amazon-sales-data-project

June 8, 2024

0.1 Project Tittle: Amazon Sales Data Analysis

0.2 Technologies : Data Science

0.3 Domain: E-commerce

0.4 Project Difficulties level : Advanced

0.4.1 PROBLEM STATEMENT:

*Sales management has gained importance to meet increasing competition and the need for improved methods of distribution to reduce cost and to increase profits. Sales management today is the most important function in a commercial and business enterprise. Do ETL: Extract-Transform-Load some Amazon dataset and find for me Sales-trend -> month-wise, year-wise, yearly_month-wise Find key metrics and factors and show the meaningful relationships between attributes. Do your own research and come up with your f*indings.

0.4.2 DATA COLLECTION

[1]:
```
# importing libraries

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

import warnings
warnings.filterwarnings('ignore')
```

[2]:
```
# importing dataset

sales= pd.read_csv("Amazon sales data.csv")
sales.head(20)
```

[2]:

	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	
5	Australia and Oceania	Solomon Islands	Baby Food	
6	Sub-Saharan Africa	Angola	Household	
7	Sub-Saharan Africa	Burkina Faso	Vegetables	
8	Sub-Saharan Africa	Republic of the Congo	Personal Care	
9	Sub-Saharan Africa	Senegal	Cereal	
10	Asia	Kyrgyzstan	Vegetables	
11	Sub-Saharan Africa	Cape Verde	Clothes	
12	Asia	Bangladesh	Clothes	
13	Central America and the Caribbean	Honduras	Household	
14	Asia	Mongolia	Personal Care	
15	Europe	Bulgaria	Clothes	
16	Asia	Sri Lanka	Cosmetics	
17	Sub-Saharan Africa	Cameroon	Beverages	
18	Asia	Turkmenistan	Household	
19	Australia and Oceania	East Timor	Meat	

	Sales Channel	Order Priority	Order Date	Order ID	Ship Date	\
0	Offline	H	5/28/2010	669165933	6/27/2010	
1	Online	C	8/22/2012	963881480	9/15/2012	
2	Offline	L	5/2/2014	341417157	5/8/2014	
3	Online	C	6/20/2014	514321792	7/5/2014	
4	Offline	L	2/1/2013	115456712	2/6/2013	
5	Online	C	2/4/2015	547995746	2/21/2015	
6	Offline	M	4/23/2011	135425221	4/27/2011	
7	Online	H	7/17/2012	871543967	7/27/2012	
8	Offline	M	7/14/2015	770463311	8/25/2015	
9	Online	H	4/18/2014	616607081	5/30/2014	
10	Online	H	6/24/2011	814711606	7/12/2011	
11	Offline	H	8/2/2014	939825713	8/19/2014	
12	Online	L	1/13/2017	187310731	3/1/2017	
13	Offline	H	2/8/2017	522840487	2/13/2017	
14	Offline	C	2/19/2014	832401311	2/23/2014	
15	Online	M	4/23/2012	972292029	6/3/2012	
16	Offline	M	11/19/2016	419123971	12/18/2016	
17	Offline	C	4/1/2015	519820964	4/18/2015	
18	Offline	L	12/30/2010	441619336	1/20/2011	
19	Online	L	7/31/2012	322067916	9/11/2012	

	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
5	2974	255.28	159.42	759202.72	474115.08	285087.64
6	4187	668.27	502.54	2798046.49	2104134.98	693911.51

```
7         8082      154.06      90.93     1245112.92   734896.26     510216.66
8         6070       81.73      56.67      496101.10   343986.90     152114.20
9         6593      205.70     117.11     1356180.10   772106.23     584073.87
10         124      154.06      90.93       19103.44    11275.32       7828.12
11        4168      109.28      35.84      455479.04   149381.12     306097.92
12        8263      109.28      35.84      902980.64   296145.92     606834.72
13        8974      668.27     502.54     5997054.98  4509793.96    1487261.02
14        4901       81.73      56.67      400558.73   277739.67     122819.06
15        1673      109.28      35.84      182825.44    59960.32     122865.12
16        6952      437.20     263.33     3039414.40  1830670.16    1208744.24
17        5430       47.45      31.79      257653.50   172619.70      85033.80
18        3830      668.27     502.54     2559474.10  1924728.20     634745.90
19        5908      421.89     364.69     2492526.12  2154588.52     337937.60
```

[3]:
```
sales.tail()
```

[3]:
```
                Region       Country      Item Type Sales Channel  \
95  Sub-Saharan Africa          Mali        Clothes        Online
96                Asia      Malaysia         Fruits       Offline
97  Sub-Saharan Africa  Sierra Leone     Vegetables       Offline
98       North America        Mexico  Personal Care       Offline
99  Sub-Saharan Africa    Mozambique      Household       Offline

   Order Priority  Order Date   Order ID   Ship Date  Units Sold  Unit Price  \
95              M   7/26/2011  512878119    9/3/2011         888      109.28
96              L  11/11/2011  810711038  12/28/2011        6267        9.33
97              C    6/1/2016  728815257   6/29/2016        1485      154.06
98              M   7/30/2015  559427106    8/8/2015        5767       81.73
99              L   2/10/2012  665095412   2/15/2012        5367      668.27

    Unit Cost  Total Revenue  Total Cost  Total Profit
95      35.84       97040.64    31825.92      65214.72
96       6.92       58471.11    43367.64      15103.47
97      90.93      228779.10   135031.05      93748.05
98      56.67      471336.91   326815.89     144521.02
99     502.54     3586605.09  2697132.18     889472.91
```

[4]:
```
# Checking the shape

sales.shape
```

[4]:
```
(100, 14)
```

[5]:
```
sales.columns
```

[5]:
```
Index(['Region', 'Country', 'Item Type', 'Sales Channel', 'Order Priority',
       'Order Date', 'Order ID', 'Ship Date', 'Units Sold', 'Unit Price',
```

```
       'Unit Cost', 'Total Revenue', 'Total Cost', 'Total Profit'],
      dtype='object')
```

[6]:
```
sales.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100 entries, 0 to 99
Data columns (total 14 columns):
 #   Column          Non-Null Count  Dtype  
---  ------          --------------  -----  
 0   Region          100 non-null    object 
 1   Country         100 non-null    object 
 2   Item Type       100 non-null    object 
 3   Sales Channel   100 non-null    object 
 4   Order Priority  100 non-null    object 
 5   Order Date      100 non-null    object 
 6   Order ID        100 non-null    int64  
 7   Ship Date       100 non-null    object 
 8   Units Sold      100 non-null    int64  
 9   Unit Price      100 non-null    float64
 10  Unit Cost       100 non-null    float64
 11  Total Revenue   100 non-null    float64
 12  Total Cost      100 non-null    float64
 13  Total Profit    100 non-null    float64
dtypes: float64(5), int64(2), object(7)
memory usage: 11.1+ KB
```

[7]:
```
sales.isnull().sum()
```

[7]:
```
Region            0
Country           0
Item Type         0
Sales Channel     0
Order Priority    0
Order Date        0
Order ID          0
Ship Date         0
Units Sold        0
Unit Price        0
Unit Cost         0
Total Revenue     0
Total Cost        0
Total Profit      0
dtype: int64
```

[8]:
```
sales.describe()
```

[8]:
```
           Order ID   Units Sold  Unit Price   Unit Cost  Total Revenue  \
count  1.000000e+02   100.000000  100.000000  100.000000   1.000000e+02
mean   5.550204e+08  5128.710000  276.761300  191.048000   1.373488e+06
std    2.606153e+08  2794.484562  235.592241  188.208181   1.460029e+06
min    1.146066e+08   124.000000    9.330000    6.920000   4.870260e+03
25%    3.389225e+08  2836.250000   81.730000   35.840000   2.687212e+05
50%    5.577086e+08  5382.500000  179.880000  107.275000   7.523144e+05
75%    7.907551e+08  7369.000000  437.200000  263.330000   2.212045e+06
max    9.940222e+08  9925.000000  668.270000  524.960000   5.997055e+06

         Total Cost  Total Profit
count  1.000000e+02  1.000000e+02
mean   9.318057e+05  4.416820e+05
std    1.083938e+06  4.385379e+05
min    3.612240e+03  1.258020e+03
25%    1.688680e+05  1.214436e+05
50%    3.635664e+05  2.907680e+05
75%    1.613870e+06  6.358288e+05
max    4.509794e+06  1.719922e+06
```

[9]:
```
sales['Item Type'].unique()
```

[9]:
```
array(['Baby Food', 'Cereal', 'Office Supplies', 'Fruits', 'Household',
       'Vegetables', 'Personal Care', 'Clothes', 'Cosmetics', 'Beverages',
       'Meat', 'Snacks'], dtype=object)
```

[10]:
```
sales['Order Year'] = pd.DatetimeIndex(sales['Order Date']).year
sales['Order Quarter']= pd.DatetimeIndex(sales['Order Date']).quarter
sales['Order Month']= pd.DatetimeIndex(sales['Order Date']).month
```

[12]:
```
sales1= sales[['Region' ,'Country', 'Item Type', 'Sales Channel', 'Order␣
↪Priority', 'Order Date', 'Order Year', 'Order Quarter', 'Order Month','Order␣
↪ID','Ship Date','Units Sold', 'Unit Price', 'Unit Cost','Total␣
↪Revenue','Total Cost','Total Profit']]
```

[13]:
```
sales1
```

[13]:
```
                              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
..                                  …                     …               …
95                   Sub-Saharan Africa                  Mali          Clothes
96                                 Asia               Malaysia           Fruits
97                   Sub-Saharan Africa          Sierra Leone       Vegetables
```

```
98                    North America                  Mexico    Personal Care
99                Sub-Saharan Africa              Mozambique        Household

   Sales Channel Order Priority  Order Date  Order Year  Order Quarter  \
0        Offline              H   5/28/2010        2010              2
1         Online              C   8/22/2012        2012              3
2        Offline              L    5/2/2014        2014              2
3         Online              C   6/20/2014        2014              2
4        Offline              L    2/1/2013        2013              1
..           …             …         …          …             …
95        Online              M   7/26/2011        2011              3
96       Offline              L  11/11/2011        2011              4
97       Offline              C    6/1/2016        2016              2
98       Offline              M   7/30/2015        2015              3
99       Offline              L   2/10/2012        2012              1

    Order Month   Order ID   Ship Date  Units Sold  Unit Price  Unit Cost  \
0             5  669165933   6/27/2010        9925      255.28     159.42
1             8  963881480   9/15/2012        2804      205.70     117.11
2             5  341417157    5/8/2014        1779      651.21     524.96
3             6  514321792    7/5/2014        8102        9.33       6.92
4             2  115456712    2/6/2013        5062      651.21     524.96
..          …        …          …          …         …        …
95            7  512878119    9/3/2011         888      109.28      35.84
96           11  810711038  12/28/2011        6267        9.33       6.92
97            6  728815257   6/29/2016        1485      154.06      90.93
98            7  559427106    8/8/2015        5767       81.73      56.67
99            2  665095412   2/15/2012        5367      668.27     502.54

    Total Revenue  Total Cost  Total Profit
0      2533654.00  1582243.50     951410.50
1       576782.80   328376.44     248406.36
2      1158502.59   933903.84     224598.75
3        75591.66    56065.84      19525.82
4      3296425.02  2657347.52     639077.50
..            …           …           …
95       97040.64    31825.92      65214.72
96       58471.11    43367.64      15103.47
97      228779.10   135031.05      93748.05
98      471336.91   326815.89     144521.02
99     3586605.09  2697132.18     889472.91

[100 rows x 17 columns]
```

[14]:
```
sort_sales1 =sales1.sort_values(by=['Units Sold'], ascending= False)
```

[16]:
```
Items_sold = sort_sales1[['Region', 'Item Type', 'Order Year', 'Order Quarter',␣
↪'Order Month', 'Units Sold']]
```

[17]:
```
Items_sold.head()
```

[17]:

	Region	Item Type	Order Year	Order Quarter	Order Month	Units Sold
0	Australia and Oceania	Baby Food	2010	2	5	9925
56	Australia and Oceania	Clothes	2010	2	6	9905
74	Middle East and North Africa	Cosmetics	2013	3	7	9892
79	Australia and Oceania	Cosmetics	2013	3	7	9654
51	Sub-Saharan Africa	Fruits	2013	3	8	9606

[18]:
```
Items_sold.tail()
```

[18]:

	Region	Item Type	Order Year	Order Quarter	Order Month	Units Sold
88	Middle East and North Africa	Fruits	2012	2	4	522
58	Europe	Household	2012	1	1	282
47	Europe	Personal Care	2010	4	12	273
43	Europe	Vegetables	2012	4	10	171
10	Asia	Vegetables	2011	2	6	124

[19]:
```
sort_sales2 =sales1.sort_values(by=['Total Revenue'], ascending= False)
```

[21]:
```
Items_revenue = sort_sales2[['Region', 'Item Type', 'Order Year', 'Order␣
↪Quarter', 'Order Month', 'Units Sold', 'Total Revenue']]
```

[22]:
```
Items_revenue.head()
```

[22]:

	Region	Item Type	Order Year
13	Central America and the Caribbean	Household	2017
33	Asia	Household	2015
68	Europe	Office Supplies	2010
75	North America	Household	2014

```
38                              Asia  Office Supplies        2012

    Order Quarter  Order Month  Units Sold  Total Revenue
13              1            2        8974     5997054.98
33              1            1        8250     5513227.50
68              4           10        8287     5396577.27
75              4           11        6954     4647149.58
38              2            4        6708     4368316.68
```

[23]:
```
Items_revenue.tail()
```

[23]:
```
                         Region      Item Type  Order Year  Order Quarter  \
47                       Europe  Personal Care        2010              4
23         Australia and Oceania         Fruits        2014              3
10                          Asia     Vegetables        2011              2
71  Middle East and North Africa         Fruits        2015              3
88  Middle East and North Africa         Fruits        2012              2

    Order Month  Units Sold  Total Revenue
47           12         273       22312.29
23            9        2187       20404.71
10            6         124       19103.44
71            8         673        6279.09
88            4         522        4870.26
```

[26]:
```
Year_wise = sales1.sort_values(by=['Order Month'])
```

[28]:
```
Year_2010 =Year_wise[(Year_wise['Order Year'] == 2010)]
```

[29]:
```
Year_2010
```

[29]:
```
                         Region       Country        Item Type Sales Channel  \
50                       Europe       Albania          Clothes        Online
41  Middle East and North Africa    Azerbaijan        Cosmetics        Online
0          Australia and Oceania        Tuvalu        Baby Food       Offline
27            Sub-Saharan Africa          Mali           Fruits        Online
56         Australia and Oceania          Fiji          Clothes       Offline
68                       Europe     Lithuania  Office Supplies       Offline
63  Middle East and North Africa         Libya          Clothes       Offline
93                       Europe       Romania        Cosmetics        Online
47                       Europe   Switzerland    Personal Care        Online
18                         Asia  Turkmenistan        Household       Offline

   Order Priority  Order Date  Order Year  Order Quarter  Order Month  \
50              C    2/2/2010        2010              1            2
41              M    2/6/2010        2010              1            2
0               H   5/28/2010        2010              2            5
```

```
27              L    5/7/2010        2010              2            5
56              C   6/30/2010        2010              2            6
68              H  10/24/2010        2010              4           10
63              H  10/30/2010        2010              4           10
93              H  11/26/2010        2010              4           11
47              M  12/23/2010        2010              4           12
18              L  12/30/2010        2010              4           12

     Order ID   Ship Date  Units Sold  Unit Price  Unit Cost  Total Revenue  \
50  385383069   3/18/2010        2269      109.28      35.84      247956.32
41  382392299   2/25/2010        7234      437.20     263.33     3162704.80
0   669165933   6/27/2010        9925      255.28     159.42     2533654.00
27  686048400   5/10/2010        5822        9.33       6.92       54319.26
56  647876489    8/1/2010        9905      109.28      35.84     1082418.40
68  166460740  11/17/2010        8287      651.21     524.96     5396577.27
63  705784308  11/17/2010        6116      109.28      35.84      668356.48
93  660643374  12/25/2010        7910      437.20     263.33     3458252.00
47  617667090   1/31/2011         273       81.73      56.67       22312.29
18  441619336   1/20/2011        3830      668.27     502.54     2559474.10

    Total Cost  Total Profit
50    81320.96     166635.36
41  1904929.22    1257775.58
0   1582243.50     951410.50
27    40288.24      14031.02
56   354995.20     727423.20
68  4350343.52    1046233.75
63   219197.44     449159.04
93  2082940.30    1375311.70
47    15470.91       6841.38
18  1924728.20     634745.90
```

