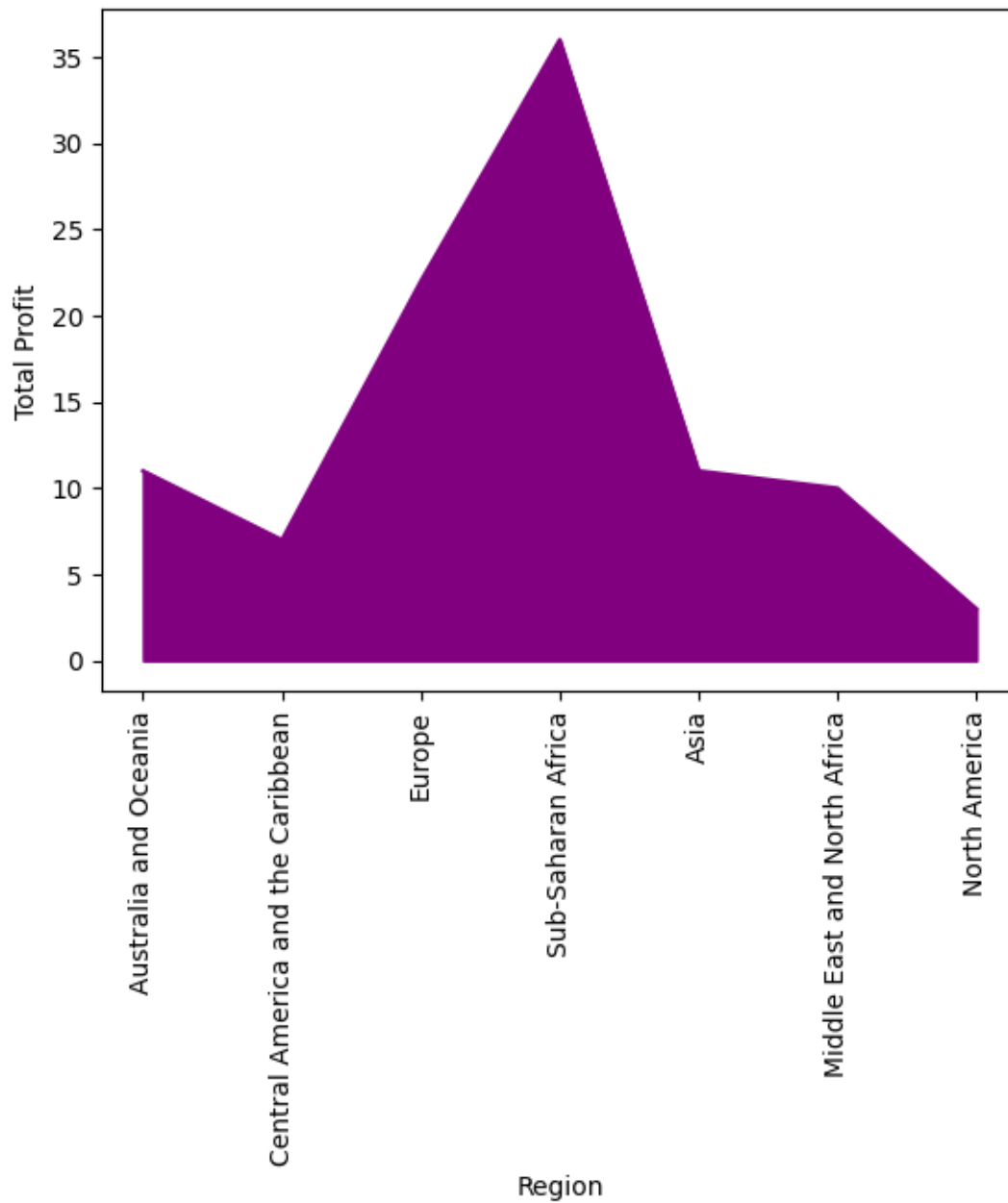
Maximum profit has been generated in the Sub-Saharan African region while minimum profit has

been generated in the North American region

```
[188]: sales_data['Order Priority'].unique()
```

```
[188]: array(['H', 'C', 'L', 'M'], dtype=object)
```

```
[189]: order_priorities = ['H', 'C', 'L', 'M']
```

```
[190]: sales_data['Order Priority'] = pd.Categorical(sales_data['Order␣
         ↪Priority'],categories=order_priorities,ordered=True)
```

```
[191]: sales_data.plot.bar(x='Order Priority',y='Total␣
         ↪Revenue',legend=None,figsize=(16,8))
       mp.xticks(rotation=90)
       mp.ylabel('Total Revenue')
```

```
[191]: Text(0, 0.5, 'Total Revenue')
```
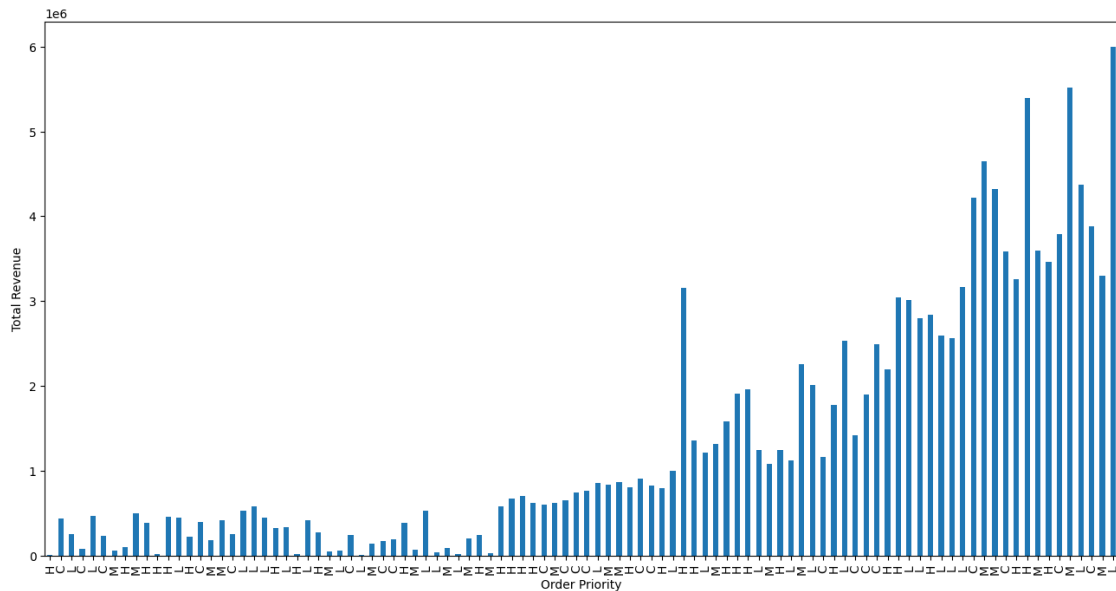


```
[192]: sales_data.groupby('Order Priority')['Total Revenue'].count().
         ↪plot(kind='bar',color=['grey','blue','green','purple','brown'])
       mp.ylabel('Total Revenue')
```

```
[192]: Text(0, 0.5, 'Total Revenue')
```
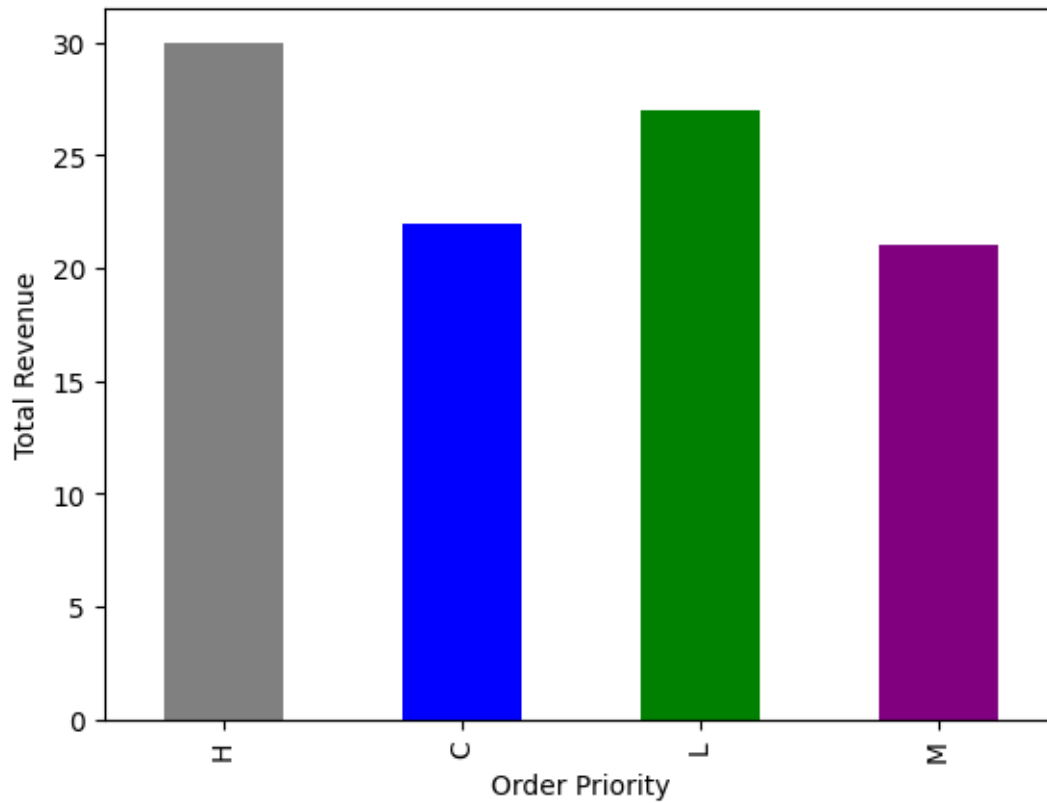
Maximum number of revenues has been generated by the products having order priority 'H' while minimum revenues has been generated by 'M' priority products

```
[193]: sales_data.groupby('Order Priority')['Total Profit'].sum().
       ↪plot(kind='bar',color=['brown','red','purple','aqua'])
       mp.ylabel('Total Profit')
       mp.title('Distribution of profit on the basis of order priority')
```