[30]:
```
Year_2012 =Year_wise[(Year_wise['Order Year'] == 2012)]
```

[31]:
```
Year_2012
```

[31]:
```
                              Region         Country        Item Type  \
49                 Sub-Saharan Africa      Mauritania  Office Supplies
58                             Europe  United Kingdom       Household
85                      North America          Mexico    Personal Care
39                             Europe        Bulgaria  Office Supplies
99                 Sub-Saharan Africa      Mozambique       Household
84                 Sub-Saharan Africa           Kenya      Vegetables
38                               Asia          Brunei  Office Supplies
15                             Europe        Bulgaria         Clothes
88       Middle East and North Africa          Kuwait          Fruits
78                             Europe          Monaco       Baby Food
```

29	Sub-Saharan Africa	The Gambia	Household
55	Sub-Saharan Africa	Cote d'Ivoire	Clothes
92	Middle East and North Africa	Azerbaijan	Office Supplies
42	Sub-Saharan Africa	The Gambia	Cereal
66	Sub-Saharan Africa	Gabon	Personal Care
7	Sub-Saharan Africa	Burkina Faso	Vegetables
19	Australia and Oceania	East Timor	Meat
1	Central America and the Caribbean	Grenada	Cereal
81	Middle East and North Africa	Lebanon	Clothes
30	Europe	Switzerland	Cosmetics
80	Europe	Spain	Household
43	Europe	Slovakia	Vegetables

	Sales Channel	Order Priority	Order Date	Order Year	Order Quarter
49	Offline	C	1/11/2012	2012	1
58	Online	L	1/5/2012	2012	1
85	Offline	L	2/17/2012	2012	1
39	Online	M	2/16/2012	2012	1
99	Offline	L	2/10/2012	2012	1
84	Online	L	3/18/2012	2012	1
38	Online	L	4/1/2012	2012	2
15	Online	M	4/23/2012	2012	2
88	Online	M	4/30/2012	2012	2
78	Offline	H	5/29/2012	2012	2
29	Offline	L	5/26/2012	2012	2
55	Online	C	6/8/2012	2012	2
92	Online	M	6/13/2012	2012	2
42	Offline	H	6/7/2012	2012	2
66	Offline	L	7/8/2012	2012	3
7	Online	H	7/17/2012	2012	3
19	Online	L	7/31/2012	2012	3
1	Online	C	8/22/2012	2012	3
81	Online	L	9/18/2012	2012	3
30	Offline	M	9/17/2012	2012	3
80	Offline	L	10/21/2012	2012	4
43	Online	H	10/6/2012	2012	4

	Order Month	Order ID	Ship Date	Units Sold	Unit Price	Unit Cost
49	1	837559306	1/13/2012	1266	651.21	524.96
58	1	955357205	2/14/2012	282	668.27	502.54
85	2	430915820	3/20/2012	6422	81.73	56.67
39	2	189965903	2/28/2012	3987	651.21	524.96
99	2	665095412	2/15/2012	5367	668.27	502.54
84	3	827844560	4/7/2012	6457	154.06	90.93
38	4	320009267	5/8/2012	6708	651.21	524.96
15	4	972292029	6/3/2012	1673	109.28	35.84
88	4	513417565	5/18/2012	522	9.33	6.92

78	5	688288152	6/2/2012	8614	255.28	159.42
29	5	886494815	6/9/2012	2370	668.27	502.54
55	6	114606559	6/27/2012	3482	109.28	35.84
92	6	423331391	7/24/2012	2021	651.21	524.96
42	6	994022214	6/8/2012	2117	205.70	117.11
66	7	228944623	7/9/2012	8656	81.73	56.67
7	7	871543967	7/27/2012	8082	154.06	90.93
19	7	322067916	9/11/2012	5908	421.89	364.69
1	8	963881480	9/15/2012	2804	205.70	117.11
81	9	663110148	10/8/2012	7884	109.28	35.84
30	9	249693334	10/20/2012	8661	437.20	263.33
80	10	213487374	11/30/2012	4513	668.27	502.54
43	10	759224212	11/10/2012	171	154.06	90.93

	Total Revenue	Total Cost	Total Profit
49	824431.86	664599.36	159832.50
58	188452.14	141716.28	46735.86
85	524870.06	363934.74	160935.32
39	2596374.27	2093015.52	503358.75
99	3586605.09	2697132.18	889472.91
84	994765.42	587135.01	407630.41
38	4368316.68	3521431.68	846885.00
15	182825.44	59960.32	122865.12
88	4870.26	3612.24	1258.02
78	2198981.92	1373243.88	825738.04
29	1583799.90	1191019.80	392780.10
55	380512.96	124794.88	255718.08
92	1316095.41	1060944.16	255151.25
42	435466.90	247921.87	187545.03
66	707454.88	490535.52	216919.36
7	1245112.92	734896.26	510216.66
19	2492526.12	2154588.52	337937.60
1	576782.80	328376.44	248406.36
81	861563.52	282562.56	579000.96
30	3786589.20	2280701.13	1505888.07
80	3015902.51	2267963.02	747939.49
43	26344.26	15549.03	10795.23

[32]:
```
Year_2013 =Year_wise[(Year_wise['Order Year'] == 2013)]
```

[33]:
```
Year_2013
```

[33]:

	Region	Country	Item Type	\
4	Sub-Saharan Africa	Rwanda	Office Supplies	
52	Middle East and North Africa	Saudi Arabia	Cereal	
70	Asia	Turkmenistan	Office Supplies	
60	Australia and Oceania	Australia	Cereal	

61	Europe	San Marino	Baby Food	
74	Middle East and North Africa	Pakistan	Cosmetics	
79	Australia and Oceania	Samoa	Cosmetics	
51	Sub-Saharan Africa	Lesotho	Fruits	
54	Sub-Saharan Africa	Sao Tome and Principe	Fruits	
65	Sub-Saharan Africa	Rwanda	Cosmetics	
64	Central America and the Caribbean	Haiti	Cosmetics	
31	Sub-Saharan Africa	South Sudan	Personal Care	

	Sales Channel	Order Priority	Order Date	Order Year	Order Quarter	\
4	Offline	L	2/1/2013	2013	1	
52	Online	M	3/25/2013	2013	1	
70	Online	M	4/23/2013	2013	2	
60	Offline	H	6/9/2013	2013	2	
61	Online	L	6/26/2013	2013	2	
74	Offline	L	7/5/2013	2013	3	
79	Online	H	7/20/2013	2013	3	
51	Online	L	8/18/2013	2013	3	
54	Offline	H	9/17/2013	2013	3	
65	Offline	H	10/11/2013	2013	4	
64	Offline	H	10/13/2013	2013	4	
31	Offline	C	12/29/2013	2013	4	

	Order Month	Order ID	Ship Date	Units Sold	Unit Price	Unit Cost	\
4	2	115456712	2/6/2013	5062	651.21	524.96	
52	3	844530045	3/28/2013	4063	205.70	117.11	
70	4	462405812	5/20/2013	5010	651.21	524.96	
60	6	450563752	7/2/2013	682	205.70	117.11	
61	6	569662845	7/1/2013	4750	255.28	159.42	
74	7	231145322	8/16/2013	9892	437.20	263.33	
79	7	670854651	8/7/2013	9654	437.20	263.33	
51	8	918419539	9/18/2013	9606	9.33	6.92	
54	9	508980977	10/24/2013	7637	9.33	6.92	
65	10	699358165	11/25/2013	4477	437.20	263.33	
64	10	505716836	11/16/2013	1705	437.20	263.33	
31	12	406502997	1/28/2014	2125	81.73	56.67	

	Total Revenue	Total Cost	Total Profit
4	3296425.02	2657347.52	639077.50
52	835759.10	475817.93	359941.17
70	3262562.10	2630049.60	632512.50
60	140287.40	79869.02	60418.38
61	1212580.00	757245.00	455335.00
74	4324782.40	2604860.36	1719922.04
79	4220728.80	2542187.82	1678540.98
51	89623.98	66473.52	23150.46
54	71253.21	52848.04	18405.17

```
65     1957344.40  1178928.41     778415.99
64      745426.00   448977.65     296448.35
31      173676.25   120423.75      53252.50
```

[34]:
```
Year_2014 =Year_wise[(Year_wise['Order Year'] == 2014)]
```

[35]:
```
Year_2014
```

[35]:
```
                  Region                          Country        Item Type  \
87     Sub-Saharan Africa                       The Gambia        Baby Food
14                   Asia                         Mongolia    Personal Care
9      Sub-Saharan Africa                          Senegal           Cereal
59     Sub-Saharan Africa                         Djibouti        Cosmetics
2                  Europe                           Russia  Office Supplies
20                 Europe                           Norway        Baby Food
3      Sub-Saharan Africa           Sao Tome and Principe           Fruits
91  Australia and Oceania                        Australia        Beverages
28                 Europe                           Norway        Beverages
11     Sub-Saharan Africa                       Cape Verde          Clothes
23  Australia and Oceania                      New Zealand           Fruits
26  Australia and Oceania                         Kiribati           Fruits
76  Australia and Oceania  Federated States of Micronesia        Beverages
48                 Europe                         Macedonia          Clothes
75          North America                           Mexico        Household

   Sales Channel Order Priority  Order Date  Order Year  Order Quarter  \
87       Offline              M    2/3/2014        2014              1
14       Offline              C   2/19/2014        2014              1
9         Online              H   4/18/2014        2014              2
59       Offline              H    4/7/2014        2014              2
2        Offline              L    5/2/2014        2014              2
20        Online              L   5/14/2014        2014              2
3         Online              C   6/20/2014        2014              2
91       Offline              H    7/7/2014        2014              3
28       Offline              C   7/18/2014        2014              3
11       Offline              H    8/2/2014        2014              3
23        Online              H    9/8/2014        2014              3
26        Online              M  10/13/2014        2014              4
76        Online              C  10/28/2014        2014              4
48       Offline              C  10/14/2014        2014              4
75       Offline              C   11/6/2014        2014              4

    Order Month   Order ID   Ship Date  Units Sold  Unit Price  Unit Cost  \
87            2  494747245   3/20/2014        5559      255.28     159.42
14            2  832401311   2/23/2014        4901       81.73      56.67
9             4  616607081   5/30/2014        6593      205.70     117.11
59            4  259353148   4/19/2014        7215      437.20     263.33
```

```
2              5  341417157    5/8/2014        1779      651.21      524.96
20             5  819028031   6/28/2014        7450      255.28      159.42
3              6  514321792    7/5/2014        8102        9.33        6.92
91             7  240470397   7/11/2014        9389       47.45       31.79
28             7  435608613   7/30/2014        5124       47.45       31.79
11             8  939825713   8/19/2014        4168      109.28       35.84
23             9  142278373   10/4/2014        2187        9.33        6.92
26            10  347140347  11/10/2014        5398        9.33        6.92
76            10  217221009  11/15/2014        9379       47.45       31.79
48            10  787399423  11/14/2014        7842      109.28       35.84
75            11  986435210  12/12/2014        6954      668.27      502.54

    Total Revenue  Total Cost  Total Profit
87     1419101.52   886215.78     532885.74
14      400558.73   277739.67     122819.06
9      1356180.10   772106.23     584073.87
59     3154398.00  1899925.95    1254472.05
2      1158502.59   933903.84     224598.75
20     1901836.00  1187679.00     714157.00
3        75591.66    56065.84      19525.82
91      445508.05   298476.31     147031.74
28      243133.80   162891.96      80241.84
11      455479.04   149381.12     306097.92
23       20404.71    15134.04       5270.67
26       50363.34    37354.16      13009.18
76      445033.55   298158.41     146875.14
48      856973.76   281057.28     575916.48
75     4647149.58  3494663.16    1152486.42
```

[36]:
```
Year_2015 =Year_wise[(Year_wise['Order Year'] == 2015)]
```

[37]:
```
Year_2015
```

[37]:
```
                          Region                Country        Item Type  \
33                          Asia                Myanmar        Household
57                        Europe                Austria        Cosmetics
5          Australia and Oceania        Solomon Islands        Baby Food
69            Sub-Saharan Africa             Madagascar          Clothes
17            Sub-Saharan Africa               Cameroon        Beverages
21                        Europe               Portugal        Baby Food
8             Sub-Saharan Africa  Republic of the Congo    Personal Care
98                 North America                 Mexico    Personal Care
71  Middle East and North Africa                  Libya           Fruits
32         Australia and Oceania              Australia  Office Supplies
44                          Asia                Myanmar          Clothes

   Sales Channel Order Priority  Order Date  Order Year  Order Quarter  \
```

```
33       Offline              H   1/16/2015        2015              1
57       Offline              H   2/23/2015        2015              1
5         Online              C    2/4/2015        2015              1
69       Offline              L   4/25/2015        2015              2
17       Offline              C    4/1/2015        2015              2
21        Online              H   7/31/2015        2015              3
8        Offline              M   7/14/2015        2015              3
98       Offline              M   7/30/2015        2015              3
71        Online              L   8/14/2015        2015              3
32        Online              C  10/27/2015        2015              4
44        Online              H  11/14/2015        2015              4

    Order Month   Order ID   Ship Date  Units Sold  Unit Price  Unit Cost  \
33            1  177713572    3/1/2015        8250      668.27     502.54
57            2  868214595    3/2/2015        2847      437.20     263.33
5             2  547995746   2/21/2015        2974      255.28     159.42
69            4  610425555   5/28/2015        7342      109.28      35.84
17            4  519820964   4/18/2015        5430       47.45      31.79
21            7  860673511    9/3/2015        1273      255.28     159.42
8             7  770463311   8/25/2015        6070       81.73      56.67
98            7  559427106    8/8/2015        5767       81.73      56.67
71            8  816200339   9/30/2015         673        9.33       6.92
32           10  158535134  11/25/2015        2924      651.21     524.96
44           11  223359620  11/18/2015        5930      109.28      35.84

    Total Revenue  Total Cost  Total Profit
33     5513227.50  4145955.00    1367272.50
57     1244708.40   749700.51     495007.89
5       759202.72   474115.08     285087.64
69      802333.76   263137.28     539196.48
17      257653.50   172619.70      85033.80
21      324971.44   202941.66     122029.78
8       496101.10   343986.90     152114.20
98      471336.91   326815.89     144521.02
71        6279.09     4657.16       1621.93
32     1904138.04  1534983.04     369155.00
44      648030.40   212531.20     435499.20
```

[38]:
```
Year_2016 =Year_wise[(Year_wise['Order Year'] == 2016)]
```

[39]:
```
Year_2016
```

[39]:
```
                              Region       Country        Item Type  \
45                 Sub-Saharan Africa       Comoros           Cereal
24                             Europe      Moldova     Personal Care
97                 Sub-Saharan Africa  Sierra Leone       Vegetables
22  Central America and the Caribbean      Honduras           Snacks
```

```
67  Central America and the Caribbean        Belize          Clothes
89                             Europe      Slovenia        Beverages
82       Middle East and North Africa          Iran        Cosmetics
16                               Asia     Sri Lanka        Cosmetics
90                  Sub-Saharan Africa  Sierra Leone  Office Supplies
46                             Europe       Iceland        Cosmetics

   Sales Channel Order Priority  Order Date  Order Year  Order Quarter  \
45       Offline              H   3/29/2016        2016              1
24        Online              L    5/7/2016        2016              2
97       Offline              C    6/1/2016        2016              2
22        Online              L   6/30/2016        2016              2
67       Offline              M   7/25/2016        2016              3
89       Offline              C  10/23/2016        2016              4
82        Online              H  11/15/2016        2016              4
16       Offline              M  11/19/2016        2016              4
90       Offline              H   12/6/2016        2016              4
46        Online              C  12/31/2016        2016              4

    Order Month   Order ID   Ship Date  Units Sold  Unit Price  Unit Cost  \
45            3  902102267   4/29/2016         962      205.70     117.11
24            5  740147912   5/10/2016        5070       81.73      56.67
97            6  728815257   6/29/2016        1485      154.06      90.93
22            6  795490682   7/26/2016        2225      152.58      97.44
67            7  807025039    9/7/2016        5498      109.28      35.84
89           10  345718562  11/25/2016        4660       47.45      31.79
82           11  286959302   12/8/2016        6489      437.20     263.33
16           11  419123971  12/18/2016        6952      437.20     263.33
90           12  621386563  12/14/2016         948      651.21     524.96
46           12  331438481  12/31/2016        8867      437.20     263.33

    Total Revenue  Total Cost  Total Profit
45      197883.40   112659.82      85223.58
24      414371.10   287316.90     127054.20
97      228779.10   135031.05      93748.05
22      339490.50   216804.00     122686.50
67      600821.44   197048.32     403773.12
89      221117.00   148141.40      72975.60
82     2836990.80  1708748.37    1128242.43
16     3039414.40  1830670.16    1208744.24
90      617347.08   497662.08     119685.00
46     3876652.40  2334947.11    1541705.29
```

[40]:
```
Year_2017 =Year_wise[(Year_wise['Order Year'] == 2017)]
```

[41]:
```
Year_2017
```

[41]:

	Region	Country	Item Type	Sales Channel	Order Priority	Order Date	Order Year	Order Quarter	Order Month	Order ID	Ship Date	Units Sold	Unit Price	Unit Cost	Total Revenue	Total Cost	Total Profit
37	Sub-Saharan Africa	The Gambia	Meat	Online	M	1/14/2017	2017	1	1	825304400	1/23/2017	4767	421.89	364.69	2011149.63	1738477.23	272672.40
12	Asia	Bangladesh	Clothes	Online	L	1/13/2017	2017	1	1	187310731	3/1/2017	8263	109.28	35.84	902980.64	296145.92	606834.72
34	Sub-Saharan Africa	Djibouti	Snacks	Online	M	2/25/2017	2017	1	2	756274640	2/25/2017	7327	152.58	97.44	1117953.66	713942.88	404010.78
13	Central America and the Caribbean	Honduras	Household	Offline	H	2/8/2017	2017	1	2	522840487	2/13/2017	8974	668.27	502.54	5997054.98	4509793.96	1487261.02
40	Sub-Saharan Africa	Niger	Personal Care	Online	H	3/11/2017	2017	1	3	699285638	3/28/2017	3015	81.73	56.67	246415.95	170860.05	75555.90
73	Sub-Saharan Africa	Djibouti	Cereal	Online	H	5/20/2017	2017	2	5	555990016	6/17/2017	8656	205.70	117.11	1780539.20	1013704.16	766835.04
25	Europe	France	Cosmetics	Online	H	5/22/2017	2017	2	5	898523128	6/5/2017	1815	437.20	263.33	793518.00	477943.95	315574.05
35	Central America and the Caribbean	Costa Rica	Personal Care	Offline	L	5/8/2017	2017	2	5	456767165	5/21/2017	6409	81.73	56.67	523807.57	363198.03	160609.54

[43]:
```
sales.axes[0]
```

[43]:
```
RangeIndex(start=0, stop=100, step=1)
```

[44]:
```
sales.axes[1]
```

[44]:
```
Index(['Region', 'Country', 'Item Type', 'Sales Channel', 'Order Priority',
       'Order Date', 'Order ID', 'Ship Date', 'Units Sold', 'Unit Price',
       'Unit Cost', 'Total Revenue', 'Total Cost', 'Total Profit',
       'Order Year', 'Order Quarter', 'Order Month'],
      dtype='object')
```

[45]:
```
sales.dtypes
```

[45]:
```
Region             object
Country            object
Item Type          object
Sales Channel      object
Order Priority     object
Order Date         object
Order ID            int64
Ship Date          object
Units Sold          int64
Unit Price        float64
Unit Cost         float64
Total Revenue     float64
Total Cost        float64
Total Profit      float64
Order Year          int32
Order Quarter       int32
Order Month         int32
dtype: object
```

[46]:
```
sales.columns.isnull()
```

[46]:
```
array([False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False])
```

[47]:
```
sales.loc[:,["Total Revenue","Total Profit"]].iloc[:]
```

[47]:

	Total Revenue	Total Profit
0	2533654.00	951410.50
1	576782.80	248406.36
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]
```

[48]:
```
np.corrcoef(sales.loc[:,'Total Revenue'].iloc[:],sales.loc[:,'Total Profit'].
↪iloc[:])
```

[48]:
```
array([[1.        , 0.89732687],
       [0.89732687, 1.        ]])
```

[49]:
```
sales.set_index('Order ID',inplace=True)
```

[52]:
```
np.average(sales['Total Profit'])
```

[52]:
```
441681.98399999994
```

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

[53]:
```
np.max(sales['Total Profit'])
```

[53]:
```
1719922.04
```

[54]:
```
np.min(sales['Total Profit'])
```

[54]:
```
1258.02
```

[55]:
```
np.var(sales['Total Profit'])
```

[55]:
```
190392340968.9648
```

Maximum and minimum profit generated are 1719922.04 and 1258.09 respectively.

[56]:
```
np.max(sales['Total Revenue'])
```

[56]:
```
5997054.98
```

[57]:
```
np.min(sales['Total Revenue'])
```

[57]:
```
4870.26
```

[58]:
```
np.mean(sales['Total Revenue'])
```

[58]:
```
1373487.6831
```

[59]:
```
np.var(sales['Total Revenue'])
```

[59]:
```
2110366986501.2166
```

[60]:
```
np.std(sales['Total Revenue'])
```

[60]: 1452710.2211044075

[61]:
```
np.median(sales['Total Revenue'])
```

[61]: 752314.36

[62]:
```
np.percentile(sales['Total Revenue'],50,axis=0,overwrite_input=True)
```

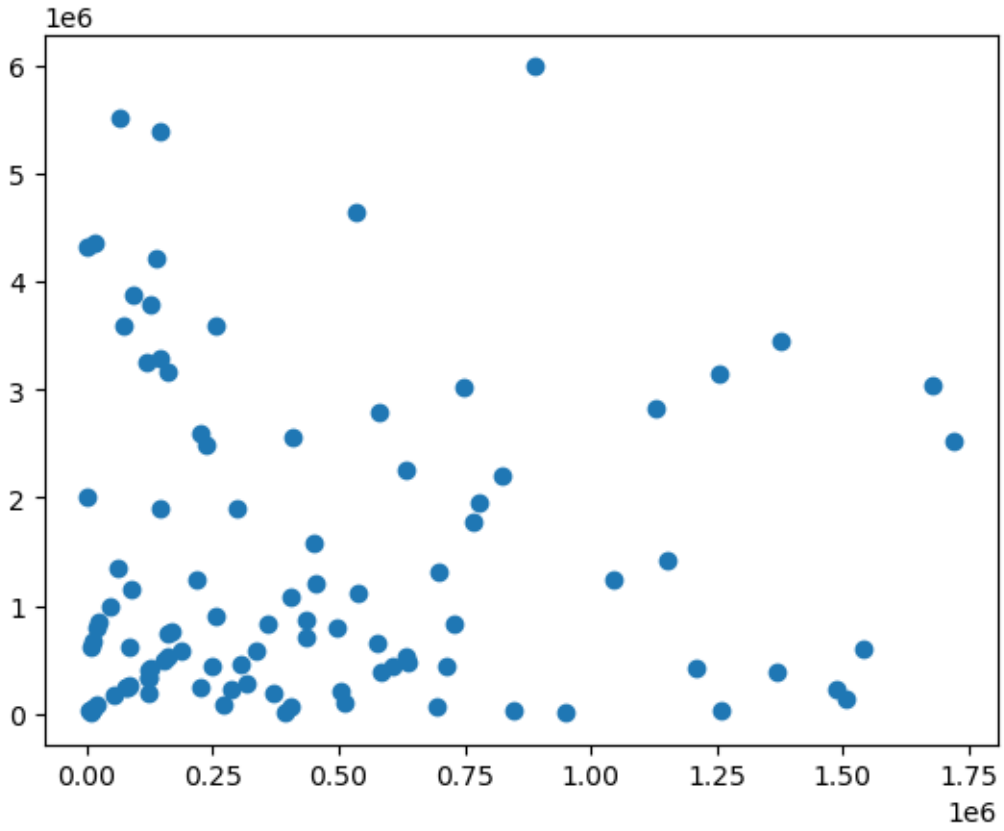
[62]: 752314.36

0.4.3 Maximum and minimum revenue generated by the product are 5997054.98 and 4870.26.

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

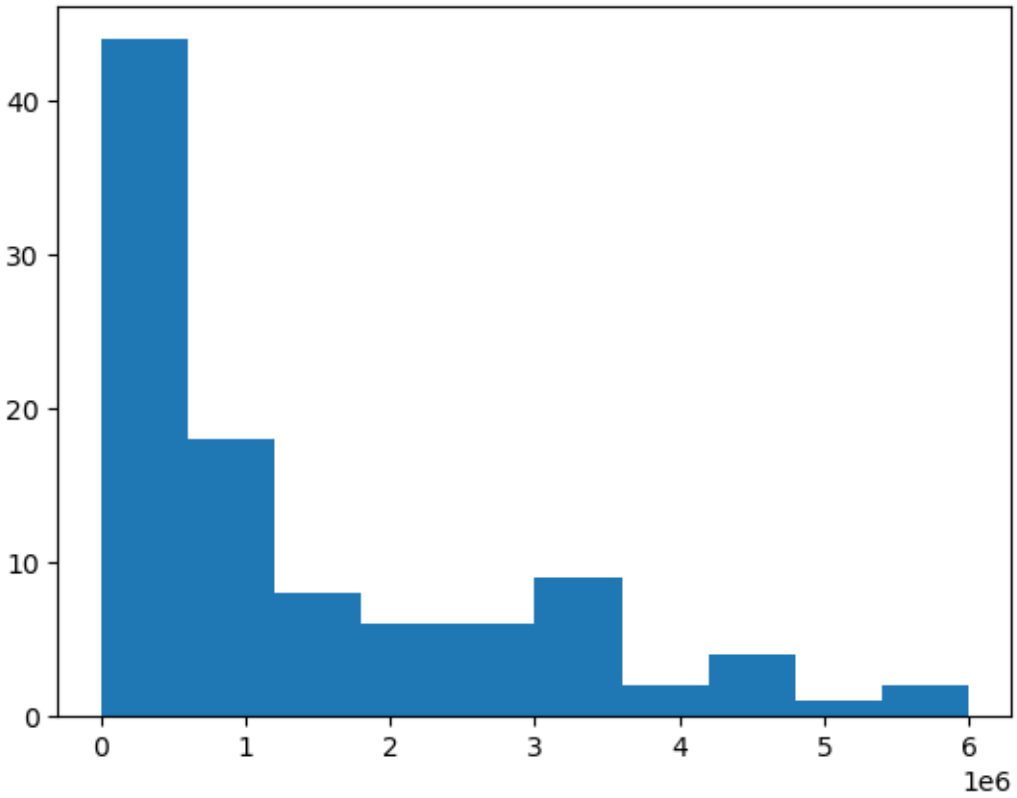
[68]:
```
plt.scatter(sales['Total Profit'],sales['Total Revenue'])
```

[68]: <matplotlib.collections.PathCollection at 0x264b03cbe50>



[70]:
```
plt.hist(sales['Total Revenue'])
```

[70]:
```
(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>)
```



[72]:
```
np.correlate(sales['Total Revenue'],sales['Total Profit'])
```

[72]:
```
array([6.25562019e+13])
```

[73]:
```
sales.head()
```

[73]:

Order ID	Region	Country
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

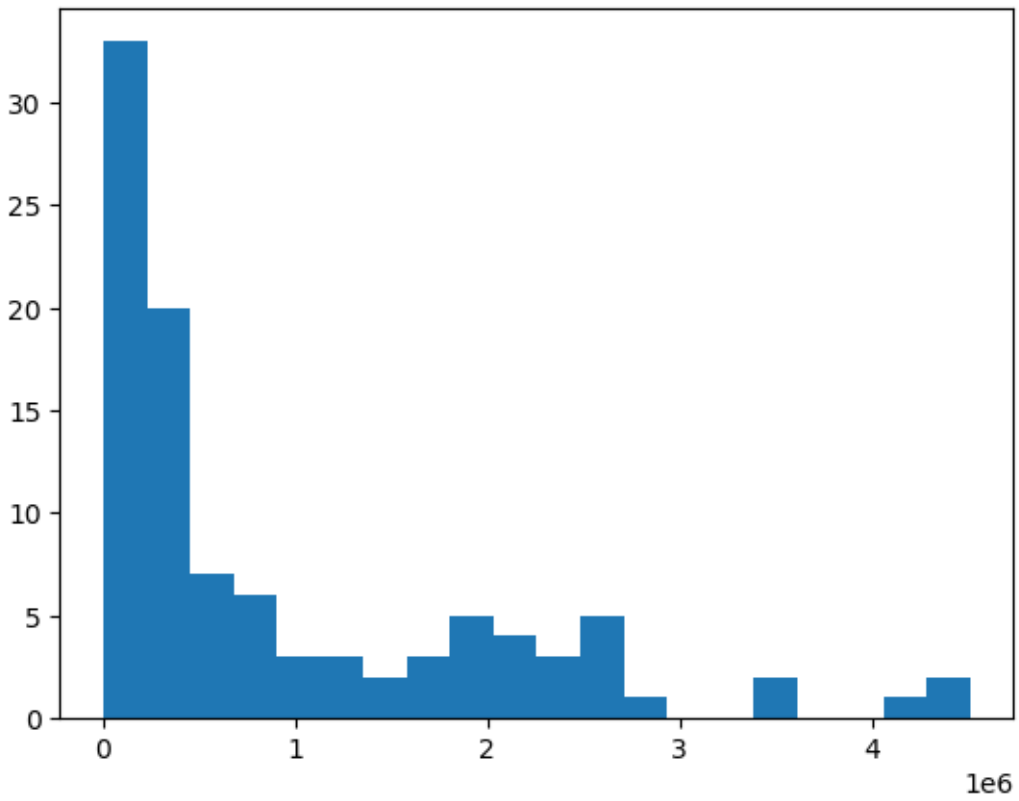
```
                Item Type Sales Channel Order Priority Order Date  Ship Date  \
Order ID
669165933       Baby Food       Offline              H  5/28/2010  6/27/2010
963881480          Cereal        Online              C  8/22/2012  9/15/2012
341417157  Office Supplies       Offline             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