[193]: Text(0.5, 1.0, 'Distribution of profit on the basis of order priority')

Distribution of profit on the basis of order priority
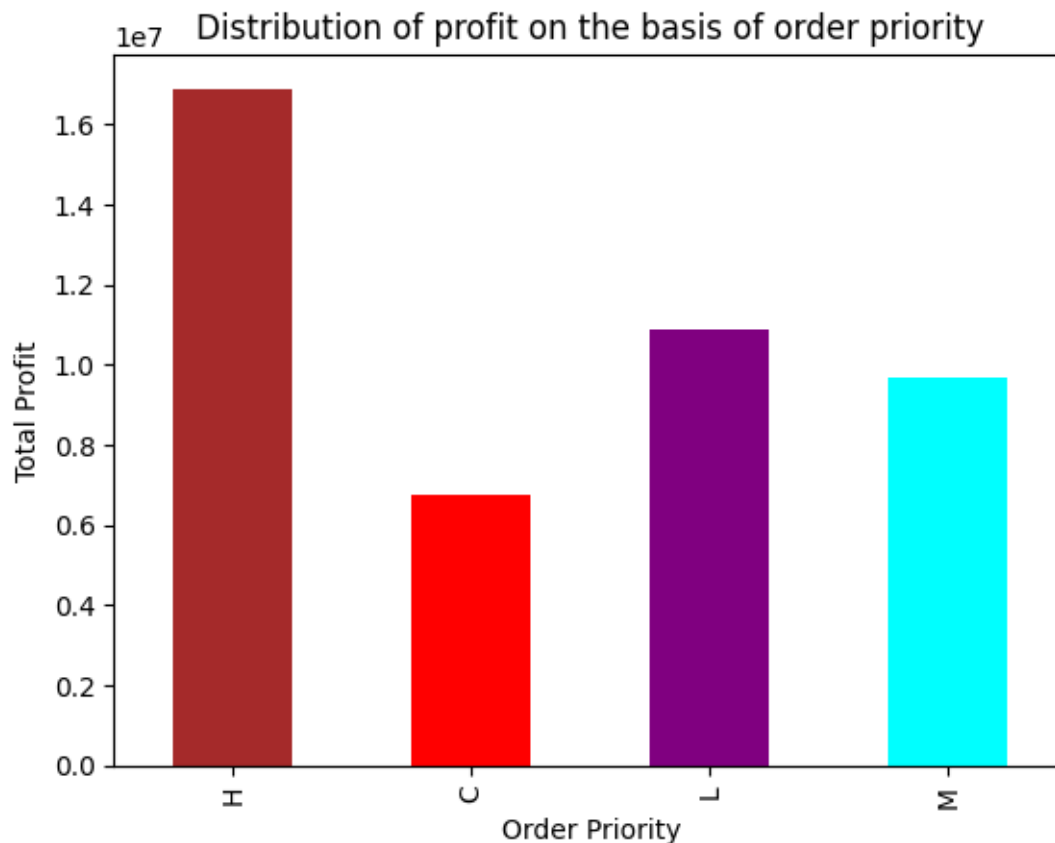
Maximum profit has been generated by products having order priority 'H' while minimum profit has been obtained in case of 'C' priority product orders

[195]: `!pip install joypy`

```
Collecting joypy
  Downloading joypy-0.2.6-py2.py3-none-any.whl (8.6 kB)
Requirement already satisfied: numpy>=1.16.5 in /usr/local/lib/python3.10/dist-
packages (from joypy) (1.25.2)
Requirement already satisfied: scipy>=0.11.0 in /usr/local/lib/python3.10/dist-
packages (from joypy) (1.11.4)
Requirement already satisfied: pandas>=0.20.0 in /usr/local/lib/python3.10/dist-
packages (from joypy) (2.0.3)
Requirement already satisfied: matplotlib in /usr/local/lib/python3.10/dist-
packages (from joypy) (3.7.1)
Requirement already satisfied: python-dateutil>=2.8.2 in
/usr/local/lib/python3.10/dist-packages (from pandas>=0.20.0->joypy) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-
packages (from pandas>=0.20.0->joypy) (2023.4)
Requirement already satisfied: tzdata>=2022.1 in /usr/local/lib/python3.10/dist-
packages (from pandas>=0.20.0->joypy) (2024.1)
```

```
Requirement already satisfied: contourpy>=1.0.1 in
/usr/local/lib/python3.10/dist-packages (from matplotlib->joypy) (1.2.1)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.10/dist-
packages (from matplotlib->joypy) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in
/usr/local/lib/python3.10/dist-packages (from matplotlib->joypy) (4.50.0)
Requirement already satisfied: kiwisolver>=1.0.1 in
/usr/local/lib/python3.10/dist-packages (from matplotlib->joypy) (1.4.5)
Requirement already satisfied: packaging>=20.0 in
/usr/local/lib/python3.10/dist-packages (from matplotlib->joypy) (24.0)
Requirement already satisfied: pillow>=6.2.0 in /usr/local/lib/python3.10/dist-
packages (from matplotlib->joypy) (9.4.0)
Requirement already satisfied: pyparsing>=2.3.1 in
/usr/local/lib/python3.10/dist-packages (from matplotlib->joypy) (3.1.2)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-
packages (from python-dateutil>=2.8.2->pandas>=0.20.0->joypy) (1.16.0)
Installing collected packages: joypy
Successfully installed joypy-0.2.6
```