           Units Sold  Unit Price  Unit Cost  Total Revenue  Total Cost  \
Order ID
669165933        9925      255.28     159.42        4870.26  1582243.50
963881480        2804      205.70     117.11      435466.90   328376.44
341417157        1779      651.21     524.96      247956.32   933903.84
514321792        8102        9.33       6.92       75591.66    56065.84
115456712        5062      651.21     524.96      471336.91  2657347.52

           Total Profit  Order Year  Order Quarter  Order Month
Order ID
669165933     951410.50        2010              2            5
963881480     248406.36        2012              3            8
341417157     224598.75        2014              2            5
514321792      19525.82        2014              2            6
115456712     639077.50        2013              1            2
```

[75]:
```
np.histogram(sales['Total Cost'],bins=10)
```

[75]:
```
(array([53, 13,  6,  5,  9,  8,  1,  2,  0,  3], dtype=int64),
 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]))
```

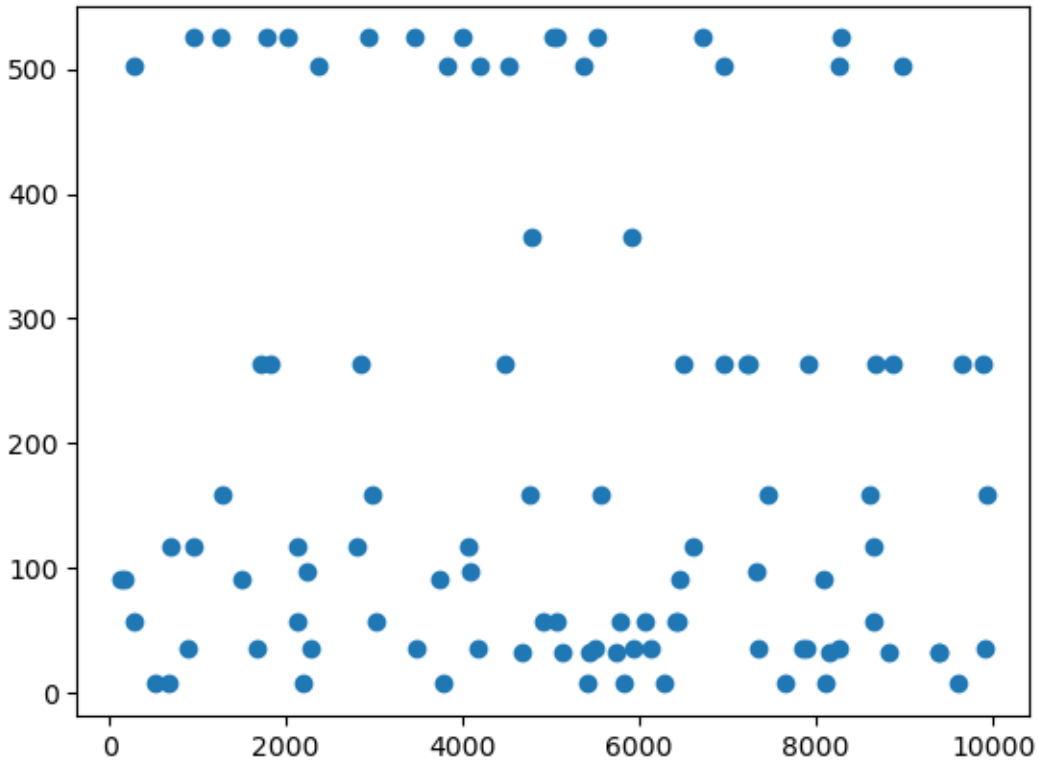
[76]:
```
plt.hist(sales['Total Cost'],bins=20)
```

[76]:
```
(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>)
```



[78]:
```
plt.scatter(sales['Units Sold'],sales['Unit Cost'])
```

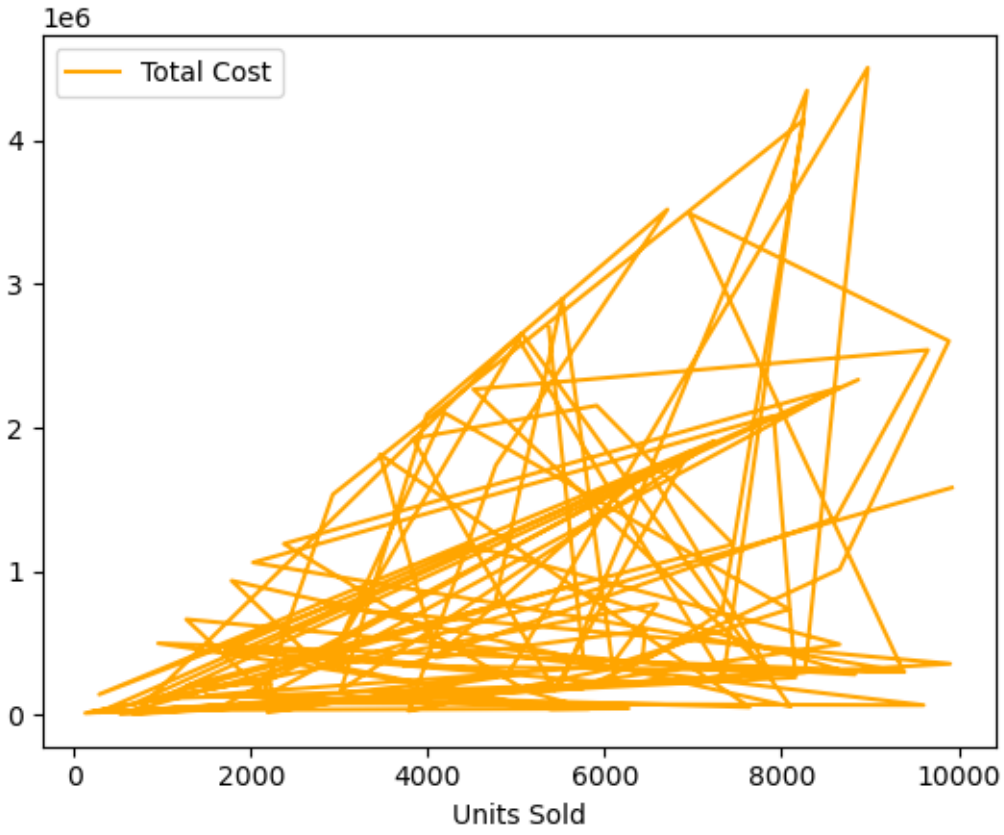
[78]:
```
<matplotlib.collections.PathCollection at 0x264b07c5f50>
```



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.

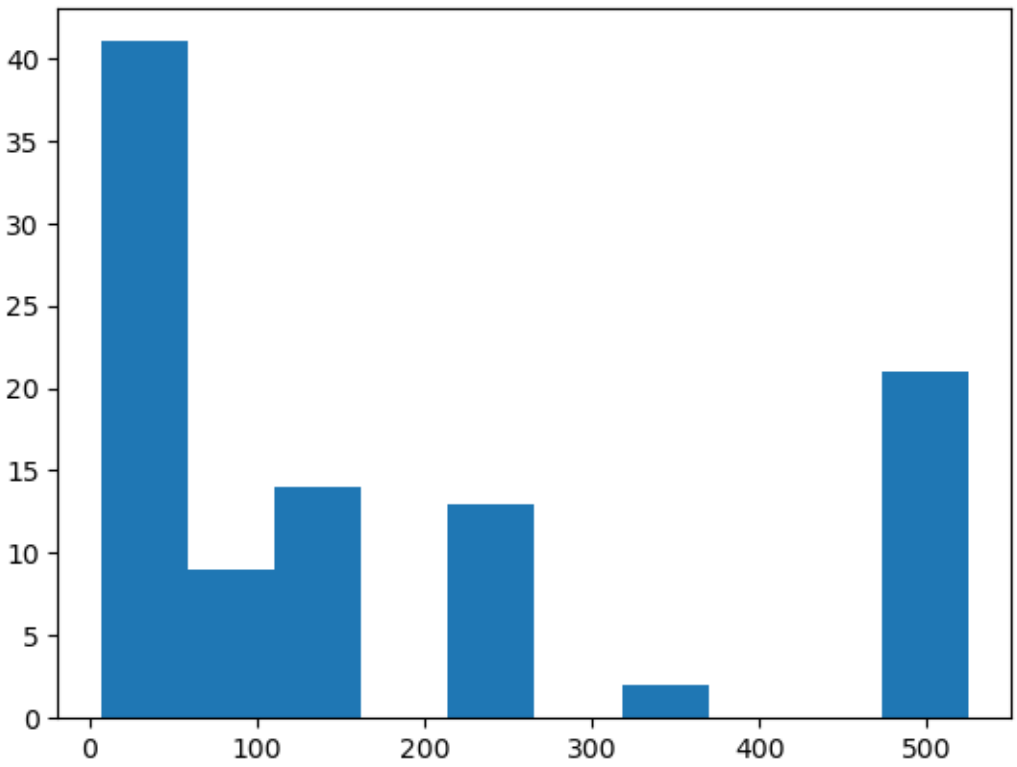
[79]:
```
sales.plot.line(x='Units Sold',y='Total Cost',subplots=True,color={'Total Cost':
↪ 'orange'})
```

[79]:
```
array([<Axes: xlabel='Units Sold'>], dtype=object)
```



[81]:
```
plt.hist(sales['Unit Cost'])
```

[81]:
```
(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>)
```



[82]:
```
np.min(sales['Unit Cost'])
```

[82]: 6.92

[83]:
```
np.max(sales['Unit Cost'])
```

[83]: 524.96

[84]:
```
np.mean(sales['Unit Cost'])
```

[84]: 191.048

[85]:
```
np.std(sales['Unit Cost'])
```

[85]: 187.2647759029979

[86]:
```
np.var(sales['Unit Cost'])
```

[86]: 35068.096294000024

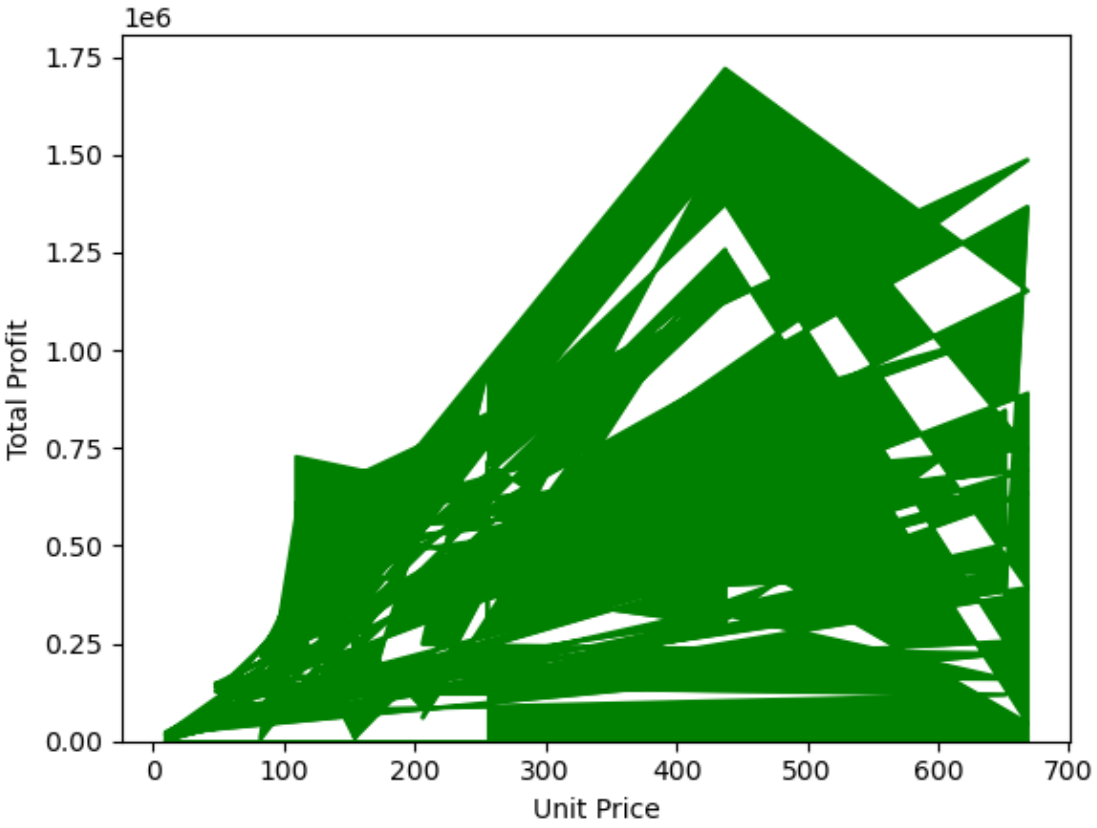
[87]:
```
np.median(sales['Unit Cost'])
```

[87]: 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.

[90]:
```
area_plot = sales.plot.area(x='Unit Price',y='Total␣
↪Profit',color='green',stacked=True,legend=None)
plt.ylabel('Total Profit')
```

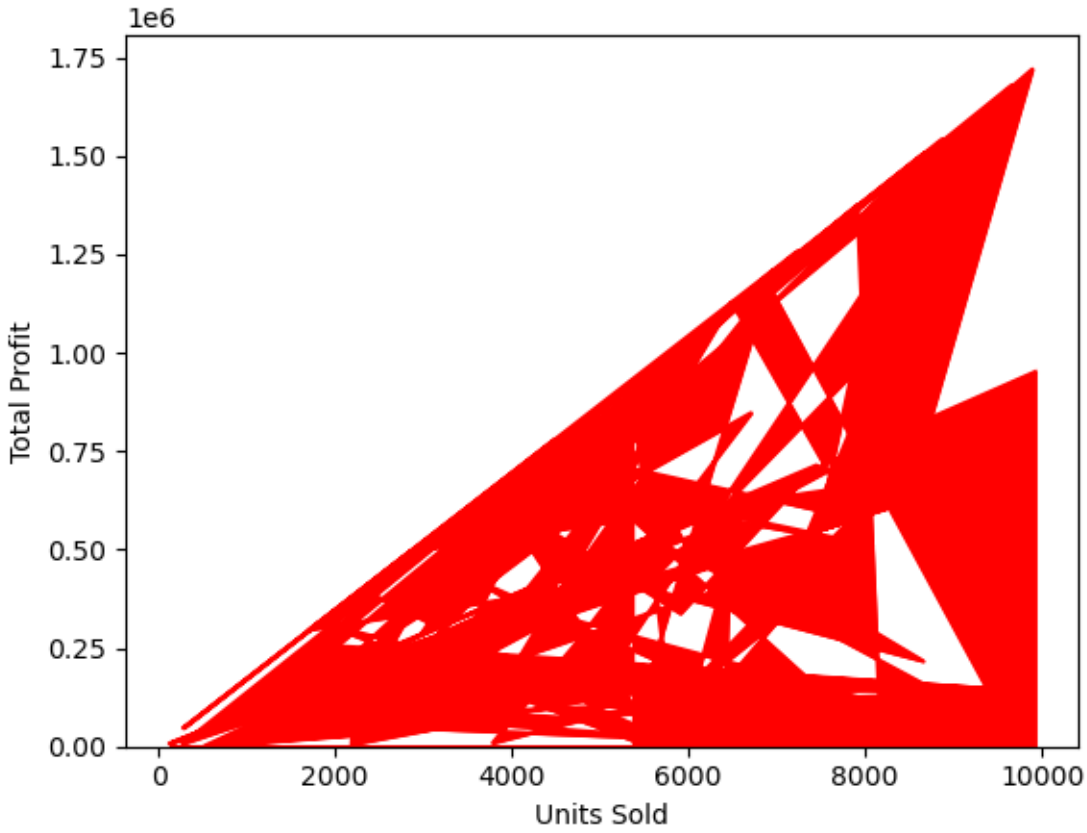
[90]: Text(0, 0.5, 'Total Profit')



0.4.5 Maximum profit has been generated in the unit price range of 400- 500.

[92]:
```
sales.plot.area(x='Units Sold',y='Total Profit',color='red',legend=None)
plt.ylabel('Total Profit')
```

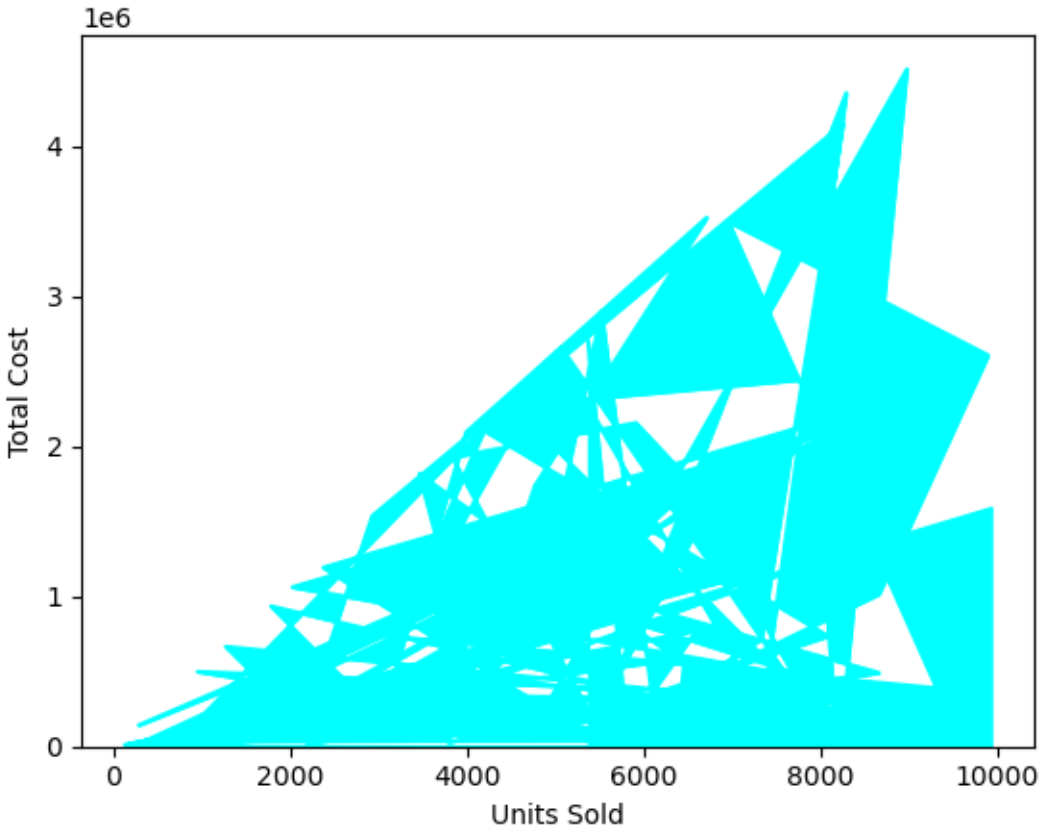
[92]: Text(0, 0.5, 'Total Profit')



0.4.6 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.

[93]:
```
sales.plot.area(x='Units Sold',y='Total Cost',color='aqua',legend=None)
plt.ylabel('Total Cost')
```

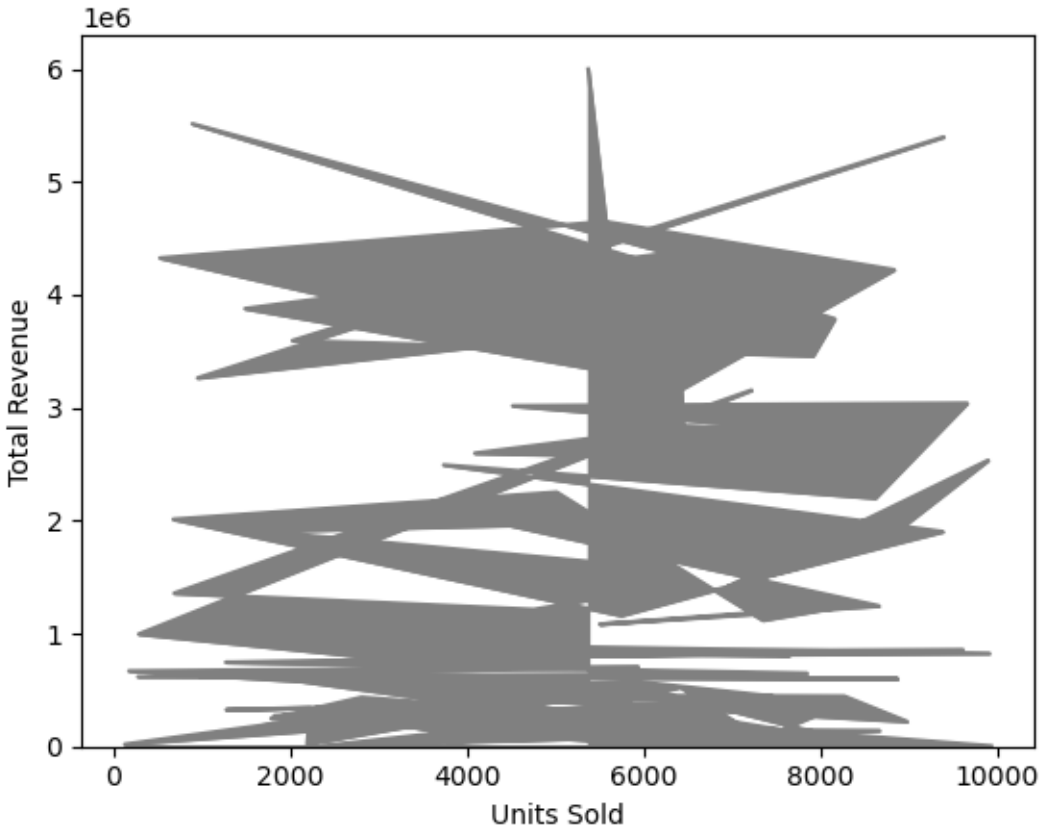
[93]:
```
Text(0, 0.5, 'Total Cost')
```



0.4.7 Maximum cost has been generated when 8000-9000 units were sold.

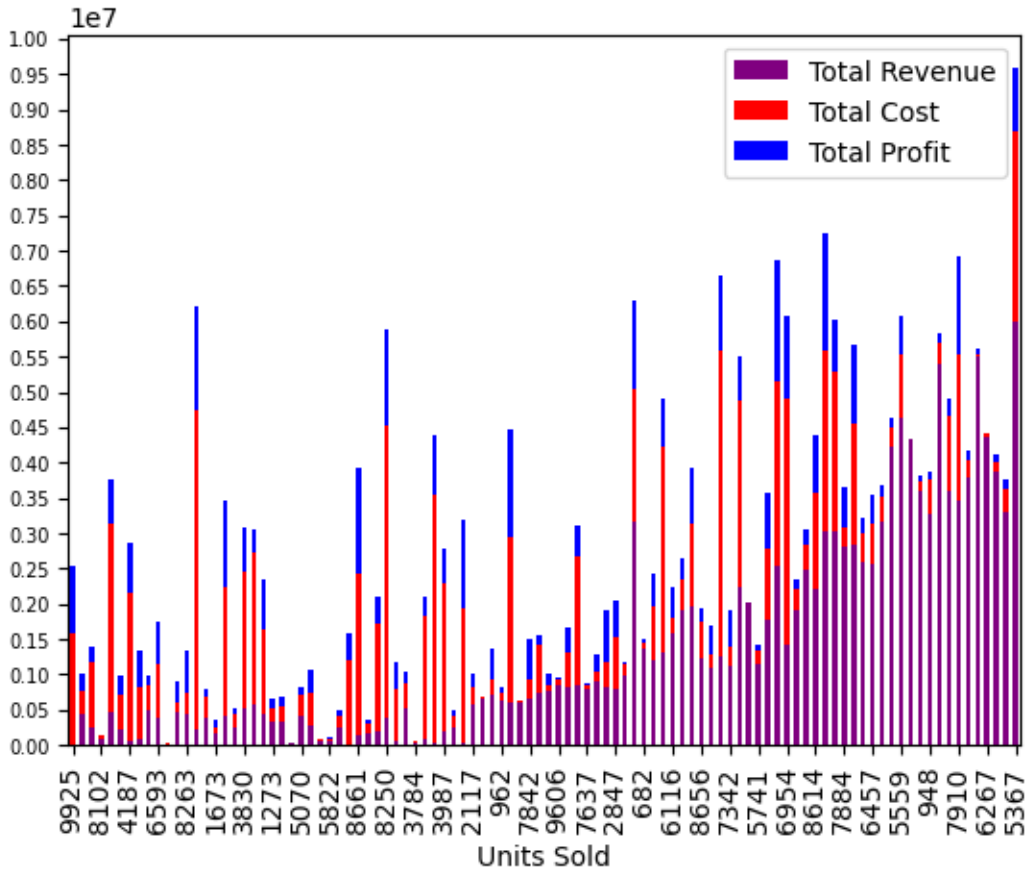
[95]:
```
sales.plot.area(x='Units Sold',y='Total Revenue',color='grey',legend=None)
plt.ylabel('Total Revenue')
```

[95]:
```
Text(0, 0.5, 'Total Revenue')
```



0.4.8 Maximum revenue has been generated when 5000-6500 units of a product were sold.

[99]:
```
bar_plot = sales.plot.bar(x='Units Sold',y=['Total Revenue','Total Cost','Total␣
↪Profit'],color=['purple','red','blue'],stacked=True,rot=True)
plt.xticks(rotation=90)
plt.locator_params(nbins=38)
plt.tick_params(axis='y', which='major', labelsize=7)
```



[100]:
```
sales.head()
```

[100]:

Order ID	Region	Country
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

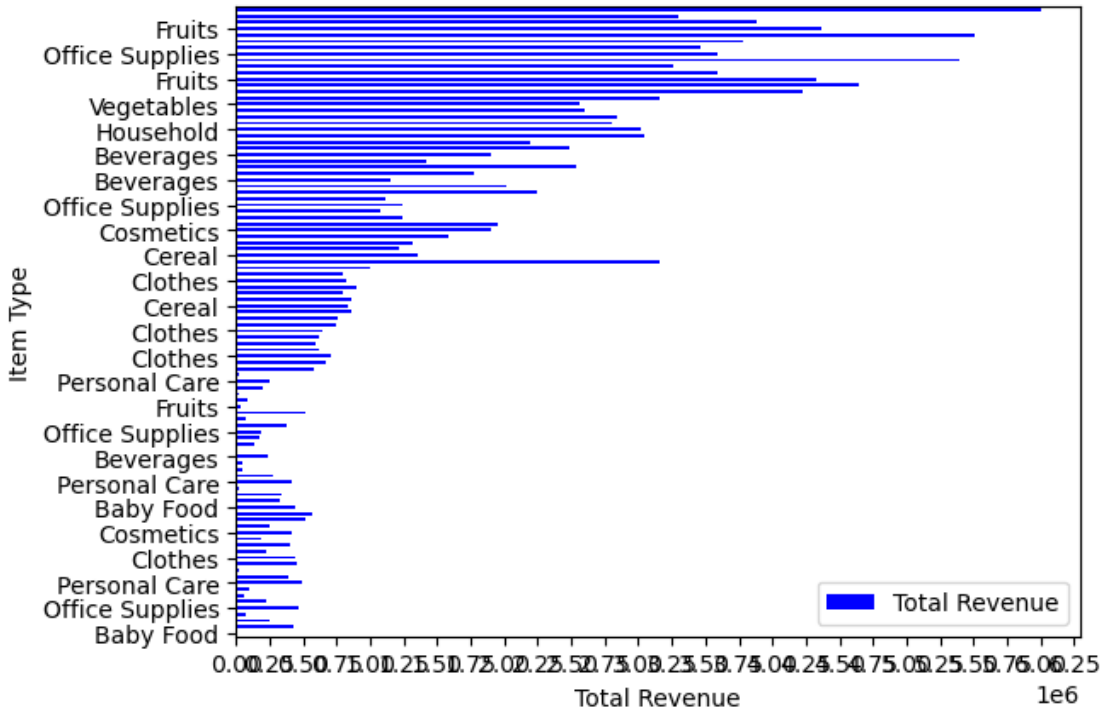
Order ID	Item Type	Sales Channel	Order Priority	Order Date	Ship Date
669165933	Baby Food	Offline	H	5/28/2010	6/27/2010
963881480	Cereal	Online	C	8/22/2012	9/15/2012
341417157	Office Supplies	Offline	L	5/2/2014	5/8/2014
514321792	Fruits	Online	C	6/20/2014	7/5/2014
115456712	Office Supplies	Offline	L	2/1/2013	2/6/2013

Order ID	Units Sold	Unit Price	Unit Cost	Total Revenue	Total Cost