[196]: `import joypy`

[197]:
```
joypy.joyplot(sales_data,column=['Total Revenue','Total Cost','Total␣
 ↪Profit'],by='Item␣
 ↪Type',figsize=(9,6),xlabelsize=13,ylim='own',ylabelsize=13,grid='both',yrot=10,fill=True,le
```

[197]:
```
(<Figure size 900x600 with 13 Axes>,
 [<Axes: >,
  <Axes: >,
  <Axes: >,
  <Axes: >,
  <Axes: >,
  <Axes: >,
  <Axes: >,
  <Axes: >,
  <Axes: >,
  <Axes: >,
  <Axes: >,
  <Axes: >,
  <Axes: >])
```

```
sales_data.head(10)
```

```
           Order Date Order Priority  Ship Date        Item Type  \
Order ID
669165933  2010-05-28              H 2010-06-27        Baby Food
963881480  2012-08-22              C 2012-09-15           Cereal
341417157  2014-05-02              L 2014-05-08  Office Supplies
514321792  2014-06-20              C 2014-07-05           Fruits
115456712  2013-02-01              L 2013-02-06  Office Supplies
547995746  2015-02-04              C 2015-02-21        Baby Food
135425221  2011-04-23              M 2011-04-27        Household
871543967  2012-07-17              H 2012-07-27       Vegetables
770463311  2015-07-14              M 2015-08-25    Personal Care
616607081  2014-04-18              H 2014-05-30           Cereal


                                Region                 Country  \
Order ID
669165933           Australia and Oceania                 Tuvalu
963881480  Central America and the Caribbean              Grenada
341417157                          Europe                 Russia
514321792              Sub-Saharan Africa  Sao Tome and Principe
115456712              Sub-Saharan Africa                 Rwanda
547995746           Australia and Oceania        Solomon Islands
135425221              Sub-Saharan Africa                 Angola
```

```
871543967                    Sub-Saharan Africa         Burkina Faso
770463311                    Sub-Saharan Africa  Republic of the Congo
616607081                    Sub-Saharan Africa              Senegal

          Sales Channel  Units Sold  Unit Price  Unit Cost  Total Revenue  \
Order ID
669165933       Offline        9925      255.28     159.42        4870.26
963881480        Online        2804      205.70     117.11      435466.90
341417157       Offline        1779      651.21     524.96      247956.32
514321792        Online        8102        9.33       6.92       75591.66
115456712       Offline        5062      651.21     524.96      471336.91
547995746        Online        2974      255.28     159.42      228779.10
135425221       Offline        4187      668.27     502.54       58471.11
871543967        Online        8082      154.06      90.93       97040.64
770463311       Offline        6070       81.73      56.67      496101.10
616607081        Online        6593      205.70     117.11      387002.20

           Total Cost  Total Profit  Ship Month
Order ID
669165933  1582243.50     951410.50           6
963881480   328376.44     248406.36           9
341417157   933903.84     224598.75           5
514321792    56065.84      19525.82           7
115456712  2657347.52     639077.50           2
547995746   474115.08     285087.64           2
135425221  2104134.98     693911.51           4
871543967   734896.26     510216.66           7
770463311   343986.90     152114.20           8
616607081   772106.23     584073.87           5
```

[202]: `np.max(sales_data.loc[:,'Total Profit'].iloc[:])`

[202]: 1719922.04

[203]: `np.set_printoptions = True`

[204]: `sales_data.head(8)`

[204]:
```
           Order Date Order Priority  Ship Date        Item Type  \
Order ID
669165933  2010-05-28              H 2010-06-27        Baby Food
963881480  2012-08-22              C 2012-09-15           Cereal
341417157  2014-05-02              L 2014-05-08  Office Supplies
514321792  2014-06-20              C 2014-07-05           Fruits
115456712  2013-02-01              L 2013-02-06  Office Supplies
547995746  2015-02-04              C 2015-02-21        Baby Food
135425221  2011-04-23              M 2011-04-27        Household
```

```
871543967 2012-07-17                    H 2012-07-27        Vegetables
```

```
                                      Region                 Country  \
Order ID
669165933              Australia and Oceania                 Tuvalu
963881480  Central America and the Caribbean                Grenada
341417157                             Europe                 Russia
514321792                Sub-Saharan Africa  Sao Tome and Principe
115456712                Sub-Saharan Africa                 Rwanda
547995746             Australia and Oceania        Solomon Islands
135425221                Sub-Saharan Africa                 Angola
871543967                Sub-Saharan Africa            Burkina Faso


          Sales Channel  Units Sold  Unit Price  Unit Cost  Total Revenue  \
Order ID
669165933       Offline        9925      255.28     159.42        4870.26
963881480        Online        2804      205.70     117.11      435466.90
341417157       Offline        1779      651.21     524.96      247956.32
514321792        Online        8102        9.33       6.92       75591.66
115456712       Offline        5062      651.21     524.96      471336.91
547995746        Online        2974      255.28     159.42      228779.10
135425221       Offline        4187      668.27     502.54       58471.11
871543967        Online        8082      154.06      90.93       97040.64


           Total Cost  Total Profit  Ship Month
Order ID
669165933  1582243.50     951410.50           6
963881480   328376.44     248406.36           9
341417157   933903.84     224598.75           5
514321792    56065.84      19525.82           7
115456712  2657347.52     639077.50           2
547995746   474115.08     285087.64           2
135425221  2104134.98     693911.51           4
871543967   734896.26     510216.66           7
```

[205]: `pd.DataFrame(sales_data.value_counts())`

[205]:
```
                                     count
Order Date Order Priority Ship Date  Item Type       Region
Country        Sales Channel Units Sold Unit Price Unit Cost Total Revenue
Total Cost Total Profit Ship Month
2010-02-02 C                 2010-03-18 Clothes        Europe
Albania        Online        2269        109.28     35.84      759202.72
81320.96   166635.36   3                 1
2014-07-07 H                 2014-07-11 Beverages      Australia and Oceania
Australia      Offline       9389        47.45      31.79      5396577.27
298476.31  147031.74   7                 1
```

```
2015-02-23 H              2015-03-02 Cosmetics        Europe
Austria       Offline       2847      437.20    263.33    793518.00
749700.51  495007.89    3             1
2015-02-04 C              2015-02-21 Baby Food       Australia and Oceania
Solomon Islands Online      2974      255.28    159.42    228779.10
474115.08  285087.64    2             1
2015-01-16 H              2015-03-01 Household       Asia
Myanmar       Offline       8250      668.27    502.54    380512.96
4145955.00 1367272.50   3             1
…

…
2012-04-30 M              2012-05-18 Fruits          Middle East and North
Africa Kuwait        Online       522       9.33       6.92      4324782.40
3612.24    1258.02     5             1
2012-04-23 M              2012-06-03 Clothes         Europe
Bulgaria      Online       1673      109.28    35.84     182825.44
59960.32   122865.12    6             1
2012-04-01 L              2012-05-08 Office Supplies Asia
Brunei        Online       6708      651.21    524.96    22312.29
3521431.68 846885.00    5             1
2012-03-18 L              2012-04-07 Vegetables      Sub-Saharan Africa
Kenya         Online       6457      154.06    90.93     2559474.10
587135.01  407630.41    4             1
2017-05-22 H              2017-06-05 Cosmetics       Europe
France        Online       1815      437.20    263.33    272410.45
477943.95  315574.05    6             1

[100 rows x 1 columns]
```

[207]:
```python
np.max(sales_data['Unit Cost'])
```

[207]: 524.96

[208]:
```python
temp = sales_data.copy()
temp['Unit Cost'] = temp['Unit Cost'].sort_values()
```

[209]:
```python
temp.head()
```

[209]:
```
            Order Date Order Priority  Ship Date       Item Type  \
Order ID
669165933   2010-05-28              H 2010-06-27        Baby Food
963881480   2012-08-22              C 2012-09-15           Cereal
341417157   2014-05-02              L 2014-05-08  Office Supplies
514321792   2014-06-20              C 2014-07-05           Fruits
115456712   2013-02-01              L 2013-02-06  Office Supplies

                            Region             Country  \
```