```
Order ID
669165933        9925      255.28     159.42        4870.26  1582243.50
963881480        2804      205.70     117.11      435466.90   328376.44
341417157        1779      651.21     524.96      247956.32   933903.84
514321792        8102        9.33       6.92       75591.66    56065.84
115456712        5062      651.21     524.96      471336.91  2657347.52

            Total Profit  Order Year  Order Quarter  Order Month
Order ID
669165933     951410.50        2010              2            5
963881480     248406.36        2012              3            8
341417157     224598.75        2014              2            5
514321792      19525.82        2014              2            6
115456712     639077.50        2013              1            2
```

[102]:
```
sales.plot.barh(x='Item Type',y='Total Revenue',color='blue')
plt.locator_params(nbins=28)
plt.xlabel('Total Revenue')
```

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



[104]:
```
sales['Item Type'].unique()
```

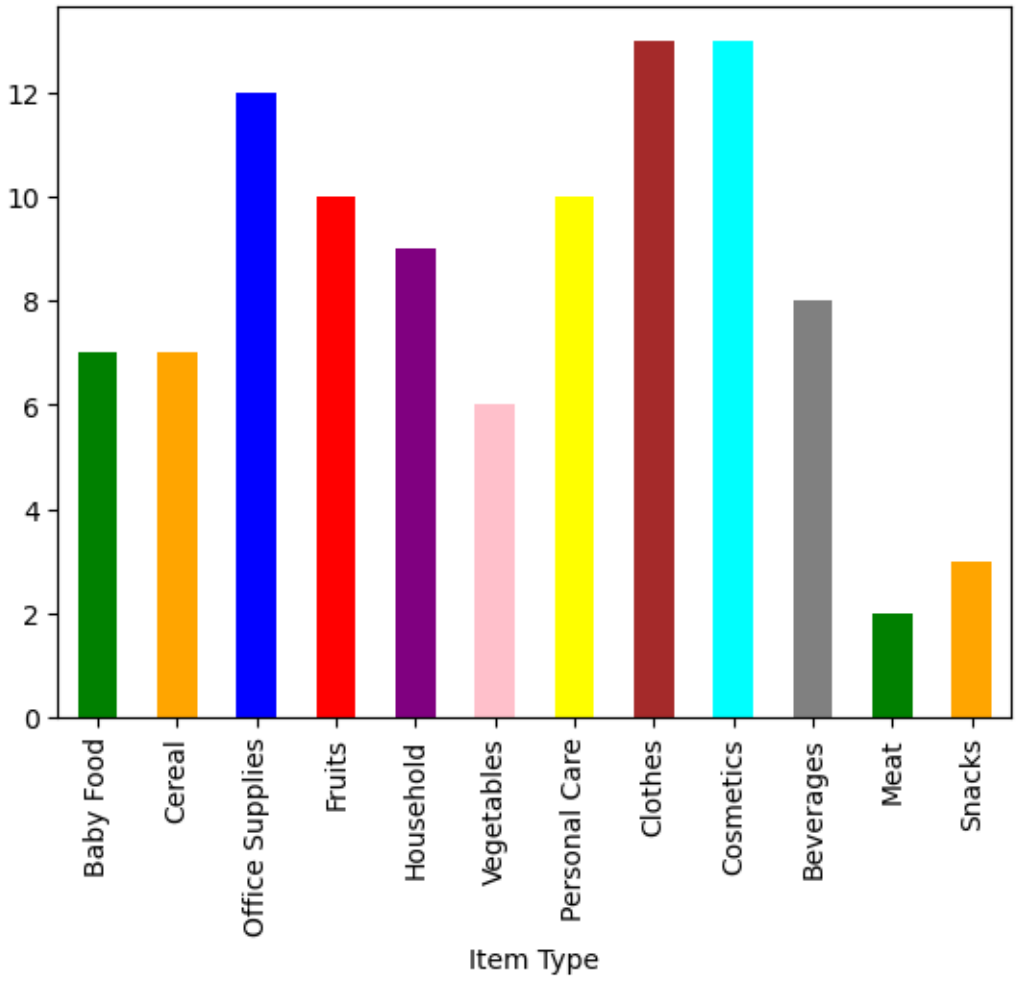
[104]:
```
array(['Baby Food', 'Cereal', 'Office Supplies', 'Fruits', 'Household',
       'Vegetables', 'Personal Care', 'Clothes', 'Cosmetics', 'Beverages',
       'Meat', 'Snacks'], dtype=object)
```

[105]:
```
items = ['Baby Food', 'Cereal', 'Office Supplies', 'Fruits', 'Household',
       'Vegetables', 'Personal Care', 'Clothes', 'Cosmetics', 'Beverages',
       'Meat', 'Snacks']
```

[107]:
```
sales['Item Type'] = pd.Categorical(sales['Item␣
↪Type'],categories=items,ordered=True)
```

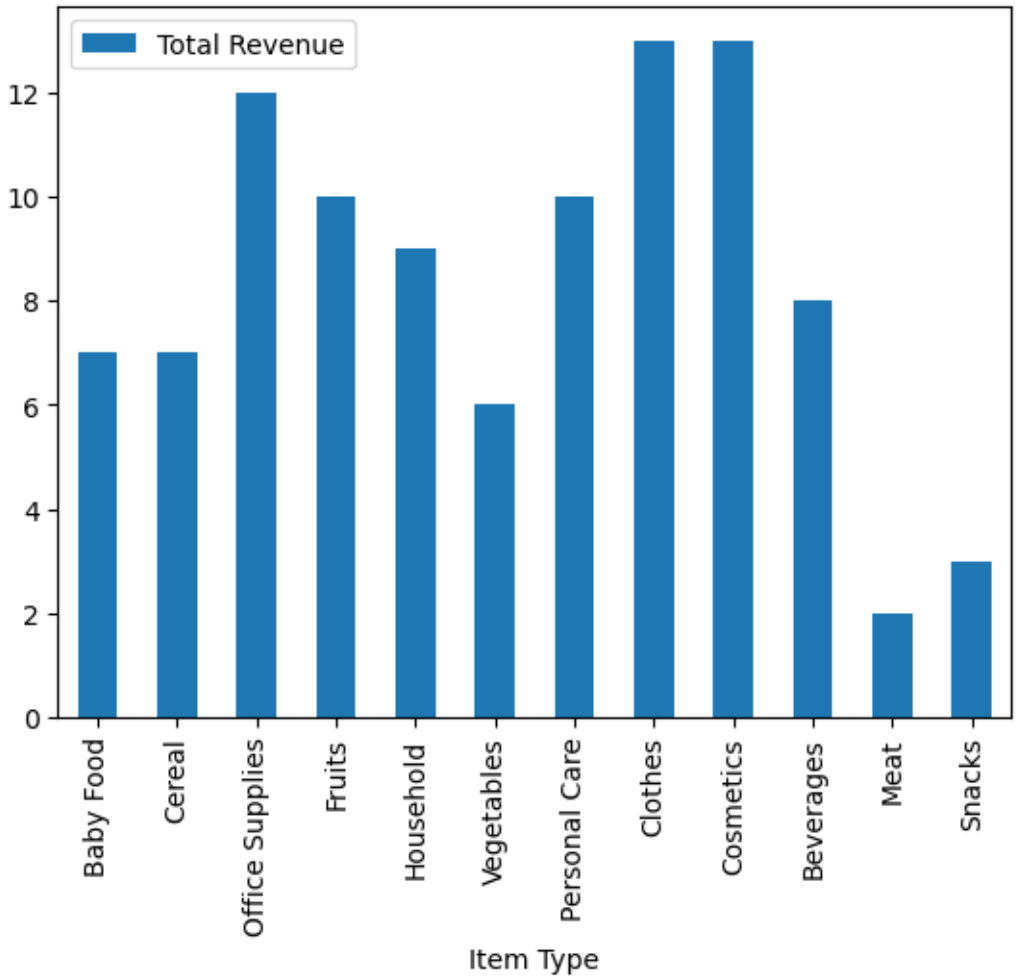
[108]:
```
sales.groupby('Item Type')['Total Revenue'].count().
↪plot(kind='bar',color=['green','orange','blue','red','purple','pink','yellow','brown','aqua
```

[108]:
```
<Axes: xlabel='Item Type'>
```



[109]:
```
pd.pivot_table(sales,values='Total Revenue',index='Item Type',aggfunc='count').
  ↪plot(kind='bar')
```

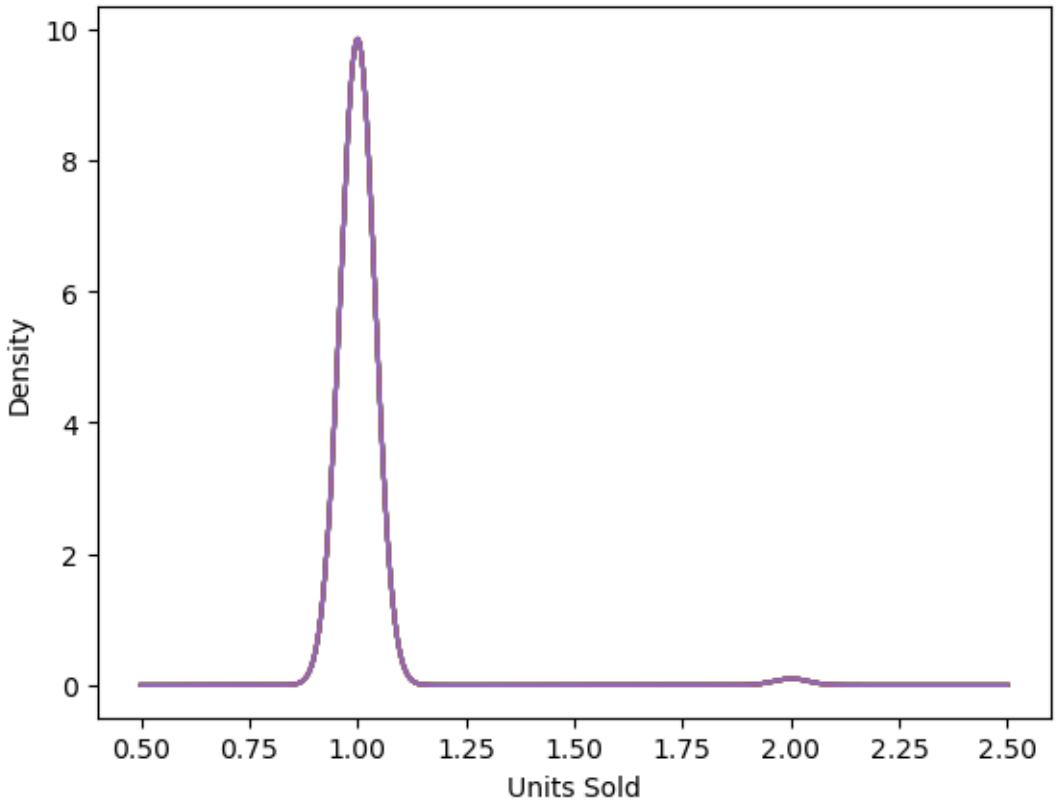
[109]:
```
<Axes: xlabel='Item Type'>
```



0.4.9 Maximum revenue has been generated from the items 'Clothes' and 'Cosmetics' closely followed by 'Office Supplies'.

[111]:
```
#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,index='Units Sold',aggfunc='count').
  ↪plot(kind='kde',stacked=True,legend=None)
plt.xlabel('Units Sold')
```

[111]:
```
Text(0.5, 0, 'Units Sold')
```



[112]:
```
sales.head()
```

[112]:
```
                                    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

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

           Units Sold  Unit Price  Unit Cost  Total Revenue  Total Cost  \
Order ID
669165933        9925      255.28     159.42        4870.26  1582243.50
```

```
963881480        2804      205.70     117.11      435466.90   328376.44
341417157        1779      651.21     524.96      247956.32   933903.84
514321792        8102        9.33       6.92       75591.66    56065.84
115456712        5062      651.21     524.96      471336.91  2657347.52

           Total Profit  Order Year  Order Quarter  Order Month
Order ID
669165933     951410.50        2010              2            5
963881480     248406.36        2012              3            8
341417157     224598.75        2014              2            5
514321792      19525.82        2014              2            6
115456712     639077.50        2013              1            2
```

[113]:
```
sales.loc[:,'Order Priority'].unique()
```

[113]:
```
array(['H', 'C', 'L', 'M'], dtype=object)
```

[114]:
```
sales['Order Date'].unique()
```

[114]:
```
array(['5/28/2010', '8/22/2012', '5/2/2014', '6/20/2014', '2/1/2013',
       '2/4/2015', '4/23/2011', '7/17/2012', '7/14/2015', '4/18/2014',
       '6/24/2011', '8/2/2014', '1/13/2017', '2/8/2017', '2/19/2014',
       '4/23/2012', '11/19/2016', '4/1/2015', '12/30/2010', '7/31/2012',
       '5/14/2014', '7/31/2015', '6/30/2016', '9/8/2014', '5/7/2016',
       '5/22/2017', '10/13/2014', '5/7/2010', '7/18/2014', '5/26/2012',
       '9/17/2012', '12/29/2013', '10/27/2015', '1/16/2015', '2/25/2017',
       '5/8/2017', '11/22/2011', '1/14/2017', '4/1/2012', '2/16/2012',
       '3/11/2017', '2/6/2010', '6/7/2012', '10/6/2012', '11/14/2015',
       '3/29/2016', '12/31/2016', '12/23/2010', '10/14/2014', '1/11/2012',
       '2/2/2010', '8/18/2013', '3/25/2013', '11/26/2011', '9/17/2013',
       '6/8/2012', '6/30/2010', '2/23/2015', '1/5/2012', '4/7/2014',
       '6/9/2013', '6/26/2013', '11/7/2011', '10/30/2010', '10/13/2013',
       '10/11/2013', '7/8/2012', '7/25/2016', '10/24/2010', '4/25/2015',
       '4/23/2013', '8/14/2015', '5/26/2011', '5/20/2017', '7/5/2013',
       '11/6/2014', '10/28/2014', '9/15/2011', '5/29/2012', '7/20/2013',
       '10/21/2012', '9/18/2012', '11/15/2016', '1/4/2011', '3/18/2012',
       '2/17/2012', '1/16/2011', '2/3/2014', '4/30/2012', '10/23/2016',
       '12/6/2016', '7/7/2014', '6/13/2012', '11/26/2010', '2/8/2011',
       '7/26/2011', '11/11/2011', '6/1/2016', '7/30/2015', '2/10/2012'],
      dtype=object)
```

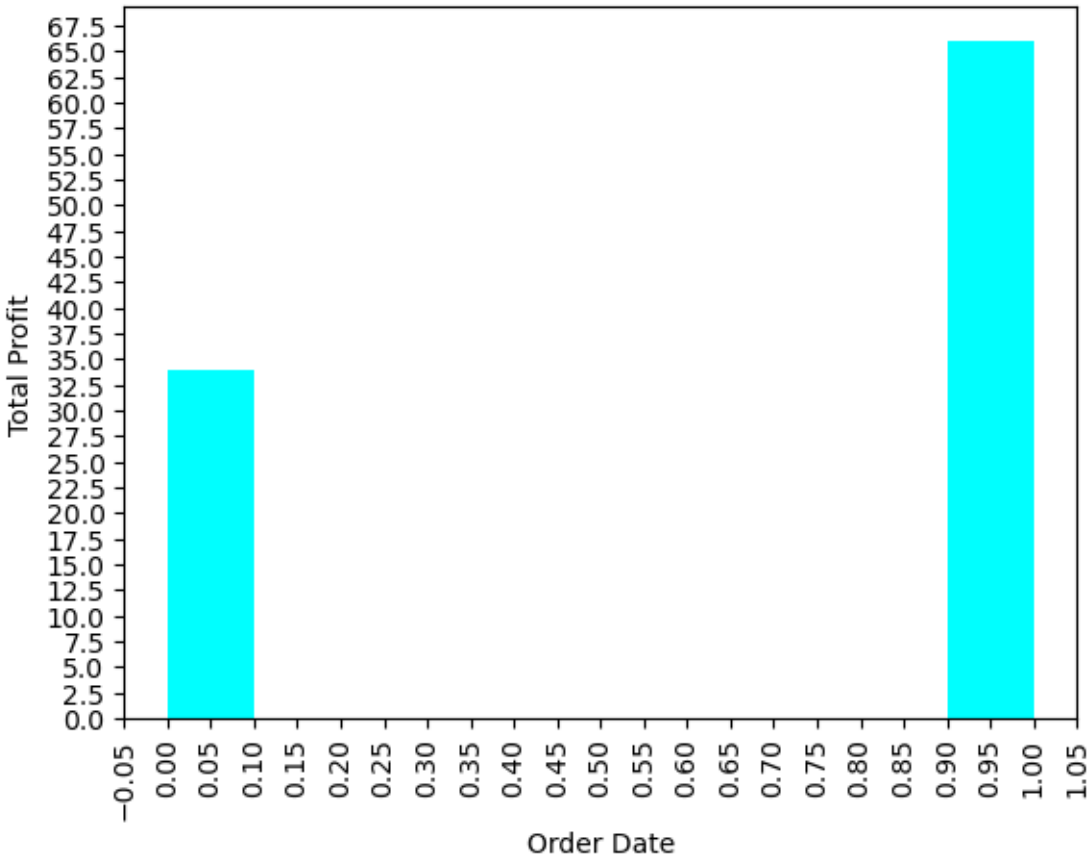
[115]:
```
order_dates = ['5/28/2010', '8/22/2012', '05-02-2014', '6/20/2014',␣
↪'02-01-2013',
       '02-04-2015', '4/23/2011', '7/17/2012', '7/14/2015', '4/18/2014',
       '6/24/2011', '08-02-2014', '1/13/2017', '02-08-2017', '2/19/2014',
       '4/23/2012', '11/19/2016', '04-01-2015', '12/30/2010', '7/31/2012',
       '5/14/2014', '7/31/2015', '6/30/2016', '09-08-2014', '05-07-2016',
```

```
       '5/22/2017', '10/13/2014', '05-07-2010', '7/18/2014', '5/26/2012',
       '9/17/2012', '12/29/2013', '10/27/2015', '1/16/2015', '2/25/2017',
       '05-08-2017', '11/22/2011', '1/14/2017', '04-01-2012', '2/16/2012',
       '03-11-2017', '02-06-2010', '06-07-2012', '10-06-2012',
       '11/14/2015', '3/29/2016', '12/31/2016', '12/23/2010',
       '10/14/2014', '01-11-2012', '02-02-2010', '8/18/2013', '3/25/2013',
       '11/26/2011', '9/17/2013', '06-08-2012', '6/30/2010', '2/23/2015',
       '01-05-2012', '04-07-2014', '06-09-2013', '6/26/2013',
       '11-07-2011', '10/30/2010', '10/13/2013', '10-11-2013',
       '07-08-2012', '7/25/2016', '10/24/2010', '4/25/2015', '4/23/2013',
       '8/14/2015', '5/26/2011', '5/20/2017', '07-05-2013', '11-06-2014',
       '10/28/2014', '9/15/2011', '5/29/2012', '7/20/2013', '10/21/2012',
       '9/18/2012', '11/15/2016', '01-04-2011', '3/18/2012', '2/17/2012',
       '1/16/2011', '02-03-2014', '4/30/2012', '10/23/2016', '12-06-2016',
       '07-07-2014', '6/13/2012', '11/26/2010', '02-08-2011', '7/26/2011',
       '11-11-2011', '06-01-2016', '7/30/2015', '02-10-2012']
```

[117]:
```
sales['Order Date'] = pd.Categorical(sales['Order␣
↪Date'],categories=order_dates,ordered=True)
```

[118]:
```
pd.pivot_table(sales,values='Total Profit',index='Order Date',aggfunc='count').
↪plot(kind='hist',color='aqua',stacked=False,legend=None)
plt.xticks(rotation=90)
plt.ylabel('Total Profit')
plt.locator_params(nbins=32)
plt.xlabel('Order Date')
```

[118]:
```
Text(0.5, 0, 'Order Date')
```



[119]:
```
sales.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 100 entries, 669165933 to 665095412
Data columns (total 16 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   Region          100 non-null    object
 1   Country         100 non-null    object
 2   Item Type       100 non-null    category
 3   Sales Channel   100 non-null    object
 4   Order Priority  100 non-null    object
 5   Order Date      66 non-null     category
 6   Ship Date       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  Order Year      100 non-null    int32
 14  Order Quarter   100 non-null    int32
 15  Order Month     100 non-null    int32
dtypes: category(2), float64(5), int32(3), int64(1), object(5)
memory usage: 16.0+ KB
```

[120]:
```
sales.describe()
```

[120]:

	Units Sold	Unit Price	Unit Cost	Total Revenue	Total Cost
count	100.000000	100.000000	100.000000	1.000000e+02	1.000000e+02
mean	5128.710000	276.761300	191.048000	1.373488e+06	9.318057e+05
std	2794.484562	235.592241	188.208181	1.460029e+06	1.083938e+06
min	124.000000	9.330000	6.920000	4.870260e+03	3.612240e+03
25%	2836.250000	81.730000	35.840000	2.687212e+05	1.688680e+05
50%	5382.500000	179.880000	107.275000	7.523144e+05	3.635664e+05
75%	7369.000000	437.200000	263.330000	2.212045e+06	1.613870e+06
max	9925.000000	668.270000	524.960000	5.997055e+06	4.509794e+06

	Total Profit	Order Year	Order Quarter	Order Month
count	1.000000e+02	100.000000	100.000000	100.000000
mean	4.416820e+05	2013.230000	2.470000	6.260000
std	4.385379e+05	2.088231	1.114233	3.353334
min	1.258020e+03	2010.000000	1.000000	1.000000
25%	1.214436e+05	2012.000000	2.000000	4.000000
50%	2.907680e+05	2013.000000	2.000000	6.000000
75%	6.358288e+05	2015.000000	3.250000	9.250000
max	1.719922e+06	2017.000000	4.000000	12.000000

[121]:
```
sales
```

[121]:

Order ID	Region	Country	Item Type	Sales Channel	Order Priority	Order Date
669165933	Australia and Oceania	Tuvalu	Baby Food	Offline	H	5/28/2010
963881480	Central America and the Caribbean	Grenada				
341417157	Europe	Russia				
514321792	Sub-Saharan Africa	Sao Tome and Principe				
115456712	Sub-Saharan Africa	Rwanda				
…	…	…				
512878119	Sub-Saharan Africa	Mali				
810711038	Asia	Malaysia				
728815257	Sub-Saharan Africa	Sierra Leone				
559427106	North America	Mexico				
665095412	Sub-Saharan Africa	Mozambique				

```
963881480          Cereal        Online              C  8/22/2012
341417157  Office Supplies       Offline              L        NaN
514321792          Fruits        Online              C  6/20/2014
115456712  Office Supplies       Offline              L        NaN
…                      …           …            …         …
512878119         Clothes        Online              M  7/26/2011
810711038          Fruits       Offline              L        NaN
728815257      Vegetables       Offline              C        NaN
559427106   Personal Care       Offline              M  7/30/2015
665095412       Household       Offline              L        NaN

            Ship Date  Units Sold  Unit Price  Unit Cost  Total Revenue  \
Order ID
669165933   6/27/2010        9925      255.28     159.42        4870.26
963881480   9/15/2012        2804      205.70     117.11      435466.90
341417157    5/8/2014        1779      651.21     524.96      247956.32
514321792    7/5/2014        8102        9.33       6.92       75591.66
115456712    2/6/2013        5062      651.21     524.96      471336.91
…                 …           …           …          …            …
512878119    9/3/2011         888      109.28      35.84     5513227.50
810711038  12/28/2011        6267        9.33       6.92     4368316.68
728815257   6/29/2016        1485      154.06      90.93     3876652.40
559427106    8/8/2015        5767       81.73      56.67     3296425.02
665095412   2/15/2012        5367      668.27     502.54     5997054.98

            Total Cost  Total Profit  Order Year  Order Quarter  Order Month
Order ID
669165933  1582243.50     951410.50        2010              2            5
963881480   328376.44     248406.36        2012              3            8
341417157   933903.84     224598.75        2014              2            5
514321792    56065.84      19525.82        2014              2            6
115456712  2657347.52     639077.50        2013              1            2
…                …           …           …              …            …
512878119    31825.92      65214.72        2011              3            7
810711038    43367.64      15103.47        2011              4           11
728815257   135031.05      93748.05        2016              2            6
559427106   326815.89     144521.02        2015              3            7
665095412  2697132.18     889472.91        2012              1            2

[100 rows x 16 columns]
```

[122]:
```
sales['Region'].unique()
```

[122]:
```
array(['Australia and Oceania', 'Central America and the Caribbean',
       'Europe', 'Sub-Saharan Africa', 'Asia',
       'Middle East and North Africa', 'North America'], dtype=object)
```

[123]:
```
regions = ['Australia and Oceania', 'Central America and the Caribbean',
       'Europe', 'Sub-Saharan Africa', 'Asia',
       'Middle East and North Africa', 'North America']
```

[124]:
```
sales['Region'] = pd.Categorical(sales['Region'],categories =␣
↪regions,ordered=True)
```

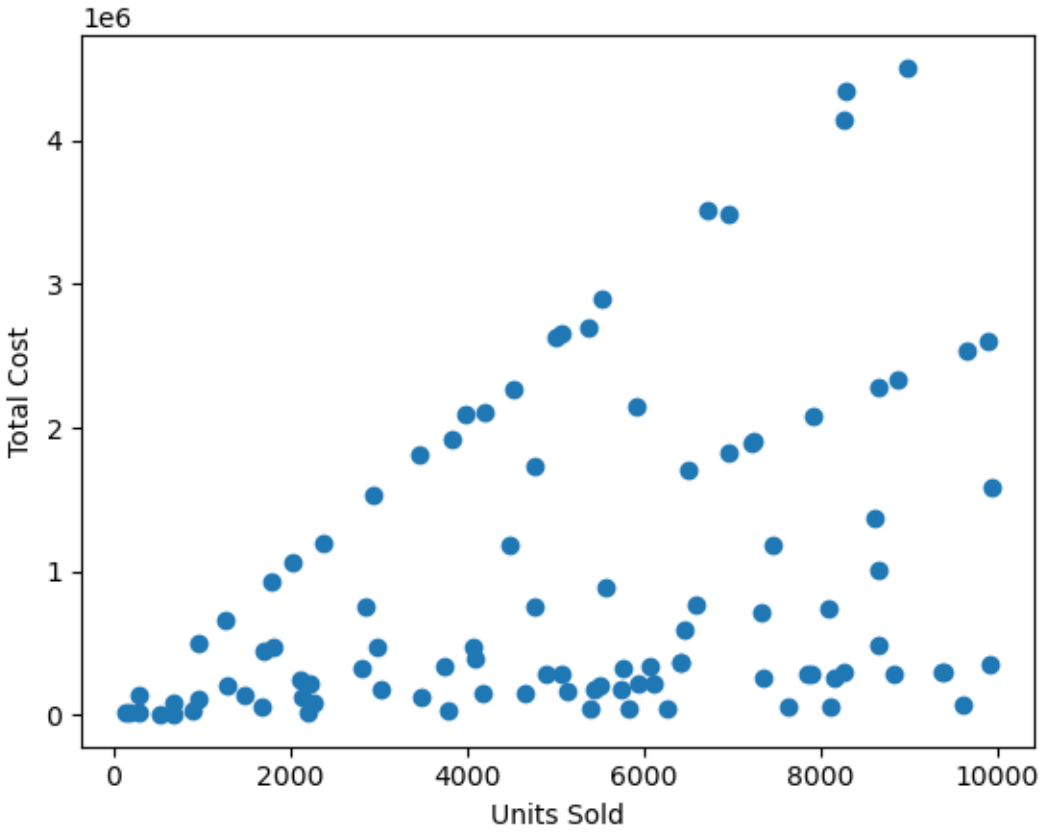
[125]:
```
sales.head()
```

[125]:

Order ID	Region	Country	Item Type	Sales Channel	Order Priority	Order Date	Ship Date	Units Sold	Unit Price	Unit Cost	Total Revenue	Total Cost	Total Profit	Order Year	Order Quarter	Order Month
669165933	Australia and Oceania	Tuvalu	Baby Food	Offline	H	5/28/2010	6/27/2010	9925	255.28	159.42	4870.26	1582243.50	951410.50	2010	2	5
963881480	Central America and the Caribbean	Grenada	Cereal	Online	C	8/22/2012	9/15/2012	2804	205.70	117.11	435466.90	328376.44	248406.36	2012	3	8
341417157	Europe	Russia	Office Supplies	Offline	L	NaN	5/8/2014	1779	651.21	524.96	247956.32	933903.84	224598.75	2014	2	5
514321792	Sub-Saharan Africa	Sao Tome and Principe	Fruits	Online	C	6/20/2014	7/5/2014	8102	9.33	6.92	75591.66	56065.84	19525.82	2014	2	6
115456712	Sub-Saharan Africa	Rwanda	Office Supplies	Offline	L	NaN	2/6/2013	5062	651.21	524.96	471336.91	2657347.52	639077.50	2013	1	2

[127]:
```
plt.scatter(sales['Units Sold'],sales['Total Cost'])
plt.xlabel('Units Sold')
plt.ylabel('Total Cost')
```

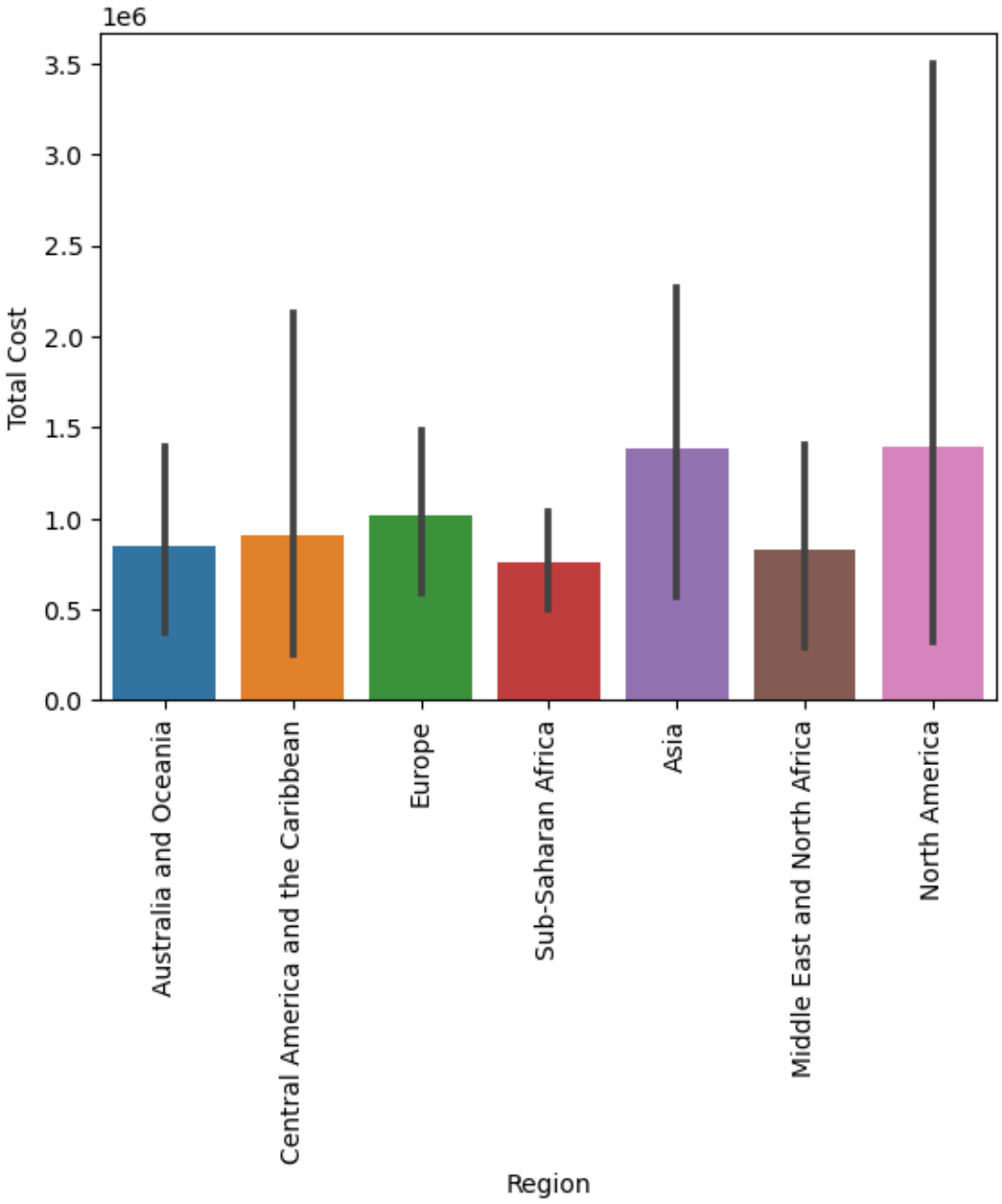
[127]:
```
Text(0, 0.5, 'Total Cost')
```



0.4.10 More the number of units sold of a product, more will be the total cost associated with it.

[130]:
```
sns.barplot(x='Region',y='Total Cost',data=sales)
plt.xticks(rotation=90)
```

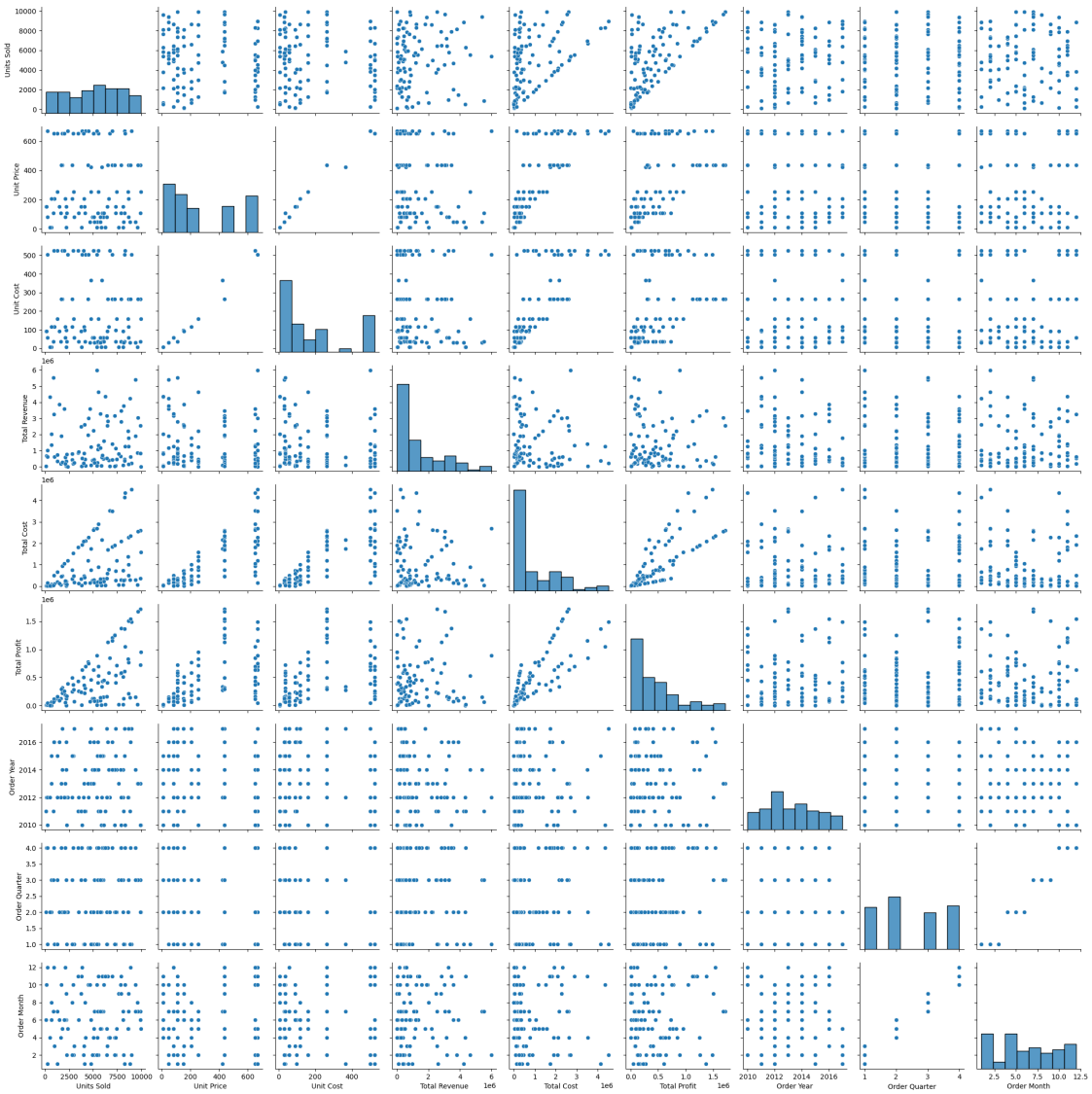
[130]:
```
(array([0, 1, 2, 3, 4, 5, 6]),
 [Text(0, 0, 'Australia and Oceania'),
  Text(1, 0, 'Central America and the Caribbean'),
  Text(2, 0, 'Europe'),
  Text(3, 0, 'Sub-Saharan Africa'),
  Text(4, 0, 'Asia'),
  Text(5, 0, 'Middle East and North Africa'),
  Text(6, 0, 'North America')])
```



0.4.11 Cost of items is maximum in Asia and North America, and minimum in Sub-Saharan Africa.

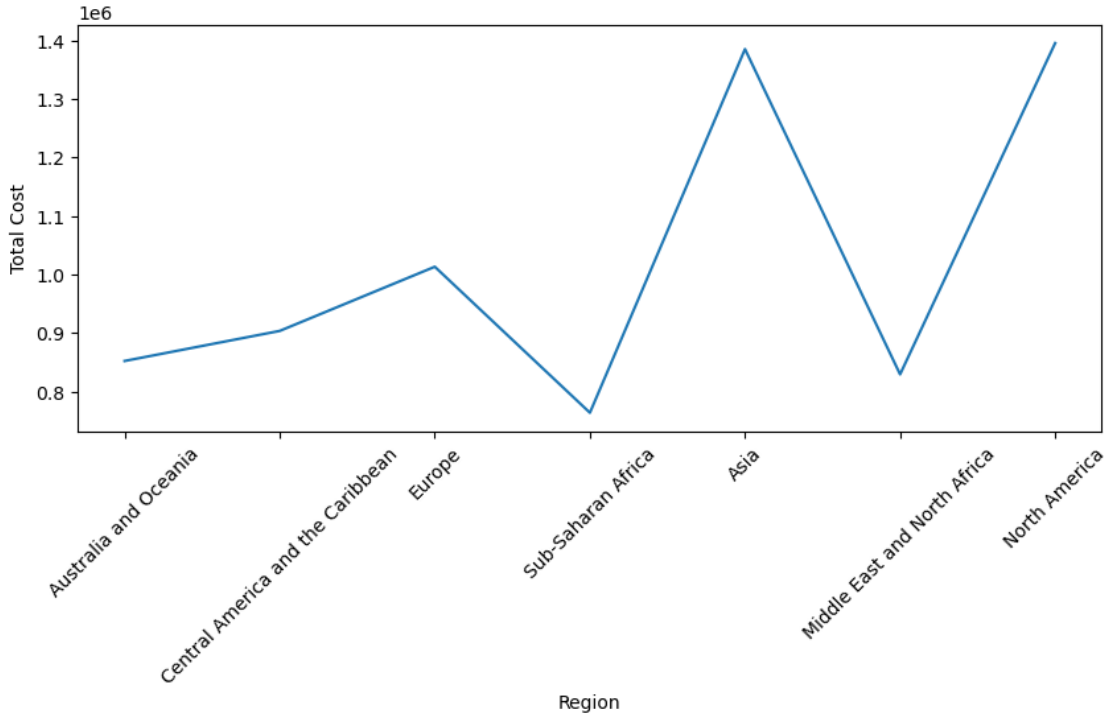
[131]:
```
sns.pairplot(sales)
```

[131]: <seaborn.axisgrid.PairGrid at 0x264b0b8e9d0>



[137]:
```
plt.figure(figsize=(10,4))
sns.lineplot(x='Region',y='Total Cost',data=sales,ci=None)
plt.xticks(rotation=45)
```

[137]:
```
([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.