	Region	Country
Order ID		
669165933	Australia and Oceania	Tuvalu
963881480	Central America and the Caribbean	Grenada
341417157	Europe	Russia
514321792	Sub-Saharan Africa	Sao Tome and Principe
115456712	Sub-Saharan Africa	Rwanda

	Sales Channel	Units Sold	Unit Price	Unit Cost	Total Revenue
Order ID					
669165933	Offline	9925	255.28	159.42	4870.26
963881480	Online	2804	205.70	117.11	435466.90
341417157	Offline	1779	651.21	524.96	247956.32
514321792	Online	8102	9.33	6.92	75591.66
115456712	Offline	5062	651.21	524.96	471336.91

	Total Cost	Total Profit	Ship Month
Order ID			
669165933	1582243.50	951410.50	6
963881480	328376.44	248406.36	9
341417157	933903.84	224598.75	5
514321792	56065.84	19525.82	7
115456712	2657347.52	639077.50	2

```python
[210]: temp.sort_values(by='Unit Cost',inplace=True)
```

```python
[211]: temp.tail(1)
```

```
[211]:
```

	Order Date	Order Priority	Ship Date	Item Type
Order ID				
837559306	2012-01-11	C	2012-01-13	Office Supplies

	Region	Country	Sales Channel	Units Sold
Order ID				
837559306	Sub-Saharan Africa	Mauritania	Offline	1266

	Unit Price	Unit Cost	Total Revenue	Total Cost	Total Profit
Order ID					
837559306	651.21	524.96	745426.0	664599.36	159832.5

	Ship Month
Order ID	
837559306	1

```python
[212]: temp.head(1)
```

```
[212]:
```

	Order Date	Order Priority	Ship Date	Item Type	Region
Order ID					

```
508980977 2013-09-17                    H 2013-10-24    Fruits  Sub-Saharan Africa

                          Country Sales Channel  Units Sold  Unit Price  \
Order ID
508980977  Sao Tome and Principe       Offline        7637        9.33

           Unit Cost  Total Revenue  Total Cost  Total Profit  Ship Month
Order ID
508980977       6.92      802333.76    52848.04      18405.17          10
```

[213]: `sales_data.head()`

[213]:
```
            Order Date Order Priority  Ship Date       Item Type  \
Order ID
669165933  2010-05-28              H 2010-06-27       Baby Food
963881480  2012-08-22              C 2012-09-15          Cereal
341417157  2014-05-02              L 2014-05-08  Office Supplies
514321792  2014-06-20              C 2014-07-05          Fruits
115456712  2013-02-01              L 2013-02-06  Office Supplies

                                     Region                Country  \
Order ID
669165933              Australia and Oceania                 Tuvalu
963881480  Central America and the Caribbean                Grenada
341417157                             Europe                 Russia
514321792                 Sub-Saharan Africa  Sao Tome and Principe
115456712                 Sub-Saharan Africa                 Rwanda

           Sales Channel  Units Sold  Unit Price  Unit Cost  Total Revenue  \
Order ID
669165933        Offline        9925      255.28     159.42        4870.26
963881480         Online        2804      205.70     117.11      435466.90
341417157        Offline        1779      651.21     524.96      247956.32
514321792         Online        8102        9.33       6.92       75591.66
115456712        Offline        5062      651.21     524.96      471336.91

           Total Cost  Total Profit  Ship Month
Order ID
669165933  1582243.50     951410.50           6
963881480   328376.44     248406.36           9
341417157   933903.84     224598.75           5
514321792    56065.84      19525.82           7
115456712  2657347.52     639077.50           2
```
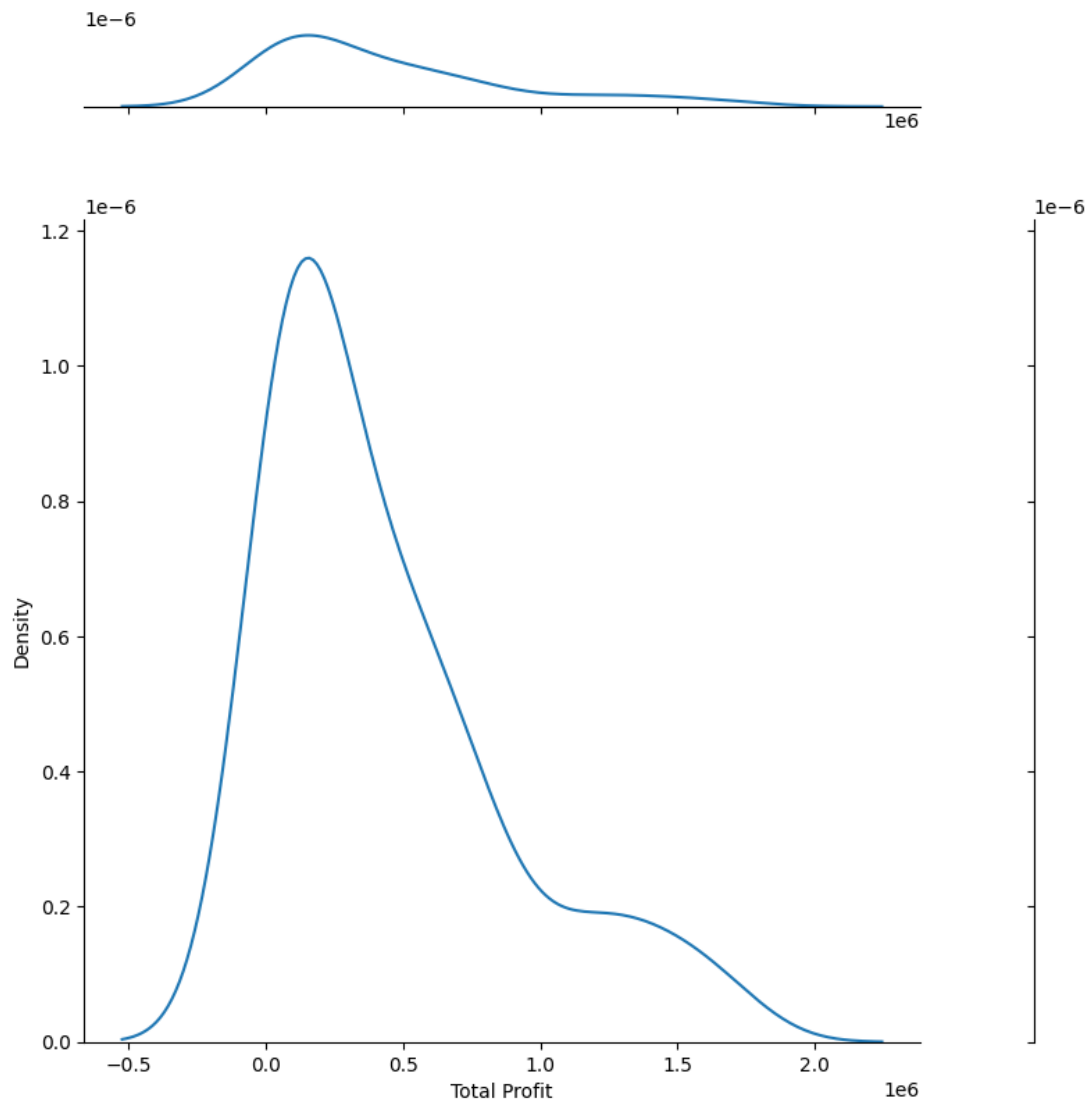
[214]:
```
sn.jointplot(x='Total
 ↪Profit',data=sales_data,height=8,ratio=5,kind='kde',space=1.5)
```

[214]: `<seaborn.axisgrid.JointGrid at 0x78950986d840>`

[215]: `sales_data['Item Type'].value_counts()`

[215]:
```
Item Type
Clothes           13
Cosmetics         13
Office Supplies   12
Fruits            10
Personal Care     10
Household          9
Beverages          8
Baby Food          7
```

```
Cereal          7
Vegetables      6
Snacks          3
Meat            2
Name: count, dtype: int64
```

[216]: `sales_data['Item Type'].dropna(inplace=True)`

[217]: `labels = sales_data['Item Type'].value_counts().index`

[218]:
```python
sizes = sales_data['Item Type'].value_counts().values
colors = ['grey','blue','red','lime','purple','brown']
```

[219]:
```python
mp.figure(figsize=(7,7))
mp.pie(sizes,labels=labels,colors=colors,autopct='%1.1f%%')
mp.title('Distribution of Item Types',fontsize=15,color='blue')
```

[219]: `Text(0.5, 1.0, 'Distribution of Item Types')`

Distribution of Item Types
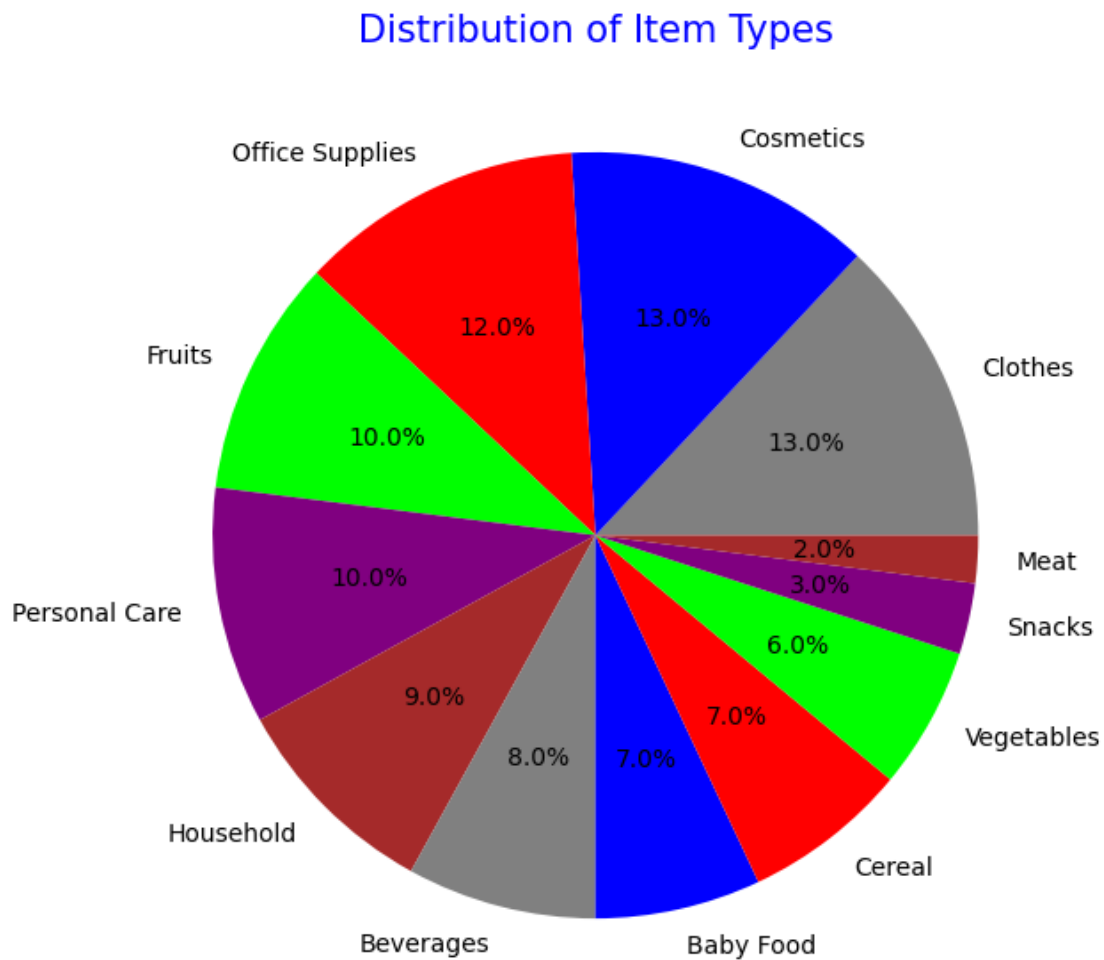
Clothes and cosmetics are the most purchased items while meat and snacks are the least purchased ones.

2.1 LM Plot

```
[220]: sales_data.head()
```

```
[220]:            Order Date Order Priority  Ship Date        Item Type  \
       Order ID
       669165933 2010-05-28              H 2010-06-27        Baby Food
       963881480 2012-08-22              C 2012-09-15           Cereal
       341417157 2014-05-02              L 2014-05-08  Office Supplies
       514321792 2014-06-20              C 2014-07-05           Fruits
       115456712 2013-02-01              L 2013-02-06  Office Supplies


                                        Region                Country  \
       Order ID
       669165933            Australia and Oceania                 Tuvalu
       963881480  Central America and the Caribbean              Grenada
       341417157                          Europe                 Russia
       514321792              Sub-Saharan Africa  Sao Tome and Principe
       115456712              Sub-Saharan Africa                 Rwanda


                 Sales Channel  Units Sold  Unit Price  Unit Cost  Total Revenue  \
       Order ID
       669165933       Offline        9925      255.28     159.42        4870.26
       963881480        Online        2804      205.70     117.11      435466.90
       341417157       Offline        1779      651.21     524.96      247956.32
       514321792        Online        8102        9.33       6.92       75591.66
       115456712       Offline        5062      651.21     524.96      471336.91


                 Total Cost  Total Profit  Ship Month
       Order ID
       669165933 1582243.50     951410.50           6
       963881480  328376.44     248406.36           9
       341417157  933903.84     224598.75           5
       514321792   56065.84      19525.82           7
       115456712 2657347.52     639077.50           2
```

```
[221]: sales_data['Country'].nunique()
```

```
[221]: 76
```

```
[222]: countries_data = pd.DataFrame(sales_data['Country'].unique())
       countries_data
```

```
[222]:                          0
      0                     Tuvalu
      1                    Grenada
      2                     Russia
      3    Sao Tome and Principe
      4                     Rwanda
      ..                        …
      71                   Slovenia
      72                    Romania
      73                  Nicaragua
      74                   Malaysia
      75                 Mozambique

      [76 rows x 1 columns]
```

```
[223]: countries = sales_data['Country'].values.tolist()
```

```
[224]: countries
```

```
[224]: ['Tuvalu',
       'Grenada',
       'Russia',
       'Sao Tome and Principe',
       'Rwanda',
       'Solomon Islands',
       'Angola',
       'Burkina Faso',
       'Republic of the Congo',
       'Senegal',
       'Kyrgyzstan',
       'Cape Verde',
       'Bangladesh',
       'Honduras',
       'Mongolia',
       'Bulgaria',
       'Sri Lanka',
       'Cameroon',
       'Turkmenistan',
       'East Timor',
       'Norway',
       'Portugal',
       'Honduras',
       'New Zealand',
       'Moldova ',
       'France',
       'Kiribati',
       'Mali',
```

```
'Norway',
'The Gambia',
'Switzerland',
'South Sudan',
'Australia',
'Myanmar',
'Djibouti',
'Costa Rica',
'Syria',
'The Gambia',
'Brunei',
'Bulgaria',
'Niger',
'Azerbaijan',
'The Gambia',
'Slovakia',
'Myanmar',
'Comoros',
'Iceland',
'Switzerland',
'Macedonia',
'Mauritania',
'Albania',
'Lesotho',
'Saudi Arabia',
'Sierra Leone',
'Sao Tome and Principe',
"Cote d'Ivoire",
'Fiji',
'Austria',
'United Kingdom',
'Djibouti',
'Australia',
'San Marino',
'Cameroon',
'Libya',
'Haiti',
'Rwanda',
'Gabon',
'Belize',
'Lithuania',
'Madagascar',
'Turkmenistan',
'Libya',
'Democratic Republic of the Congo',
'Djibouti',
'Pakistan',
```

```
    'Mexico',
    'Federated States of Micronesia',
    'Laos',
    'Monaco',
    'Samoa ',
    'Spain',
    'Lebanon',
    'Iran',
    'Zambia',
    'Kenya',
    'Mexico',
    'Sao Tome and Principe',
    'The Gambia',
    'Kuwait',
    'Slovenia',
    'Sierra Leone',
    'Australia',
    'Azerbaijan',
    'Romania',
    'Nicaragua',
    'Mali',
    'Malaysia',
    'Sierra Leone',
    'Mexico',
    'Mozambique']
```
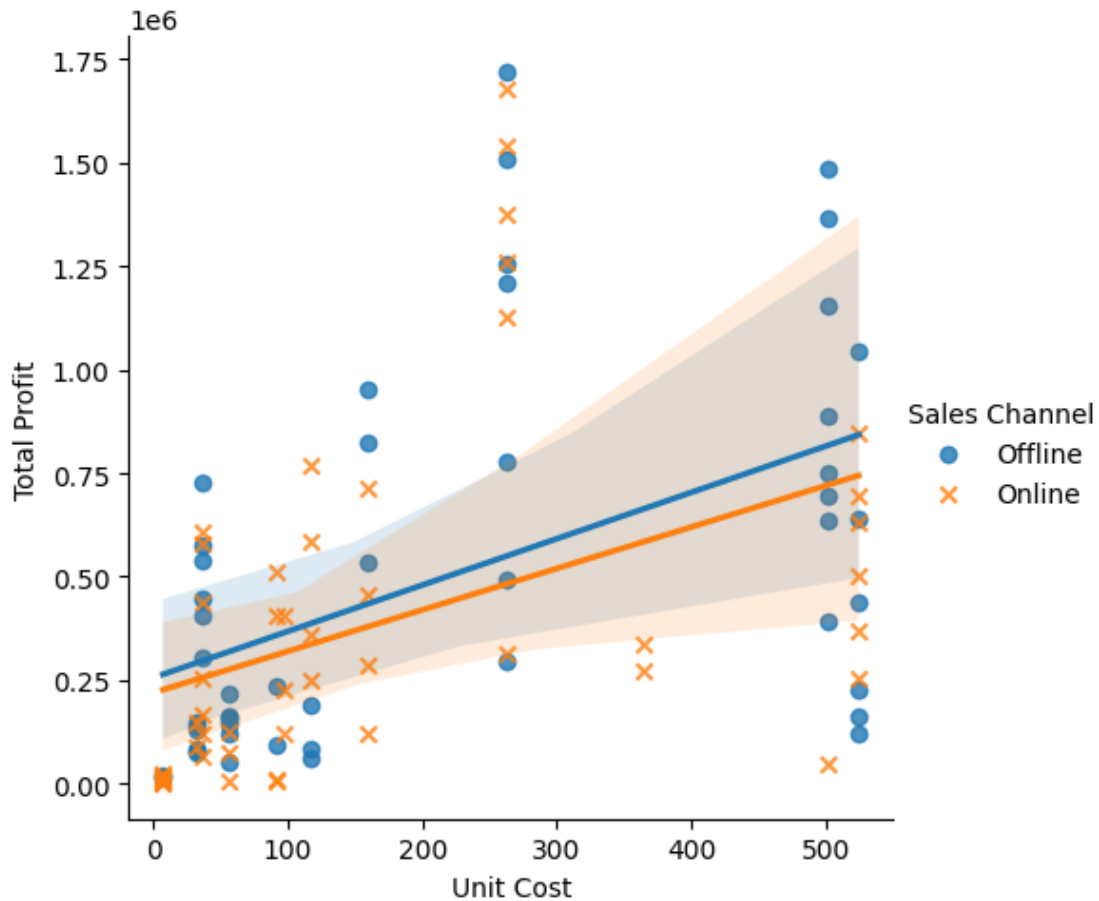
[225]:
```python
sn.lmplot(x='Unit Cost',y='Total
 ↪Profit',data=sales_data,height=5,aspect=1,hue='Sales
 ↪Channel',logx=False,truncate=True,ci=100,y_jitter=2.
 ↪2,scatter=True,fit_reg=True,markers=['o','x'])
```

[225]: <seaborn.axisgrid.FacetGrid at 0x78950155f400>

2.1.1 From the above LM plot, we can infer that profit keeps on increasing with increase in unit cost.

3 KDE Plot

```
[226]: sales_data['Order Date'] = pd.to_datetime(sales_data['Order Date'])
```
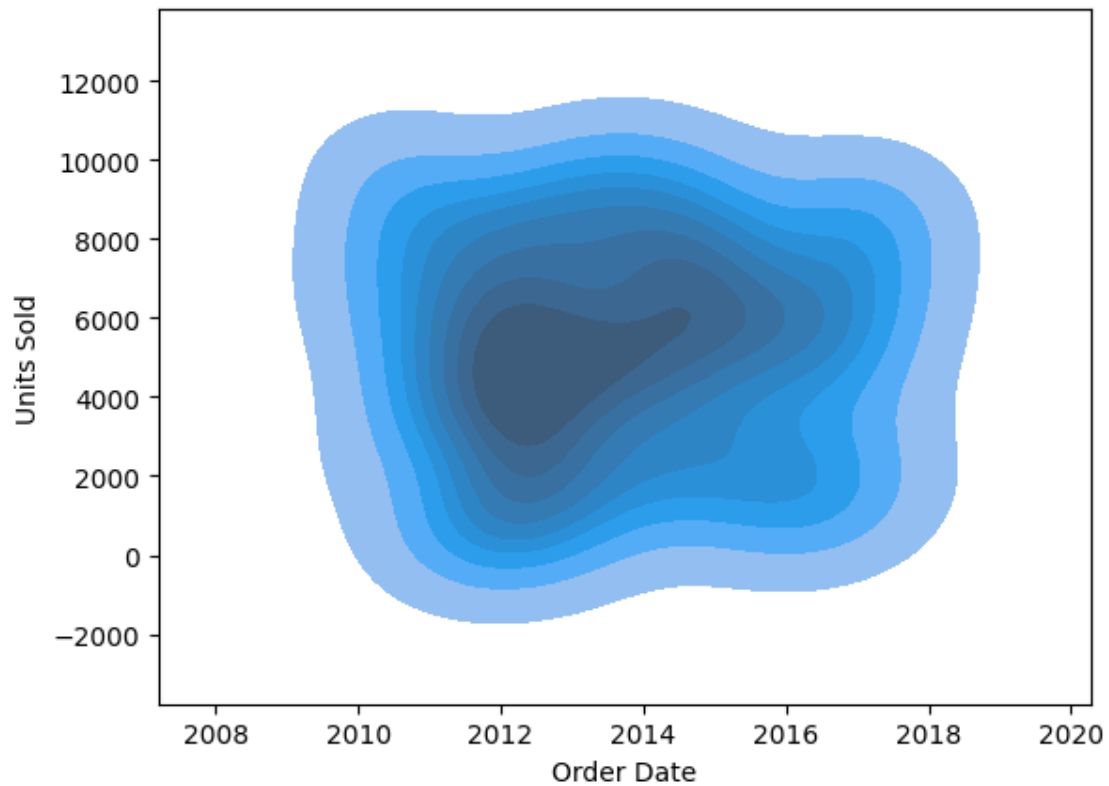
```
[227]: sn.kdeplot(x='Order Date',y='Units Sold',data=sales_data,shade=True,cut=3)
```

```
<ipython-input-227-3e7f08ef35e3>:1: FutureWarning:

`shade` is now deprecated in favor of `fill`; setting `fill=True`.
This will become an error in seaborn v0.14.0; please update your code.

  sn.kdeplot(x='Order Date',y='Units Sold',data=sales_data,shade=True,cut=3)
```

```
[227]: <Axes: xlabel='Order Date', ylabel='Units Sold'>
```

[228]:
```
sales_data['Ship Date'] = pd.to_datetime(sales_data['Ship Date'])
```
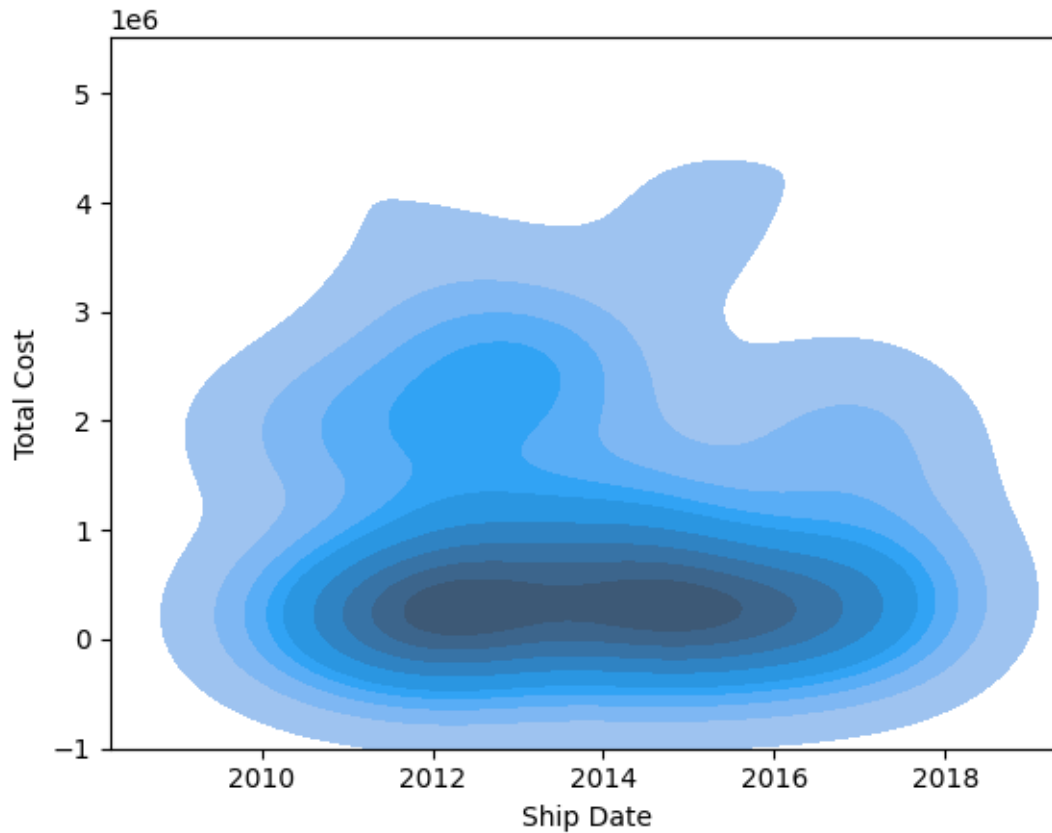
[229]:
```
sn.kdeplot(x='Ship Date',y='Total Cost',data=sales_data,shade=True,cut=2)
```

<ipython-input-229-4545e61a0963>:1: FutureWarning:

`shade` is now deprecated in favor of `fill`; setting `fill=True`.
This will become an error in seaborn v0.14.0; please update your code.

  sn.kdeplot(x='Ship Date',y='Total Cost',data=sales_data,shade=True,cut=2)

[229]: <Axes: xlabel='Ship Date', ylabel='Total Cost'>

4 Violin Plot

```
[230]: pal = sn.cubehelix_palette(2, rot=-.5, dark=.3)
```

```
[231]: sn.violinplot(data=sales_data,palette=pal,inner="points")
       list_a = list(range(12))
       mp.locator_params(axis='y',nbins=10)
       mp.locator_params(axis='x',nbins=10)
```
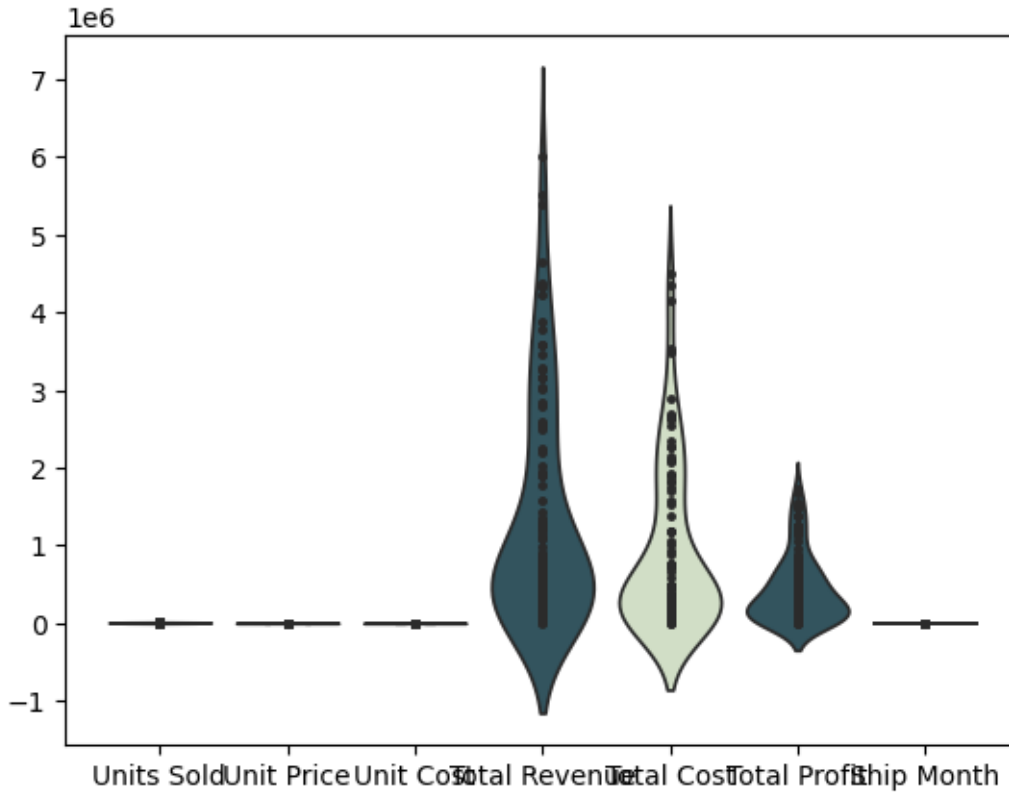
```
<ipython-input-231-5c87db5b1579>:1: UserWarning:
The palette list has fewer values (2) than needed (7) and will cycle, which may
produce an uninterpretable plot.
  sn.violinplot(data=sales_data,palette=pal,inner="points")
<ipython-input-231-5c87db5b1579>:4: UserWarning: 'set_params()' not defined for
locator of type <class 'matplotlib.category.StrCategoryLocator'>
  mp.locator_params(axis='x',nbins=10)
```

```
[232]: sales_data.head()
```

```
[232]:            Order Date Order Priority  Ship Date       Item Type  \
       Order ID
       669165933  2010-05-28              H 2010-06-27       Baby Food
       963881480  2012-08-22              C 2012-09-15          Cereal
       341417157  2014-05-02              L 2014-05-08  Office Supplies
       514321792  2014-06-20              C 2014-07-05          Fruits
       115456712  2013-02-01              L 2013-02-06  Office Supplies

                                     Region                 Country  \
       Order ID
       669165933            Australia and Oceania                  Tuvalu
       963881480  Central America and the Caribbean               Grenada
       341417157                           Europe                  Russia
       514321792              Sub-Saharan Africa  Sao Tome and Principe
       115456712              Sub-Saharan Africa                  Rwanda

                 Sales Channel  Units Sold  Unit Price  Unit Cost  Total Revenue  \
       Order ID
       669165933       Offline        9925      255.28     159.42        4870.26
```

Order ID					
963881480	Online	2804	205.70	117.11	435466.90
341417157	Offline	1779	651.21	524.96	247956.32
514321792	Online	8102	9.33	6.92	75591.66
115456712	Offline	5062	651.21	524.96	471336.91

|	Total Cost	Total Profit	Ship Month
Order ID			
669165933	1582243.50	951410.50	6
963881480	328376.44	248406.36	9
341417157	933903.84	224598.75	5
514321792	56065.84	19525.82	7
115456712	2657347.52	639077.50	2

```
[233]: sales_data.plot(x='Unit Cost', y='Total Profit', kind='bar')
       mp.xticks(rotation=90,fontsize=10)
       mp.yticks(rotation=90,fontsize=10)
       mp.locator_params(axis='x',nbins=40)
       mp.locator_params(axis='y',nbins=30)
       mp.ylabel('Total Profit')
```

```
[233]: Text(0, 0.5, 'Total Profit')
```
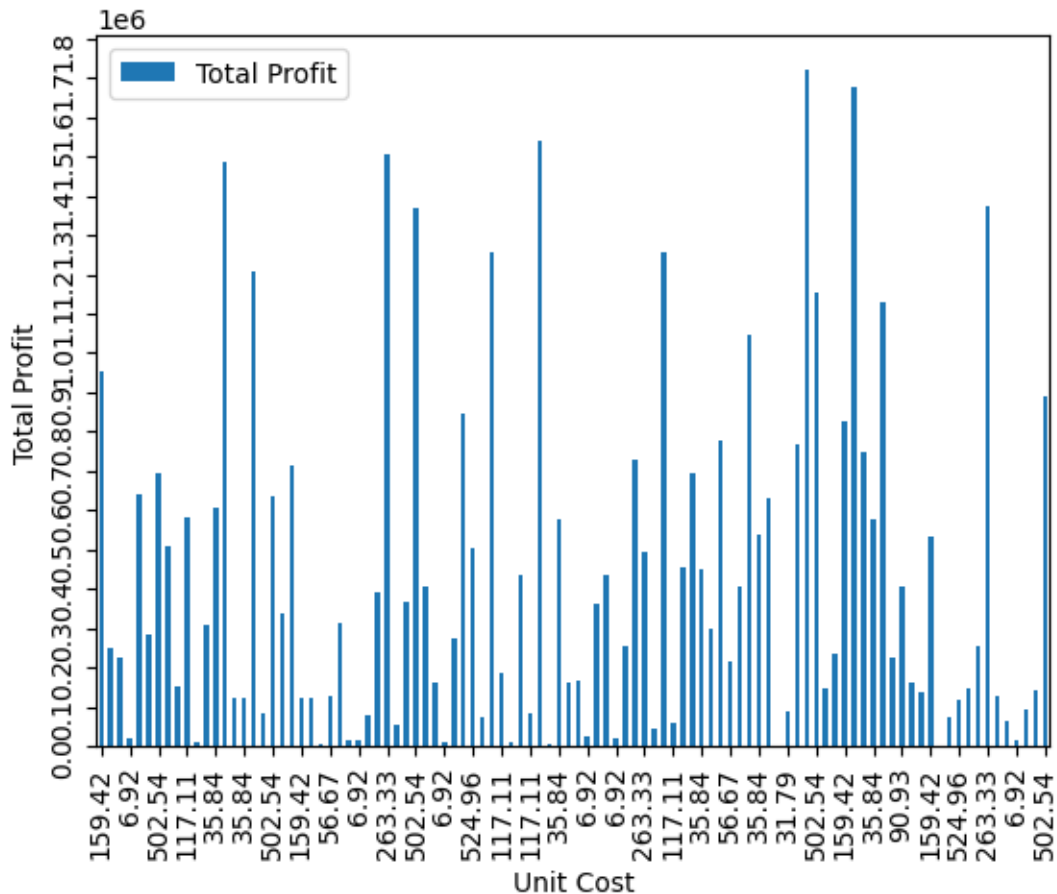
```
[234]: pd.pivot_table(index='Order Date',values='Total␣
        ↪Revenue',data=sales_data,aggfunc='count').
        ↪plot(kind='line',color='blue',legend=False)
       mp.ylabel('Total Revenue')
       mp.yticks(fontsize=12,color='green')
       mp.xticks(fontsize=12,color='purple')
       mp.title('Revenue generated per Order Date',fontsize=16,color='grey')
```

[234]: Text(0.5, 1.0, 'Revenue generated per Order Date')
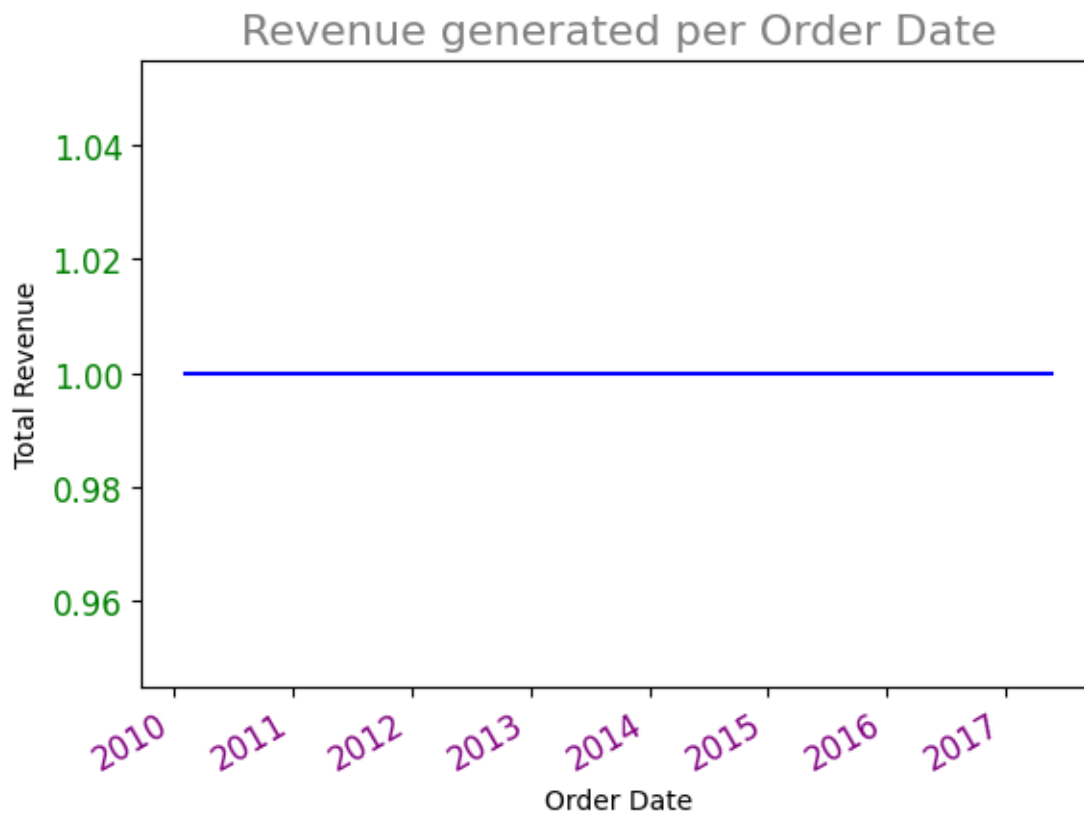
4.0.1 Total revenue remains constant for every year from 2010 to 2017