[139]:
```
sales.head()
```

[139]:

Order ID	Region	Country	Item Type	Sales Channel	Order Priority	Order Date	Ship Date	Units Sold	Unit Price	Unit Cost	Total Revenue	Total Cost
669165933	Australia and Oceania	Tuvalu	Baby Food	Offline	H	5/28/2010	6/27/2010	9925	255.28	159.42	4870.26	1582243.50
963881480	Central America and the Caribbean	Grenada	Cereal	Online	C	8/22/2012	9/15/2012					
341417157	Europe	Russia	Office Supplies	Offline	L	NaN	5/8/2014					
514321792	Sub-Saharan Africa	Sao Tome and Principe	Fruits	Online	C	6/20/2014	7/5/2014					
115456712	Sub-Saharan Africa	Rwanda	Office Supplies	Offline	L	NaN	2/6/2013					

```
963881480        2804      205.70     117.11      435466.90   328376.44
341417157        1779      651.21     524.96      247956.32   933903.84
514321792        8102        9.33       6.92       75591.66    56065.84
115456712        5062      651.21     524.96      471336.91  2657347.52
```

Order ID	Total Profit	Order Year	Order Quarter	Order Month
669165933	951410.50	2010	2	5
963881480	248406.36	2012	3	8
341417157	224598.75	2014	2	5
514321792	19525.82	2014	2	6
115456712	639077.50	2013	1	2

[140]:
```
sales['Country'].unique()
```

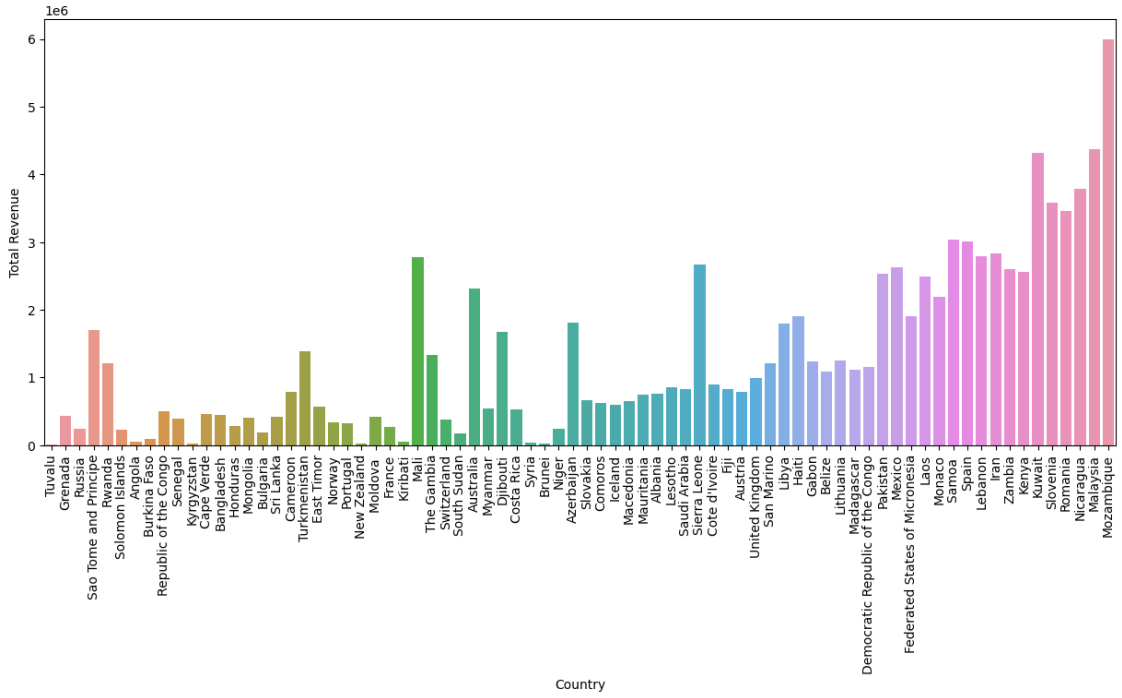
[140]:
```
array(['Tuvalu', 'Grenada', 'Russia', 'Sao Tome and Principe', 'Rwanda',
       'Solomon Islands', 'Angola', 'Burkina Faso',
       'Republic of the Congo', 'Senegal', 'Kyrgyzstan', 'Cape Verde',
       'Bangladesh', 'Honduras', 'Mongolia', 'Bulgaria', 'Sri Lanka',
       'Cameroon', 'Turkmenistan', '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)
```

[141]:
```
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']
```

[143]:
```
sales['Country'] = pd.
↪Categorical(sales['Country'],categories=countries,ordered=True)
```

[145]:
```
plt.figure(figsize=(15,6))
sns.barplot(x='Country', y='Total Revenue', data=sales, ci=None)
plt.xticks(rotation=90)
plt.tick_params(axis='x', which='major', labelsize=10)
```



Mozambique is the country where maximum revenue has been generated followed by Kenya.

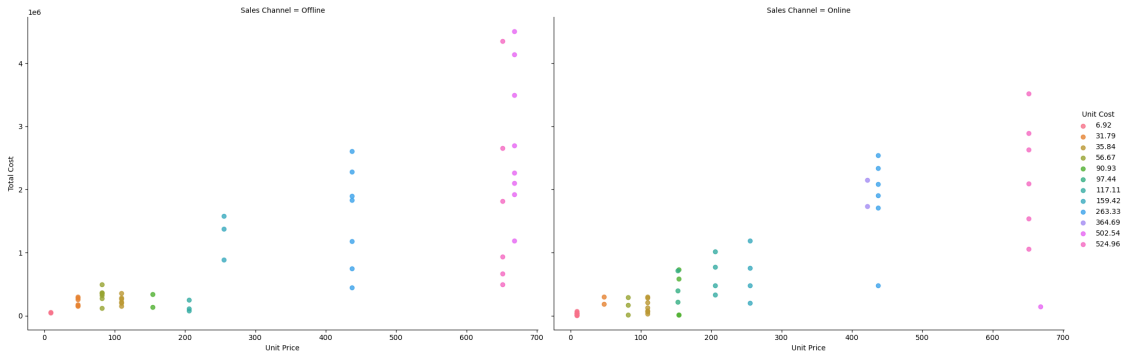
[146]:
```
sales.head()
```

[146]:
```
                                    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

                Item Type Sales Channel Order Priority Order Date  Ship Date  \
```

```
Order ID
669165933        Baby Food       Offline              H  5/28/2010  6/27/2010
963881480           Cereal        Online              C  8/22/2012  9/15/2012
341417157  Office Supplies       Offline              L        NaN   5/8/2014
514321792           Fruits        Online              C  6/20/2014   7/5/2014
115456712  Office Supplies       Offline              L        NaN   2/6/2013

           Units Sold  Unit Price  Unit Cost  Total Revenue  Total Cost  \
Order ID
669165933        9925      255.28     159.42        4870.26  1582243.50
963881480        2804      205.70     117.11      435466.90   328376.44
341417157        1779      651.21     524.96      247956.32   933903.84
514321792        8102        9.33       6.92       75591.66    56065.84
115456712        5062      651.21     524.96      471336.91  2657347.52

           Total Profit  Order Year  Order Quarter  Order Month
Order ID
669165933     951410.50        2010              2            5
963881480     248406.36        2012              3            8
341417157     224598.75        2014              2            5
514321792      19525.82        2014              2            6
115456712     639077.50        2013              1            2
```

[148]:
```
sns.lmplot(x='Unit Price',y='Total Cost',data=sales,col='Sales␣
↪Channel',hue='Unit Cost',aspect=1.5,height=7)
```

[148]:
```
<seaborn.axisgrid.FacetGrid at 0x264ba011b50>
```



[ ]: