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

data=pd.read_csv("Downloads\\Amazon Sales data.csv")

data
```

```
                                Region
Country  \
0               Australia and Oceania
Tuvalu
1   Central America and the Caribbean
Grenada
2                               Europe
Russia
3                  Sub-Saharan Africa             Sao Tome and
Principe
4                  Sub-Saharan Africa
Rwanda
5               Australia and Oceania                    Solomon
Islands
6                  Sub-Saharan Africa
Angola
7                  Sub-Saharan Africa                       Burkina
Faso
8                  Sub-Saharan Africa             Republic of the
Congo
9                  Sub-Saharan Africa
Senegal
10                               Asia
Kyrgyzstan
11                 Sub-Saharan Africa                         Cape
Verde
12                               Asia
Bangladesh
13  Central America and the Caribbean
Honduras
14                               Asia
Mongolia
15                              Europe
Bulgaria
16                               Asia                           Sri
Lanka
17                 Sub-Saharan Africa
Cameroon
18                               Asia
Turkmenistan
```

```
19               Australia and Oceania                           East
Timor
20                                Europe
Norway
21                                Europe
Portugal
22  Central America and the Caribbean
Honduras
23               Australia and Oceania                          New
Zealand
24                                Europe                            Moldova

25                                Europe
France
26               Australia and Oceania
Kiribati
27                   Sub-Saharan Africa
Mali
28                                Europe
Norway
29                   Sub-Saharan Africa                           The
Gambia
30                                Europe
Switzerland
31                   Sub-Saharan Africa                          South
Sudan
32               Australia and Oceania
Australia
33                                  Asia
Myanmar
34                   Sub-Saharan Africa
Djibouti
35  Central America and the Caribbean                            Costa
Rica
36       Middle East and North Africa
Syria
37                   Sub-Saharan Africa                           The
Gambia
38                                  Asia
Brunei
39                                Europe
Bulgaria
40                   Sub-Saharan Africa
Niger
41       Middle East and North Africa
Azerbaijan
42                   Sub-Saharan Africa                           The
Gambia
43                                Europe
```

```
Slovakia
44                                Asia
Myanmar
45                  Sub-Saharan Africa
Comoros
46                              Europe
Iceland
47                              Europe
Switzerland
48                              Europe
Macedonia
49                  Sub-Saharan Africa
Mauritania
50                              Europe
Albania
51                  Sub-Saharan Africa
Lesotho
52       Middle East and North Africa                       Saudi
Arabia
53                  Sub-Saharan Africa                      Sierra
Leone
54                  Sub-Saharan Africa             Sao Tome and
Principe
55                  Sub-Saharan Africa                     Cote
d'Ivoire
56               Australia and Oceania
Fiji
57                              Europe
Austria
58                              Europe                    United
Kingdom
59                  Sub-Saharan Africa
Djibouti
60               Australia and Oceania
Australia
61                              Europe                         San
Marino
62                  Sub-Saharan Africa
Cameroon
63       Middle East and North Africa
Libya
64  Central America and the Caribbean
Haiti
65                  Sub-Saharan Africa
Rwanda
66                  Sub-Saharan Africa
Gabon
67  Central America and the Caribbean
Belize
```

```
68                                Europe
Lithuania
69                  Sub-Saharan Africa
Madagascar
70                                  Asia
Turkmenistan
71       Middle East and North Africa
Libya
72                  Sub-Saharan Africa  Democratic Republic of the
Congo
73                  Sub-Saharan Africa
Djibouti
74       Middle East and North Africa
Pakistan
75                        North America
Mexico
76               Australia and Oceania    Federated States of
Micronesia
77                                  Asia
Laos
78                                Europe
Monaco
79               Australia and Oceania                              Samoa

80                                Europe
Spain
81       Middle East and North Africa
Lebanon
82       Middle East and North Africa
Iran
83                  Sub-Saharan Africa
Zambia
84                  Sub-Saharan Africa
Kenya
85                        North America
Mexico
86                  Sub-Saharan Africa              Sao Tome and
Principe
87                  Sub-Saharan Africa                          The
Gambia
88       Middle East and North Africa
Kuwait
89                                Europe
Slovenia
90                  Sub-Saharan Africa                       Sierra
Leone
91               Australia and Oceania
Australia
92       Middle East and North Africa
```

```
Azerbaijan
93                              Europe
Romania
94  Central America and the Caribbean
Nicaragua
95                  Sub-Saharan Africa
Mali
96                                Asia
Malaysia
97                  Sub-Saharan Africa                        Sierra
Leone
98                        North America
Mexico
99                  Sub-Saharan Africa
Mozambique

          Item Type Sales Channel Order Priority  Order Date   Order
ID  \
0         Baby Food       Offline              H   5/28/2010
669165933
1            Cereal        Online              C   8/22/2012
963881480
2   Office Supplies       Offline              L    5/2/2014
341417157
3            Fruits        Online              C   6/20/2014
514321792
4   Office Supplies       Offline              L    2/1/2013
115456712
5         Baby Food        Online              C    2/4/2015
547995746
6         Household       Offline              M   4/23/2011
135425221
7        Vegetables        Online              H   7/17/2012
871543967
8     Personal Care       Offline              M   7/14/2015
770463311
9            Cereal        Online              H   4/18/2014
616607081
10       Vegetables        Online              H   6/24/2011
814711606
11          Clothes       Offline              H    8/2/2014
939825713
12          Clothes        Online              L   1/13/2017
187310731
13        Household       Offline              H    2/8/2017
522840487
14    Personal Care       Offline              C   2/19/2014
832401311
15          Clothes        Online              M   4/23/2012
```

```
972292029
16        Cosmetics       Offline              M  11/19/2016
419123971
17        Beverages       Offline              C    4/1/2015
519820964
18        Household       Offline              L  12/30/2010
441619336
19             Meat        Online              L   7/31/2012
322067916
20        Baby Food        Online              L   5/14/2014
819028031
21        Baby Food        Online              H   7/31/2015
860673511
22           Snacks        Online              L   6/30/2016
795490682
23           Fruits        Online              H    9/8/2014
142278373
24    Personal Care        Online              L    5/7/2016
740147912
25        Cosmetics        Online              H   5/22/2017
898523128
26           Fruits        Online              M  10/13/2014
347140347
27           Fruits        Online              L    5/7/2010
686048400
28        Beverages       Offline              C   7/18/2014
435608613
29        Household       Offline              L   5/26/2012
886494815
30        Cosmetics       Offline              M   9/17/2012
249693334
31    Personal Care       Offline              C  12/29/2013
406502997
32  Office Supplies        Online              C  10/27/2015
158535134
33        Household       Offline              H   1/16/2015
177713572
34           Snacks        Online              M   2/25/2017
756274640
35    Personal Care       Offline              L    5/8/2017
456767165
36           Fruits        Online              L  11/22/2011
162052476
37             Meat        Online              M   1/14/2017
825304400
38  Office Supplies        Online              L    4/1/2012
320009267
39  Office Supplies        Online              M   2/16/2012
189965903
```

```
40    Personal Care        Online              H   3/11/2017
699285638
41        Cosmetics        Online              M    2/6/2010
382392299
42           Cereal       Offline              H    6/7/2012
994022214
43       Vegetables        Online              H   10/6/2012
759224212
44          Clothes        Online              H  11/14/2015
223359620
45           Cereal       Offline              H   3/29/2016
902102267
46        Cosmetics        Online              C  12/31/2016
331438481
47    Personal Care        Online              M  12/23/2010
617667090
48          Clothes       Offline              C  10/14/2014
787399423
49  Office Supplies       Offline              C   1/11/2012
837559306
50          Clothes        Online              C    2/2/2010
385383069
51           Fruits        Online              L   8/18/2013
918419539
52           Cereal        Online              M   3/25/2013
844530045
53  Office Supplies       Offline              M  11/26/2011
441888415
54           Fruits       Offline              H   9/17/2013
508980977
55          Clothes        Online              C    6/8/2012
114606559
56          Clothes       Offline              C   6/30/2010
647876489
57        Cosmetics       Offline              H   2/23/2015
868214595
58        Household        Online              L    1/5/2012
955357205
59        Cosmetics       Offline              H    4/7/2014
259353148
60           Cereal       Offline              H    6/9/2013
450563752
61        Baby Food        Online              L   6/26/2013
569662845
62  Office Supplies        Online              M   11/7/2011
177636754
63          Clothes       Offline              H  10/30/2010
705784308
64        Cosmetics       Offline              H  10/13/2013
```

```
505716836
65        Cosmetics       Offline              H  10/11/2013
699358165
66    Personal Care       Offline              L    7/8/2012
228944623
67          Clothes       Offline              M   7/25/2016
807025039
68  Office Supplies       Offline              H  10/24/2010
166460740
69          Clothes       Offline              L   4/25/2015
610425555
70  Office Supplies        Online              M   4/23/2013
462405812
71           Fruits        Online              L   8/14/2015
816200339
72        Beverages        Online              C   5/26/2011
585920464
73           Cereal        Online              H   5/20/2017
555990016
74        Cosmetics       Offline              L    7/5/2013
231145322
75        Household       Offline              C   11/6/2014
986435210
76        Beverages        Online              C  10/28/2014
217221009
77       Vegetables       Offline              C   9/15/2011
789176547
78        Baby Food       Offline              H   5/29/2012
688288152
79        Cosmetics        Online              H   7/20/2013
670854651
80        Household       Offline              L  10/21/2012
213487374
81          Clothes        Online              L   9/18/2012
663110148
82        Cosmetics        Online              H  11/15/2016
286959302
83           Snacks        Online              L    1/4/2011
122583663
84       Vegetables        Online              L   3/18/2012
827844560
85    Personal Care       Offline              L   2/17/2012
430915820
86        Beverages       Offline              C   1/16/2011
180283772
87        Baby Food       Offline              M    2/3/2014
494747245
88           Fruits        Online              M   4/30/2012
513417565
```

```
89        Beverages       Offline              C  10/23/2016
345718562
90  Office Supplies       Offline              H   12/6/2016
621386563
91        Beverages       Offline              H    7/7/2014
240470397
92  Office Supplies        Online              M   6/13/2012
423331391
93        Cosmetics        Online              H  11/26/2010
660643374
94        Beverages       Offline              C    2/8/2011
963392674
95          Clothes        Online              M   7/26/2011
512878119
96           Fruits       Offline              L  11/11/2011
810711038
97       Vegetables       Offline              C    6/1/2016
728815257
98    Personal Care       Offline              M   7/30/2015
559427106
99        Household       Offline              L   2/10/2012
665095412

     Ship Date  Units Sold  Unit Price  Unit Cost  Total Revenue
Total Cost  \
0    6/27/2010        9925      255.28     159.42     2533654.00
1582243.50
1    9/15/2012        2804      205.70     117.11      576782.80
328376.44
2     5/8/2014        1779      651.21     524.96     1158502.59
933903.84
3     7/5/2014        8102        9.33       6.92       75591.66
56065.84
4     2/6/2013        5062      651.21     524.96     3296425.02
2657347.52
5    2/21/2015        2974      255.28     159.42      759202.72
474115.08
6    4/27/2011        4187      668.27     502.54     2798046.49
2104134.98
7    7/27/2012        8082      154.06      90.93     1245112.92
734896.26
8    8/25/2015        6070       81.73      56.67      496101.10
343986.90
9    5/30/2014        6593      205.70     117.11     1356180.10
772106.23
10   7/12/2011         124      154.06      90.93       19103.44
11275.32
11   8/19/2014        4168      109.28      35.84      455479.04
149381.12
```

12	3/1/2017	8263	109.28	35.84	902980.64	296145.92
13	2/13/2017	8974	668.27	502.54	5997054.98	4509793.96
14	2/23/2014	4901	81.73	56.67	400558.73	277739.67
15	6/3/2012	1673	109.28	35.84	182825.44	59960.32
16	12/18/2016	6952	437.20	263.33	3039414.40	1830670.16
17	4/18/2015	5430	47.45	31.79	257653.50	172619.70
18	1/20/2011	3830	668.27	502.54	2559474.10	1924728.20
19	9/11/2012	5908	421.89	364.69	2492526.12	2154588.52
20	6/28/2014	7450	255.28	159.42	1901836.00	1187679.00
21	9/3/2015	1273	255.28	159.42	324971.44	202941.66
22	7/26/2016	2225	152.58	97.44	339490.50	216804.00
23	10/4/2014	2187	9.33	6.92	20404.71	15134.04
24	5/10/2016	5070	81.73	56.67	414371.10	287316.90
25	6/5/2017	1815	437.20	263.33	793518.00	477943.95
26	11/10/2014	5398	9.33	6.92	50363.34	37354.16
27	5/10/2010	5822	9.33	6.92	54319.26	40288.24
28	7/30/2014	5124	47.45	31.79	243133.80	162891.96
29	6/9/2012	2370	668.27	502.54	1583799.90	1191019.80
30	10/20/2012	8661	437.20	263.33	3786589.20	2280701.13
31	1/28/2014	2125	81.73	56.67	173676.25	120423.75
32	11/25/2015	2924	651.21	524.96	1904138.04	1534983.04
33	3/1/2015	8250	668.27	502.54	5513227.50	4145955.00
34	2/25/2017	7327	152.58	97.44	1117953.66	713942.88
35	5/21/2017	6409	81.73	56.67	523807.57	363198.03
36	12/3/2011	3784	9.33	6.92	35304.72	

26185.28
37 1/23/2017 4767 421.89 364.69 2011149.63
1738477.23
38 5/8/2012 6708 651.21 524.96 4368316.68
3521431.68
39 2/28/2012 3987 651.21 524.96 2596374.27
2093015.52
40 3/28/2017 3015 81.73 56.67 246415.95
170860.05
41 2/25/2010 7234 437.20 263.33 3162704.80
1904929.22
42 6/8/2012 2117 205.70 117.11 435466.90
247921.87
43 11/10/2012 171 154.06 90.93 26344.26
15549.03
44 11/18/2015 5930 109.28 35.84 648030.40
212531.20
45 4/29/2016 962 205.70 117.11 197883.40
112659.82
46 12/31/2016 8867 437.20 263.33 3876652.40
2334947.11
47 1/31/2011 273 81.73 56.67 22312.29
15470.91
48 11/14/2014 7842 109.28 35.84 856973.76
281057.28
49 1/13/2012 1266 651.21 524.96 824431.86
664599.36
50 3/18/2010 2269 109.28 35.84 247956.32
81320.96
51 9/18/2013 9606 9.33 6.92 89623.98
66473.52
52 3/28/2013 4063 205.70 117.11 835759.10
475817.93
53 1/7/2012 3457 651.21 524.96 2251232.97
1814786.72
54 10/24/2013 7637 9.33 6.92 71253.21
52848.04
55 6/27/2012 3482 109.28 35.84 380512.96
124794.88
56 8/1/2010 9905 109.28 35.84 1082418.40
354995.20
57 3/2/2015 2847 437.20 263.33 1244708.40
749700.51
58 2/14/2012 282 668.27 502.54 188452.14
141716.28
59 4/19/2014 7215 437.20 263.33 3154398.00
1899925.95
60 7/2/2013 682 205.70 117.11 140287.40
79869.02
61 7/1/2013 4750 255.28 159.42 1212580.00

757245.00
62 11/15/2011 5518 651.21 524.96 3593376.78
2896729.28
63 11/17/2010 6116 109.28 35.84 668356.48
219197.44
64 11/16/2013 1705 437.20 263.33 745426.00
448977.65
65 11/25/2013 4477 437.20 263.33 1957344.40
1178928.41
66 7/9/2012 8656 81.73 56.67 707454.88
490535.52
67 9/7/2016 5498 109.28 35.84 600821.44
197048.32
68 11/17/2010 8287 651.21 524.96 5396577.27
4350343.52
69 5/28/2015 7342 109.28 35.84 802333.76
263137.28
70 5/20/2013 5010 651.21 524.96 3262562.10
2630049.60
71 9/30/2015 673 9.33 6.92 6279.09
4657.16
72 7/15/2011 5741 47.45 31.79 272410.45
182506.39
73 6/17/2017 8656 205.70 117.11 1780539.20
1013704.16
74 8/16/2013 9892 437.20 263.33 4324782.40
2604860.36
75 12/12/2014 6954 668.27 502.54 4647149.58
3494663.16
76 11/15/2014 9379 47.45 31.79 445033.55
298158.41
77 10/23/2011 3732 154.06 90.93 574951.92
339350.76
78 6/2/2012 8614 255.28 159.42 2198981.92
1373243.88
79 8/7/2013 9654 437.20 263.33 4220728.80
2542187.82
80 11/30/2012 4513 668.27 502.54 3015902.51
2267963.02
81 10/8/2012 7884 109.28 35.84 861563.52
282562.56
82 12/8/2016 6489 437.20 263.33 2836990.80
1708748.37
83 1/5/2011 4085 152.58 97.44 623289.30
398042.40
84 4/7/2012 6457 154.06 90.93 994765.42
587135.01
85 3/20/2012 6422 81.73 56.67 524870.06
363934.74

```
86   1/21/2011        8829       47.45      31.79      418936.05
280673.91
87   3/20/2014        5559      255.28     159.42     1419101.52
886215.78
88   5/18/2012         522        9.33       6.92        4870.26
3612.24
89  11/25/2016        4660       47.45      31.79      221117.00
148141.40
90  12/14/2016         948      651.21     524.96      617347.08
497662.08
91   7/11/2014        9389       47.45      31.79      445508.05
298476.31
92   7/24/2012        2021      651.21     524.96     1316095.41
1060944.16
93  12/25/2010        7910      437.20     263.33     3458252.00
2082940.30
94   3/21/2011        8156       47.45      31.79      387002.20
259279.24
95    9/3/2011         888      109.28      35.84       97040.64
31825.92
96  12/28/2011        6267        9.33       6.92       58471.11
43367.64
97   6/29/2016        1485      154.06      90.93      228779.10
135031.05
98    8/8/2015        5767       81.73      56.67      471336.91
326815.89
99   2/15/2012        5367      668.27     502.54     3586605.09
2697132.18

    Total Profit
0      951410.50
1      248406.36
2      224598.75
3       19525.82
4      639077.50
5      285087.64
6      693911.51
7      510216.66
8      152114.20
9      584073.87
10       7828.12
11     306097.92
12     606834.72
13    1487261.02
14     122819.06
15     122865.12
16    1208744.24
17      85033.80
18     634745.90
```

```
19     337937.60
20     714157.00
21     122029.78
22     122686.50
23       5270.67
24     127054.20
25     315574.05
26      13009.18
27      14031.02
28      80241.84
29     392780.10
30    1505888.07
31      53252.50
32     369155.00
33    1367272.50
34     404010.78
35     160609.54
36       9119.44
37     272672.40
38     846885.00
39     503358.75
40      75555.90
41    1257775.58
42     187545.03
43      10795.23
44     435499.20
45      85223.58
46    1541705.29
47       6841.38
48     575916.48
49     159832.50
50     166635.36
51      23150.46
52     359941.17
53     436446.25
54      18405.17
55     255718.08
56     727423.20
57     495007.89
58      46735.86
59    1254472.05
60      60418.38
61     455335.00
62     696647.50
63     449159.04
64     296448.35
65     778415.99
66     216919.36
67     403773.12
```

```
68    1046233.75
69     539196.48
70     632512.50
71       1621.93
72      89904.06
73     766835.04
74    1719922.04
75    1152486.42
76     146875.14
77     235601.16
78     825738.04
79    1678540.98
80     747939.49
81     579000.96
82    1128242.43
83     225246.90
84     407630.41
85     160935.32
86     138262.14
87     532885.74
88       1258.02
89      72975.60
90     119685.00
91     147031.74
92     255151.25
93    1375311.70
94     127722.96
95      65214.72
96      15103.47
97      93748.05
98     144521.02
99     889472.91
```

```
data.shape
```

```
(100, 14)
```

```
data.columns
```

```
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')
```

```
data.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
```

```
data.isnull().sum()
```

```
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
```

```
data.dtypes
```

```
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
dtype: object
```

```
# Changing the data type of different column for model training and
analysis
data['Order Date'] = pd.to_datetime(data['Order Date'])
data['Ship Date'] = pd.to_datetime(data['Ship Date'])

data['Region'] = data['Region'].astype(str)
data['Country'] = data['Country'].astype(str)
data['Item Type'] = data['Item Type'].astype(str)
data['Sales Channel'] = data['Sales Channel'].astype(str)
data['Order Priority'] = data['Order Priority'].astype(str)

# Using describe function on dataframe for getting basic stats of
numerical dataset
data[['Units Sold', 'Unit Price',     'Unit Cost', 'Total Revenue',
'Total Cost',   'Total Profit']].describe()
```

```
        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
count  1.000000e+02
mean   4.416820e+05
std    4.385379e+05
min    1.258020e+03
25%    1.214436e+05
50%    2.907680e+05
75%    6.358288e+05
max    1.719922e+06
```

```
# Adding extra column to dataframe which contain only month, year and
month with year
data['Order Month'] = data['Order Date'].dt.month
data['Order Year'] = data['Order Date'].dt.year

data['Order Date MonthYear'] = data['Order Date'].dt.strftime('%Y-%m')

 data = data.drop(columns=['Order Date'])

# Saving the data dataframe to df
df = data

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

```
Region                  0
Country                 0
Item Type               0
Sales Channel           0
Order Priority          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
Order Month             0
Order Year              0
Order Date MonthYear    0
dtype: int64
```

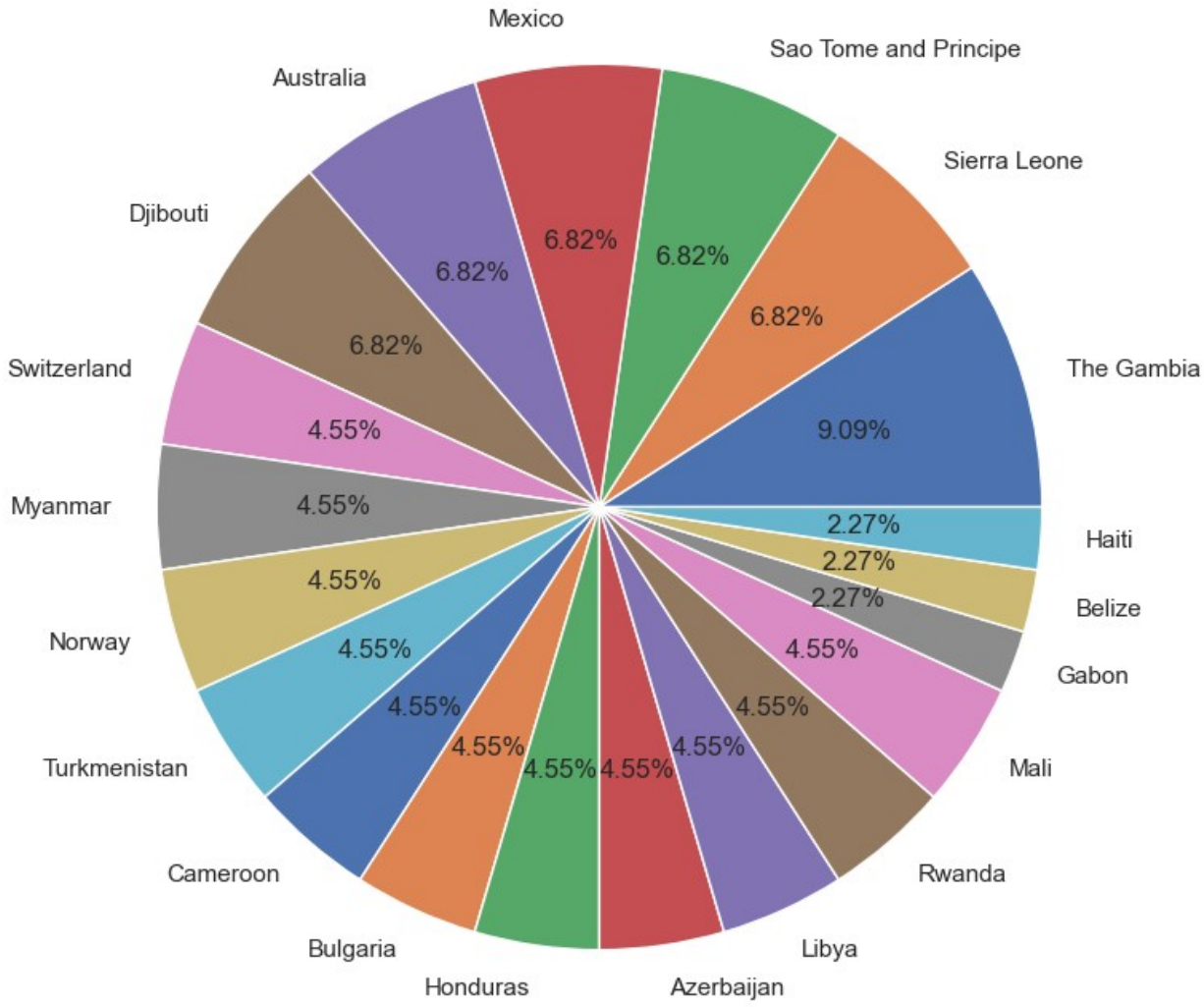
```
# Display total values of all country
pd.set_option('display.max_rows', None)
df['Country'].value_counts()
```

```
Country
The Gambia                          4
Sierra Leone                        3
Sao Tome and Principe               3
Mexico                              3
Australia                           3
Djibouti                            3
Switzerland                         2
Myanmar                             2
Norway                              2
Turkmenistan                        2
Cameroon                            2
Bulgaria                            2
Honduras                            2
Azerbaijan                          2
```

Libya	2
Rwanda	2
Mali	2
Gabon	1
Belize	1
Haiti	1
Lithuania	1
San Marino	1
United Kingdom	1
Austria	1
Fiji	1
Madagascar	1
Cote d'Ivoire	1
Tuvalu	1
Democratic Republic of the Congo	1
Zambia	1
Malaysia	1
Nicaragua	1
Romania	1
Slovenia	1
Kuwait	1
Kenya	1
Iran	1
Pakistan	1
Lebanon	1
Spain	1
Samoa	1
Monaco	1
Laos	1
Saudi Arabia	1
Federated States of Micronesia	1
Slovakia	1
Lesotho	1
Albania	1
Russia	1
Solomon Islands	1
Angola	1
Burkina Faso	1
Republic of the Congo	1
Senegal	1
Kyrgyzstan	1
Cape Verde	1
Bangladesh	1
Mongolia	1
Sri Lanka	1
East Timor	1
Portugal	1
New Zealand	1
Moldova	1

```
France                                  1
Kiribati                                1
South Sudan                             1
Costa Rica                              1
Syria                                   1
Brunei                                  1
Niger                                   1
Grenada                                 1
Comoros                                 1
Iceland                                 1
Macedonia                               1
Mauritania                              1
Mozambique                              1
Name: count, dtype: int64
```

```
import matplotlib.pyplot as plt
country_names = df.Country.value_counts().index
country_val = df.Country.value_counts().values
# Pie Chart for top 20 country
fig,ax = plt.subplots(figsize=(9,9))
ax.pie(country_val[:20],labels=country_names[:20],autopct='%1.2f%%')
plt.show()
```




```
#Total Profit
np.mean(df['Total Profit'])
```

```
441681.98399999994
```

```
np.max(df['Total Profit'])
```

```
1719922.04
```

```
np.min(df['Total Profit'])
```

```
1258.02
```

```
np.var(df['Total Profit'])
```

```
190392340968.9648
```

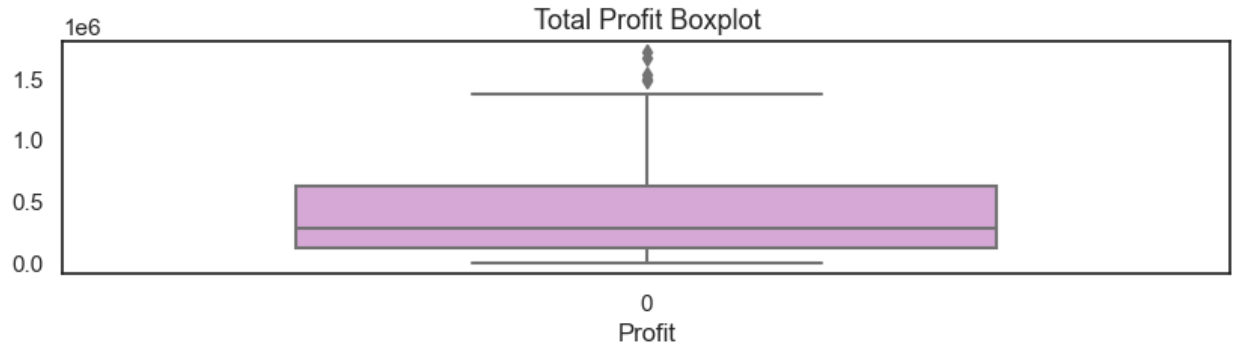
```
np.std(df['Total Profit'])
```

```
436339.7082193699
```

```
np.average(df['Total Profit'])
```

```
441681.98399999994
```

```
import matplotlib.pyplot as plt
sns.set(style='white')
fig, ax = plt.subplots(figsize=(10, 2))
sns.boxplot(data['Total Profit'], color="plum", width=.6)

plt.title('Total Profit Boxplot', fontsize=13)
plt.xlabel('Profit')
plt.show()
```




```
def detect_outliers(dataframe, column):
    threshold = 2     ## 2rd standard deviation
    mean = np.mean(column)
    std = np.std(column)
    outliers = []

    for i, value in enumerate(column):
        z_score = (value - mean) / std
        if np.abs(z_score) > threshold:
            outliers.append(i)
            print(dataframe.loc[i])

    return outliers

outliers = detect_outliers(df, df["Total Profit"])
```

```
Region                          Central America and the Caribbean
Country                                                  Honduras
Item Type                                                Household
Sales Channel                                             Offline
```

```
Order Priority                                      H
Order ID                                    522840487
Ship Date                     2017-02-13 00:00:00
Units Sold                                       8974
Unit Price                                     668.27
Unit Cost                                      502.54
Total Revenue                                5997054.98
Total Cost                                   4509793.96
Total Profit                                 1487261.02
Order Month                                         2
Order Year                                       2017
Order Date MonthYear                          2017-02
Name: 13, dtype: object
Region                              Europe
Country                         Switzerland
Item Type                         Cosmetics
Sales Channel                       Offline
Order Priority                            M
Order ID                          249693334
Ship Date                2012-10-20 00:00:00
Units Sold                             8661
Unit Price                            437.2
Unit Cost                            263.33
Total Revenue                     3786589.2
Total Cost                       2280701.13
Total Profit                     1505888.07
Order Month                               9
Order Year                             2012
Order Date MonthYear                2012-09
Name: 30, dtype: object
Region                                 Asia
Country                             Myanmar
Item Type                         Household
Sales Channel                       Offline
Order Priority                            H
Order ID                          177713572
Ship Date                2015-03-01 00:00:00
Units Sold                             8250
Unit Price                           668.27
Unit Cost                            502.54
Total Revenue                     5513227.5
Total Cost                        4145955.0
Total Profit                      1367272.5
Order Month                               1
Order Year                             2015
Order Date MonthYear                2015-01
Name: 33, dtype: object
Region                              Europe
Country                            Iceland
```

```
Item Type                            Cosmetics
Sales Channel                           Online
Order Priority                               C
Order ID                             331438481
Ship Date                  2016-12-31 00:00:00
Units Sold                                8867
Unit Price                               437.2
Unit Cost                               263.33
Total Revenue                          3876652.4
Total Cost                            2334947.11
Total Profit                          1541705.29
Order Month                                 12
Order Year                                2016
Order Date MonthYear                   2016-12
Name: 46, dtype: object
Region                  Middle East and North Africa
Country                                     Pakistan
Item Type                                   Cosmetics
Sales Channel                                 Offline
Order Priority                                      L
Order ID                                    231145322
Ship Date                       2013-08-16 00:00:00
Units Sold                                     9892
Unit Price                                    437.2
Unit Cost                                    263.33
Total Revenue                               4324782.4
Total Cost                                 2604860.36
Total Profit                               1719922.04
Order Month                                         7
Order Year                                       2013
Order Date MonthYear                          2013-07
Name: 74, dtype: object
Region                  Australia and Oceania
Country                                     Samoa
Item Type                                 Cosmetics
Sales Channel                                Online
Order Priority                                    H
Order ID                                  670854651
Ship Date                   2013-08-07 00:00:00
Units Sold                                     9654
Unit Price                                    437.2
Unit Cost                                    263.33
Total Revenue                              4220728.8
Total Cost                                2542187.82
Total Profit                              1678540.98
Order Month                                         7
Order Year                                       2013
Order Date MonthYear                          2013-07
Name: 79, dtype: object
```

```
Region                                 Europe
Country                                Romania
Item Type                             Cosmetics
Sales Channel                            Online
Order Priority                                H
Order ID                              660643374
Ship Date                   2010-12-25 00:00:00
Units Sold                                 7910
Unit Price                                437.2
Unit Cost                                263.33
Total Revenue                          3458252.0
Total Cost                             2082940.3
Total Profit                           1375311.7
Order Month                                   11
Order Year                                  2010
Order Date MonthYear                     2010-11
Name: 93, dtype: object
```

```
# Print rows where outlier is present for the Total Profit column
value
print(outliers)
```

```
[13, 30, 33, 46, 74, 79, 93]
```

```
list_length = len(outliers)
```

```
# Print the number of values in the list
print("The list has", list_length, "outliers in Total Profit column of
dataframe data ")
```

```
The list has 7 outliers in Total Profit column of dataframe data
```

```
#Total cost
```

```
np.mean(df['Total Cost'])
```

```
931805.6991000001
```

```
np.median(df['Total Cost'])
```

```
363566.385
```

```
np.std(df['Total Cost'])
```

```
1078504.9435267276
```

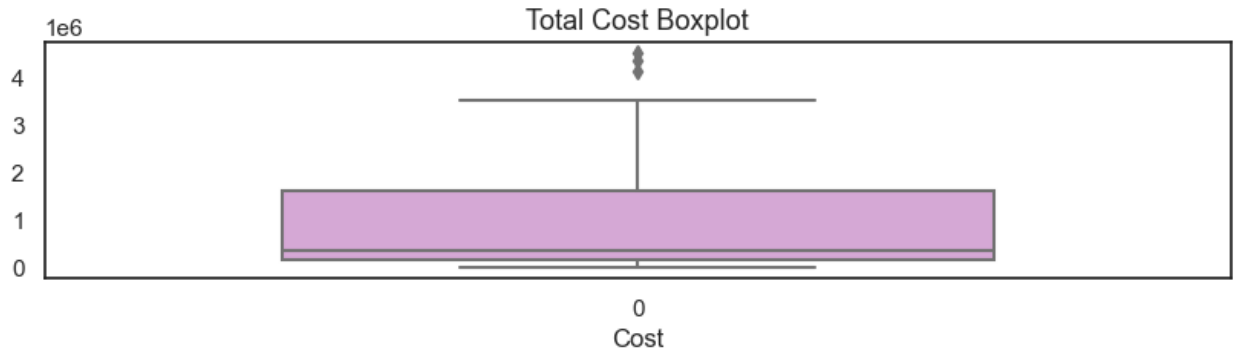
```
np.var(df['Total Cost'])
```

```
1163172913211.59
```

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

```
752314.36
```

```
sns.set(style='white')
fig, ax = plt.subplots(figsize=(10, 2))
sns.boxplot(data['Total Cost'], color="plum", width=.6)

plt.title('Total Cost Boxplot', fontsize=13)
plt.xlabel('Cost')
plt.show()
```




```
def detect_outliers(dataframe, column):
    threshold = 2     ## 3rd standard deviation
    mean = np.mean(column)
    std = np.std(column)
    outliers = []

    for i, value in enumerate(column):
        z_score = (value - mean) / std
        if np.abs(z_score) > threshold:
            outliers.append(i)
            print(dataframe.loc[i])

    return outliers

outliers = detect_outliers(df, df["Total Cost"])
```

```
Region                  Central America and the Caribbean
Country                                           Honduras
Item Type                                        Household
Sales Channel                                      Offline
Order Priority                                           H
Order ID                                         522840487
Ship Date                              2017-02-13 00:00:00
Units Sold                                            8974
Unit Price                                          668.27
Unit Cost                                           502.54
Total Revenue                                     221117.0
Total Cost                                      4509793.96
Total Profit                                    1487261.02
Order Month                                              2
```

```
Order Year                                          2017
Order Date MonthYear                             2017-02
Name: 13, dtype: object
Region                                             Asia
Country                                         Myanmar
Item Type                                     Household
Sales Channel                                   Offline
Order Priority                                        H
Order ID                                      177713572
Ship Date                         2015-03-01 00:00:00
Units Sold                                         8250
Unit Price                                       668.27
Unit Cost                                        502.54
Total Revenue                                 380512.96
Total Cost                                    4145955.0
Total Profit                                  1367272.5
Order Month                                           1
Order Year                                         2015
Order Date MonthYear                            2015-01
Name: 33, dtype: object
Region                                             Asia
Country                                          Brunei
Item Type                           Office Supplies
Sales Channel                                    Online
Order Priority                                        L
Order ID                                      320009267
Ship Date                         2012-05-08 00:00:00
Units Sold                                         6708
Unit Price                                       651.21
Unit Cost                                        524.96
Total Revenue                                 22312.29
Total Cost                                  3521431.68
Total Profit                                  846885.0
Order Month                                           4
Order Year                                         2012
Order Date MonthYear                            2012-04
Name: 38, dtype: object
Region                                           Europe
Country                                       Lithuania
Item Type                           Office Supplies
Sales Channel                                   Offline
Order Priority                                        H
Order ID                                      166460740
Ship Date                         2010-11-17 00:00:00
Units Sold                                         8287
Unit Price                                       651.21
Unit Cost                                        524.96
Total Revenue                                1245112.92
Total Cost                                   4350343.52
```

```
Total Profit                        1046233.75
Order Month                             10
Order Year                            2010
Order Date MonthYear                2010-10
Name: 68, dtype: object
Region                         North America
Country                               Mexico
Item Type                          Household
Sales Channel                        Offline
Order Priority                             C
Order ID                           986435210
Ship Date                2014-12-12 00:00:00
Units Sold                            6954
Unit Price                          668.27
Unit Cost                           502.54
Total Revenue                      1419101.52
Total Cost                         3494663.16
Total Profit                       1152486.42
Order Month                             11
Order Year                            2014
Order Date MonthYear                2014-11
Name: 75, dtype: object
```

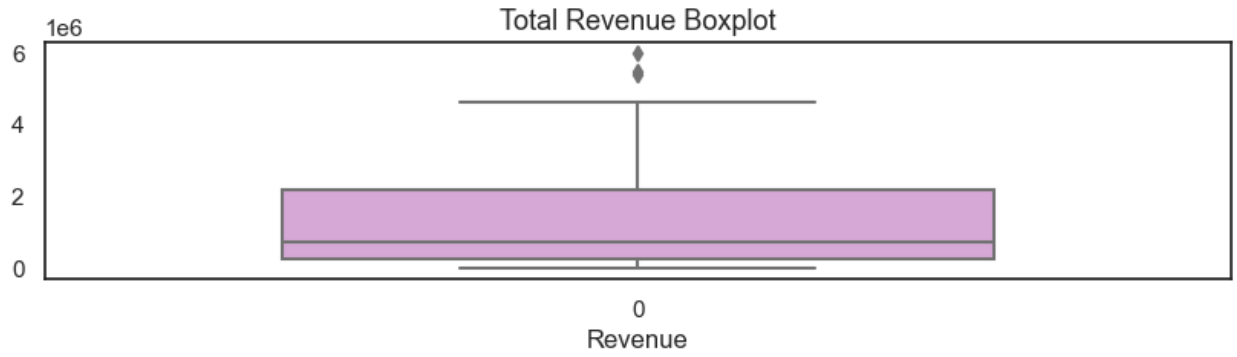
```
print(outliers)
```

```
[13, 33, 38, 68, 75]
```

```
list_length = len(outliers)

# Print the number of values in the list
print("The list has", list_length, "outliers in Total Cost column of
dataframe data ")
```

```
The list has 5 outliers in Total Cost column of dataframe data
```

```
#REVENUE

sns.set(style='white')
fig, ax = plt.subplots(figsize=(10, 2))
sns.boxplot(data['Total Revenue'], color="plum", width=.6)

plt.title('Total Revenue Boxplot', fontsize=13)
plt.xlabel('Revenue')
plt.show()
```




```python
def detect_outliers(dataframe, column):
    threshold = 2     ## 3rd standard deviation
    mean = np.mean(column)
    std = np.std(column)
    outliers = []

    for i, value in enumerate(column):
        z_score = (value - mean) / std
        if np.abs(z_score) > threshold:
            outliers.append(i)
            print(dataframe.loc[i])

    return outliers

outliers = detect_outliers(df, df["Total Revenue"])
```

```
Region                      Sub-Saharan Africa
Country                            The Gambia
Item Type                           Baby Food
Sales Channel                         Offline
Order Priority                              M
Order ID                            494747245
Ship Date                 2014-03-20 00:00:00
Units Sold                               5559
Unit Price                             255.28
Unit Cost                              159.42
Total Revenue                         4647149.58
Total Cost                             886215.78
Total Profit                           532885.74
Order Month                                 2
Order Year                               2014
Order Date MonthYear                  2014-02
Name: 87, dtype: object
Region                     Middle East and North Africa
Country                                          Kuwait
Item Type                                        Fruits
Sales Channel                                    Online
Order Priority                                        M
```

```
Order ID                                  513417565
Ship Date                       2012-05-18 00:00:00
Units Sold                                    522
Unit Price                                   9.33
Unit Cost                                    6.92
Total Revenue                             4324782.4
Total Cost                                  3612.24
Total Profit                                1258.02
Order Month                                       4
Order Year                                     2012
Order Date MonthYear                          2012-04
Name: 88, dtype: object
Region                   Australia and Oceania
Country                                Australia
Item Type                              Beverages
Sales Channel                            Offline
Order Priority                                 H
Order ID                               240470397
Ship Date                    2014-07-11 00:00:00
Units Sold                                  9389
Unit Price                                 47.45
Unit Cost                                  31.79
Total Revenue                           5396577.27
Total Cost                               298476.31
Total Profit                             147031.74
Order Month                                    7
Order Year                                  2014
Order Date MonthYear                       2014-07
Name: 91, dtype: object
Region                     Sub-Saharan Africa
Country                                  Mali
Item Type                              Clothes
Sales Channel                           Online
Order Priority                               M
Order ID                             512878119
Ship Date                  2011-09-03 00:00:00
Units Sold                                  888
Unit Price                               109.28
Unit Cost                                 35.84
Total Revenue                          5513227.5
Total Cost                              31825.92
Total Profit                            65214.72
Order Month                                    7
Order Year                                  2011
Order Date MonthYear                       2011-07
Name: 95, dtype: object
Region                                      Asia
Country                                Malaysia
Item Type                                 Fruits
```

```
Sales Channel                          Offline
Order Priority                               L
Order ID                             810711038
Ship Date                2011-12-28 00:00:00
Units Sold                                6267
Unit Price                                9.33
Unit Cost                                 6.92
Total Revenue                       4368316.68
Total Cost                            43367.64
Total Profit                          15103.47
Order Month                                 11
Order Year                                2011
Order Date MonthYear                   2011-11
Name: 96, dtype: object
Region                      Sub-Saharan Africa
Country                               Mozambique
Item Type                              Household
Sales Channel                            Offline
Order Priority                               L
Order ID                              665095412
Ship Date                2012-02-15 00:00:00
Units Sold                                5367
Unit Price                              668.27
Unit Cost                               502.54
Total Revenue                       5997054.98
Total Cost                          2697132.18
Total Profit                         889472.91
Order Month                                 2
Order Year                                2012
Order Date MonthYear                   2012-02
Name: 99, dtype: object
```

```
print(outliers)
```

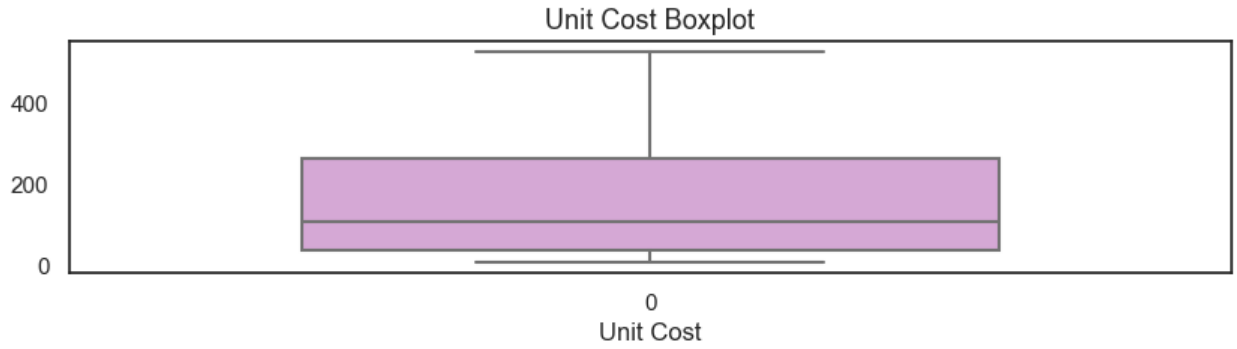
```
[87, 88, 91, 95, 96, 99]
```

```
list_length = len(outliers)

# Print the number of values in the list
print("The list has", list_length, "outliers in Total Revenue column
of dataframe data ")
```

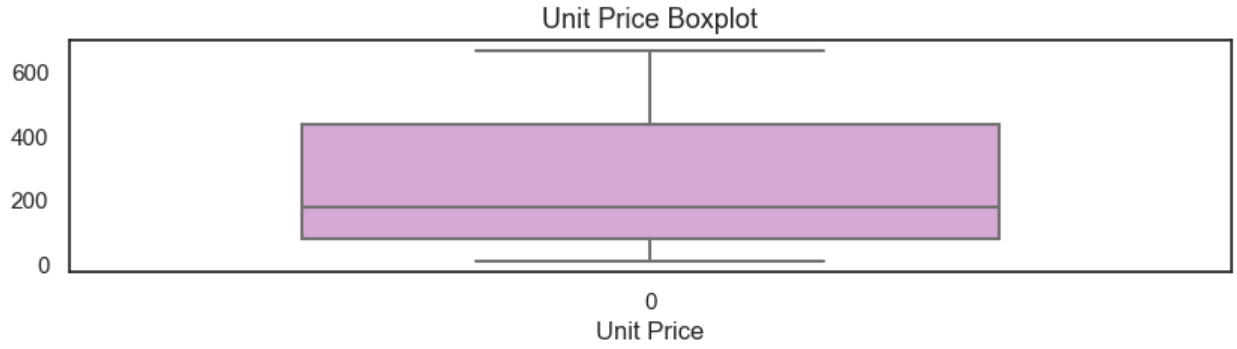
```
The list has 6 outliers in Total Revenue column of dataframe data
```

```
#unit cost

sns.set(style='white')
fig, ax = plt.subplots(figsize=(10, 2))
sns.boxplot(data['Unit Cost'], color="plum", width=.6)
plt.title('Unit Cost Boxplot', fontsize=13)
```

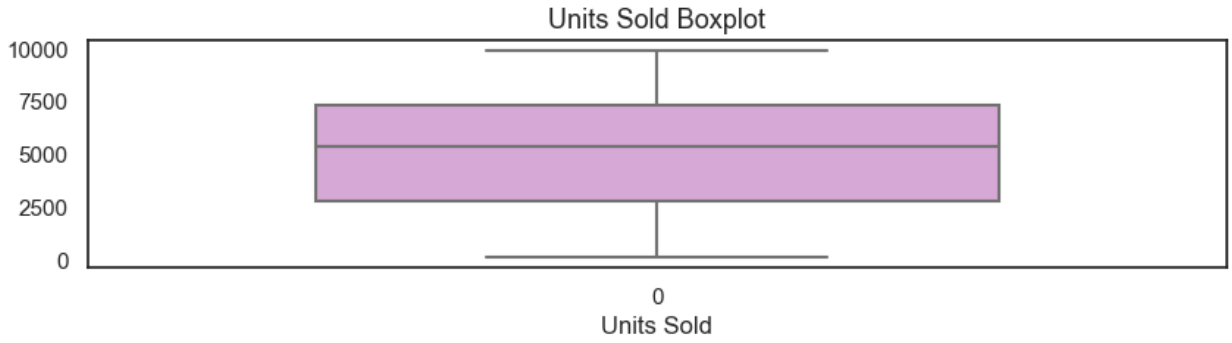
```
plt.xlabel('Unit Cost')
plt.show()
```



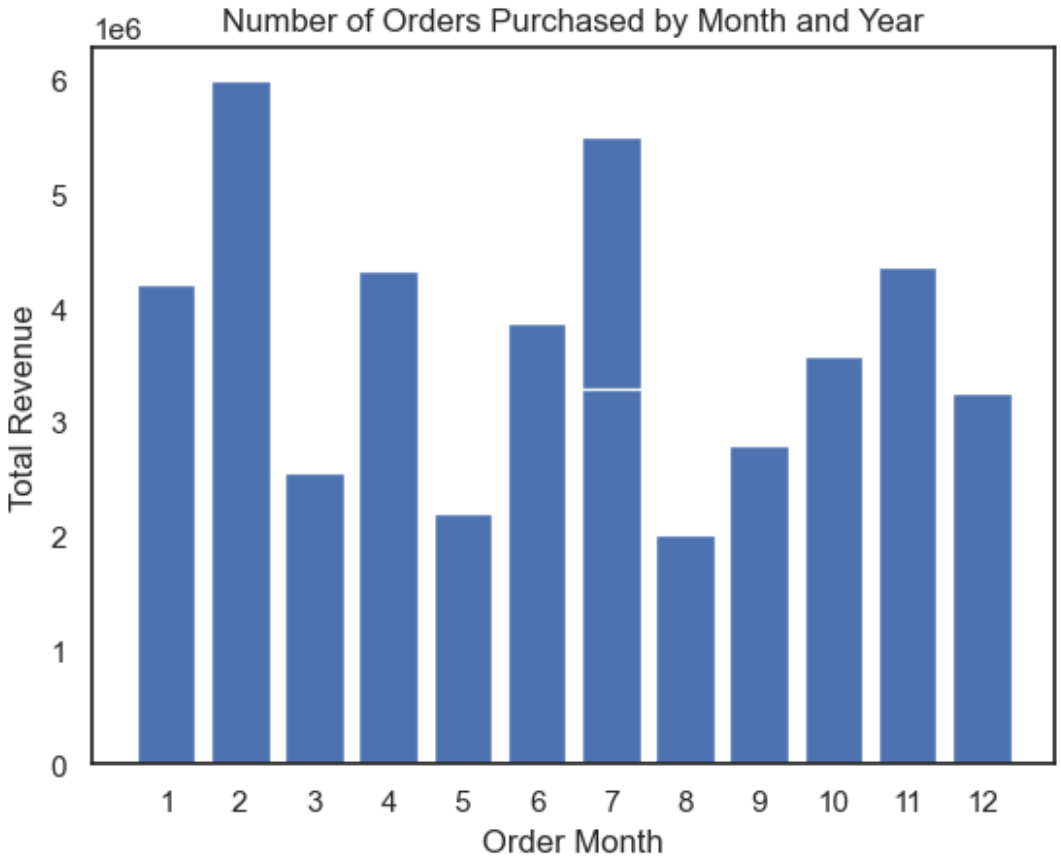

```
#unit price

sns.set(style='white')
fig, ax = plt.subplots(figsize=(10, 2))
sns.boxplot(data['Unit Price'], color="plum", width=.6)
plt.title('Unit Price Boxplot', fontsize=13)
plt.xlabel('Unit Price')
plt.show()
```




```
#unit sold

sns.set(style='white')
fig, ax = plt.subplots(figsize=(10, 2))
sns.boxplot(data['Units Sold'], color="plum", width=.6)

plt.title('Units Sold Boxplot', fontsize=13)
plt.xlabel('Units Sold')
plt.show()
```



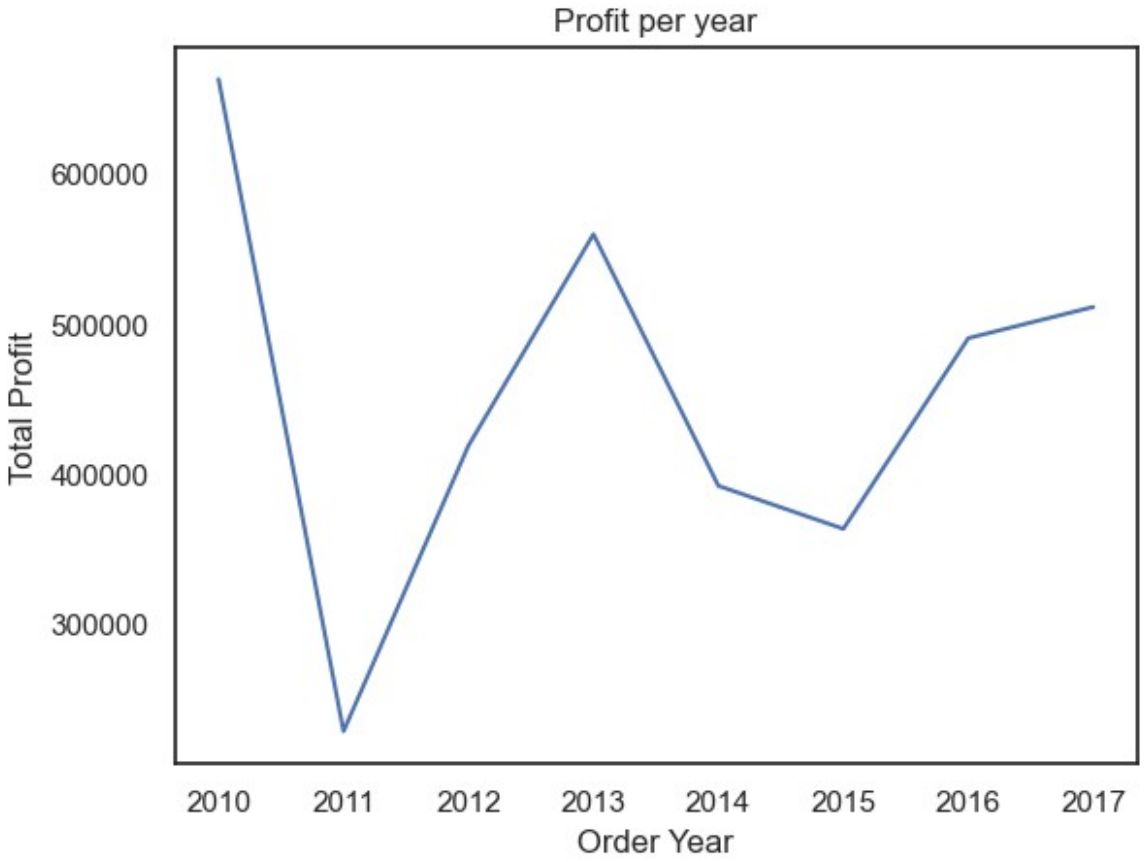

```
# A bar chart for Total Revenue and Order Month
plt.bar(df['Order Month'], df['Total Revenue'])
plt.title('Number of Orders Purchased by Month and Year')
plt.xticks([1,2,3,4,5,6,7,8,9,10,11,12])
plt.xlabel('Order Month')
plt.ylabel('Total Revenue')
plt.show()
```




```
# Plot line graph of Total Profit and Order Year
df.groupby('Order Year')['Total Profit'].mean().plot()
plt.xlabel('Order Year')
```

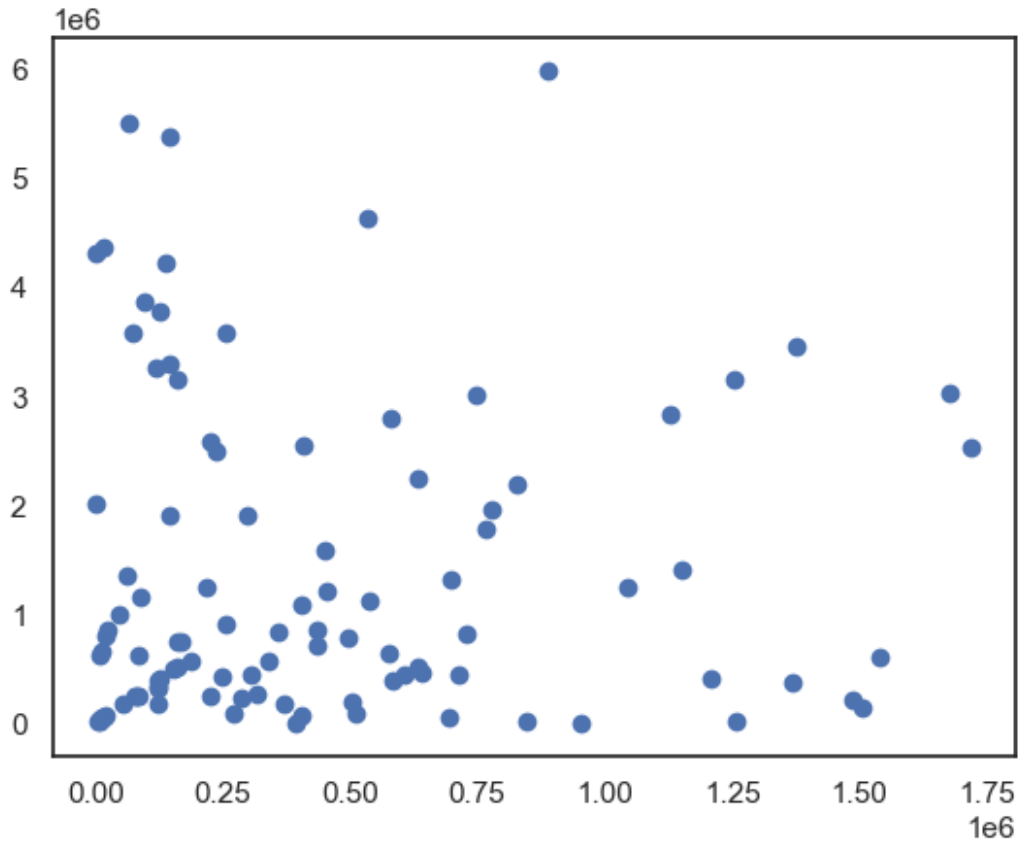
```
plt.ylabel('Total Profit')
plt.title('Profit per year')
```

```
Text(0.5, 1.0, 'Profit per year')
```



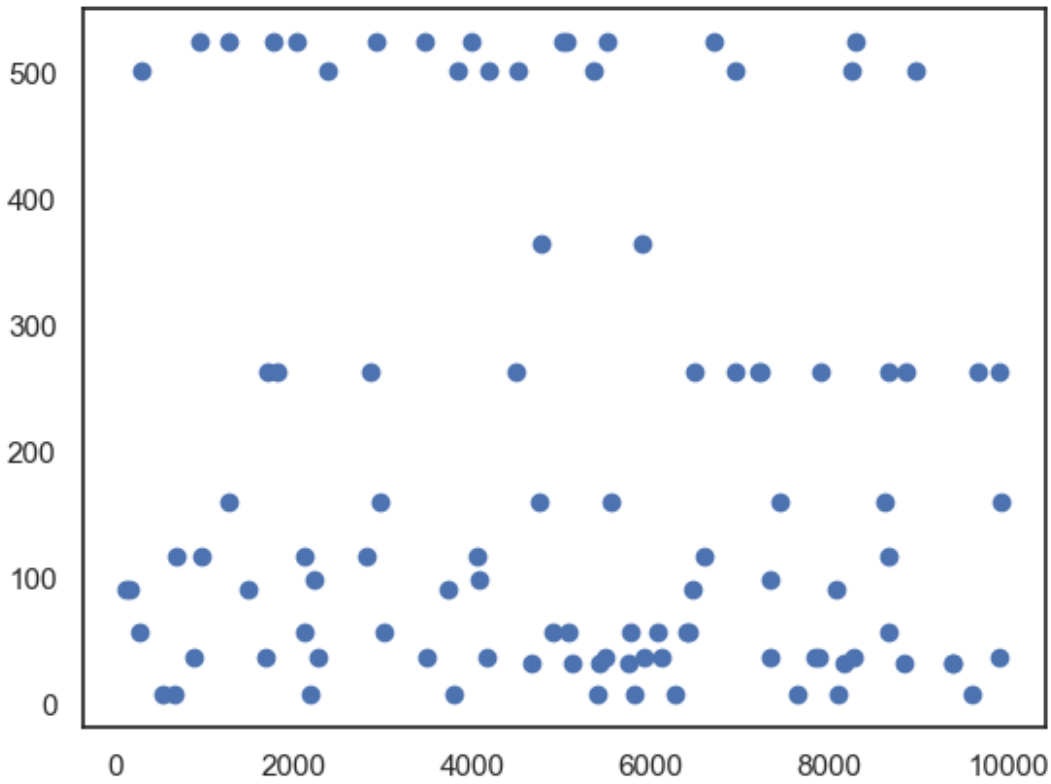

```
plt.scatter(df['Total Profit'],df['Total Revenue'])
```

```
<matplotlib.collections.PathCollection at 0x253f60110d0>
```



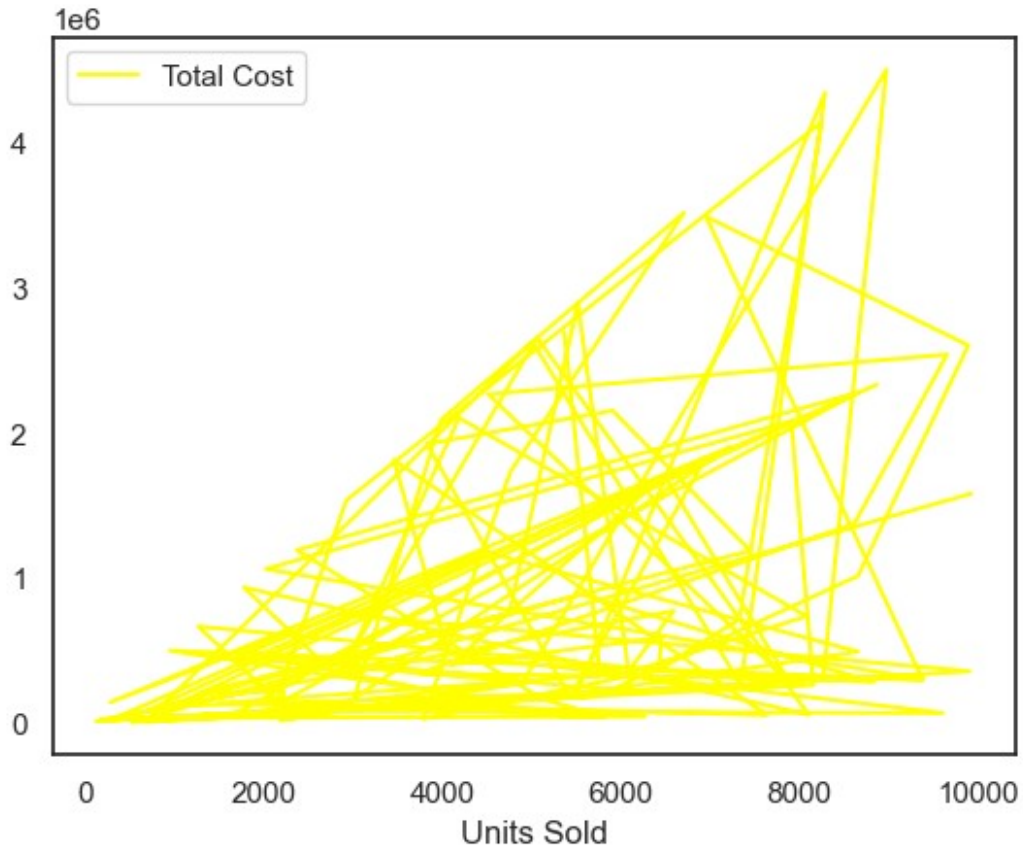

```
plt.scatter(df['Units Sold'],df['Unit Cost'])
```

```
<matplotlib.collections.PathCollection at 0x253f79fdb50>
```



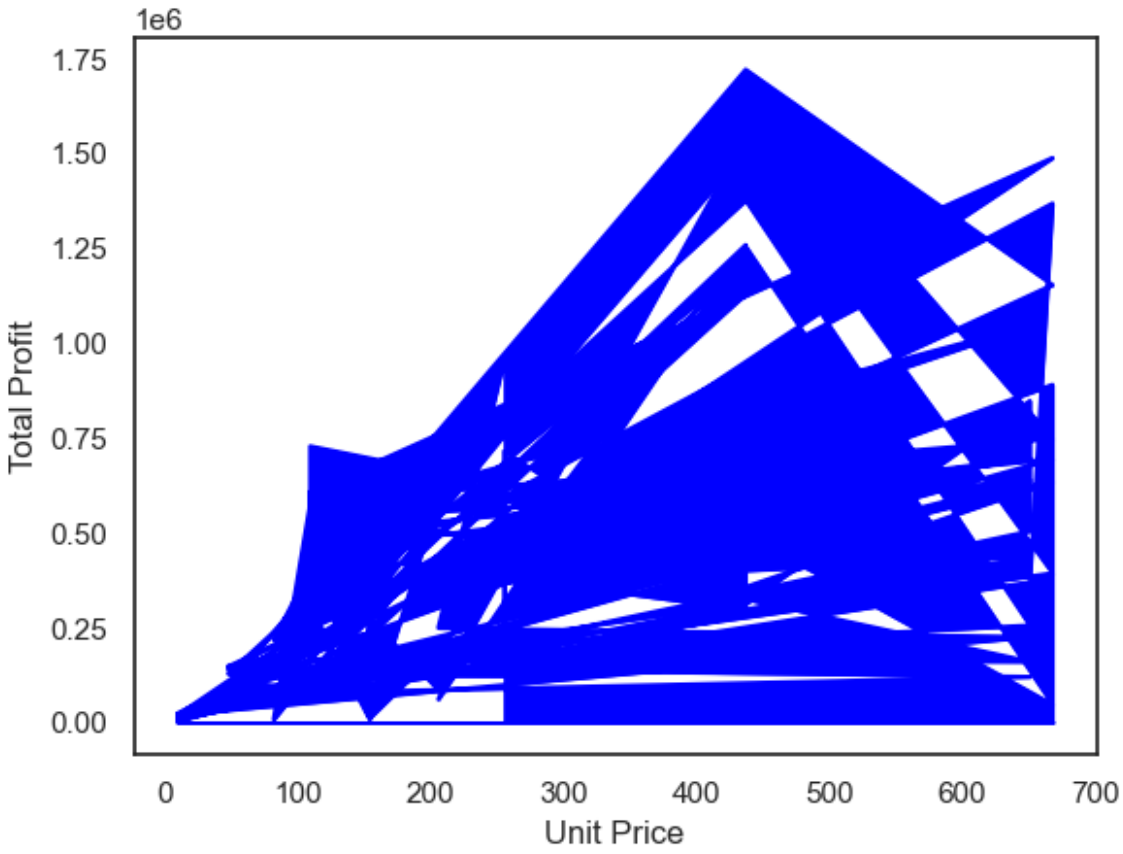

```
df.plot.line(x='Units Sold',y='Total Cost',subplots=True,color={'Total
Cost': 'yellow'})
```

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



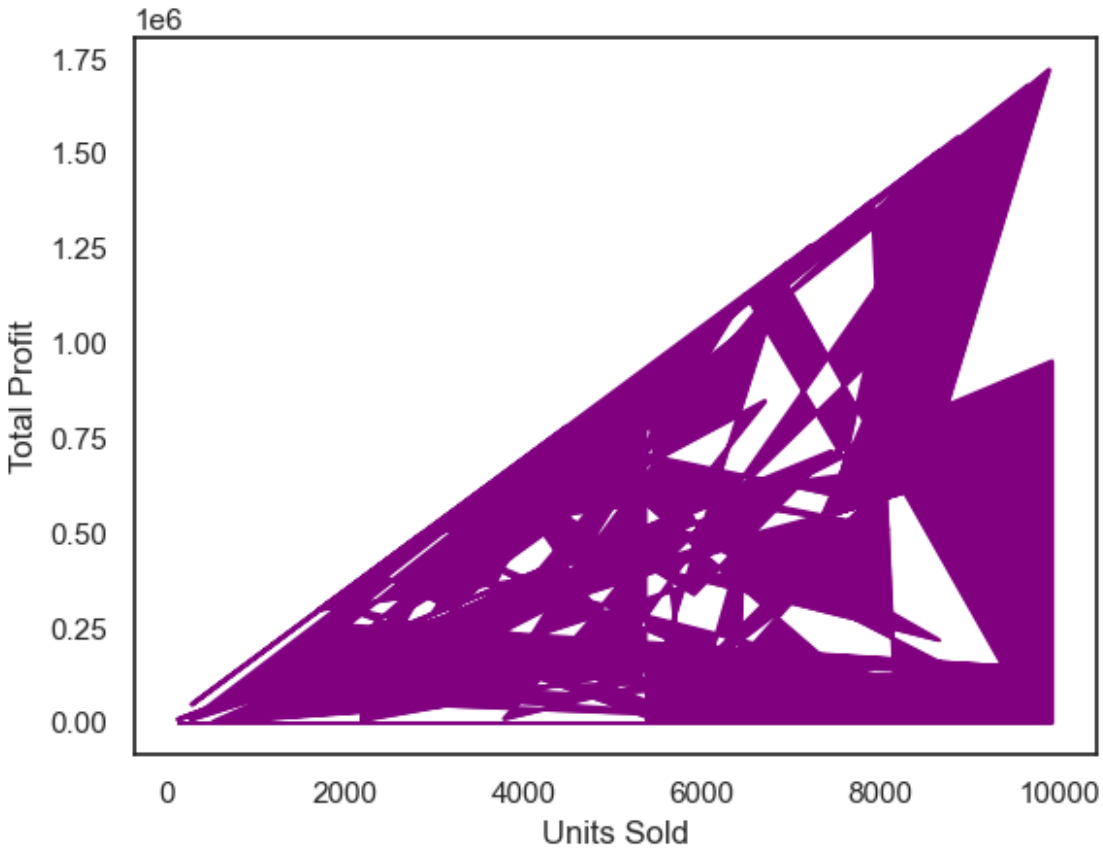

```
area_plot = df.plot.area(x='Unit Price',y='Total
Profit',color='blue',stacked=True,legend=None)
plt.ylabel('Total Profit')
```

```
Text(0, 0.5, 'Total Profit')
```

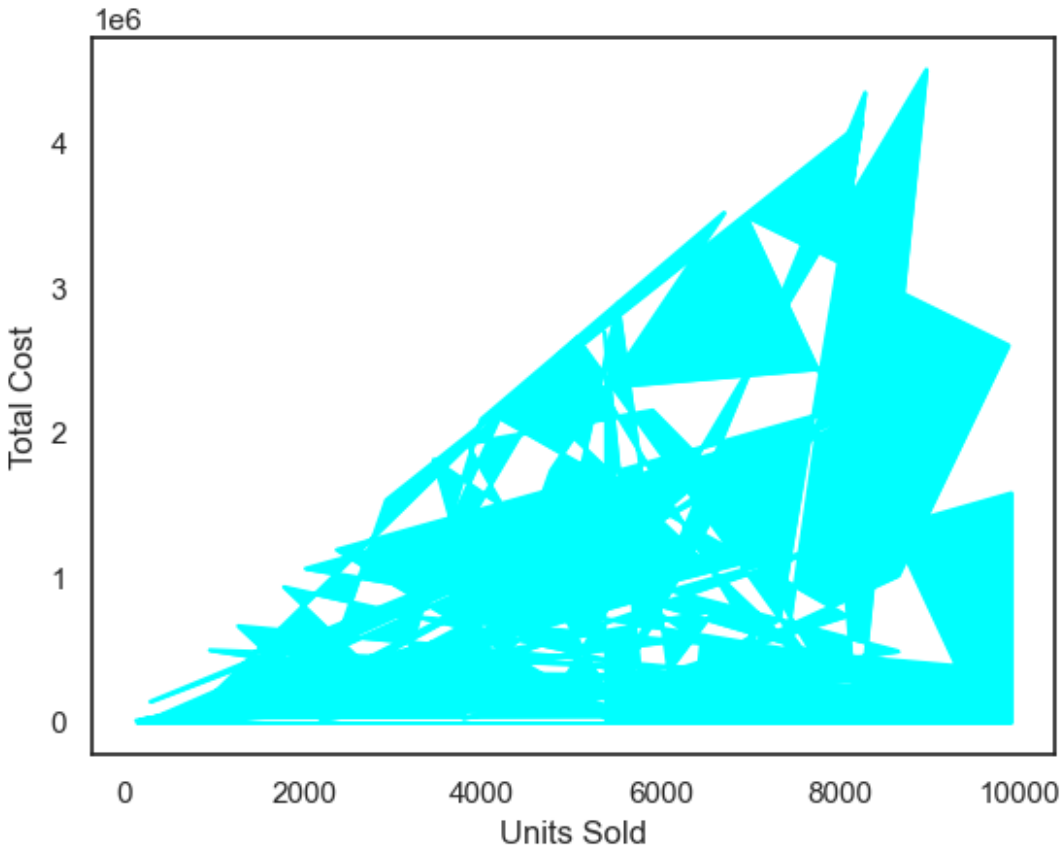



```
df.plot.area(x='Units Sold',y='Total
Profit',color='purple',legend=None)
plt.ylabel('Total Profit')
```

```
Text(0, 0.5, 'Total Profit')
```

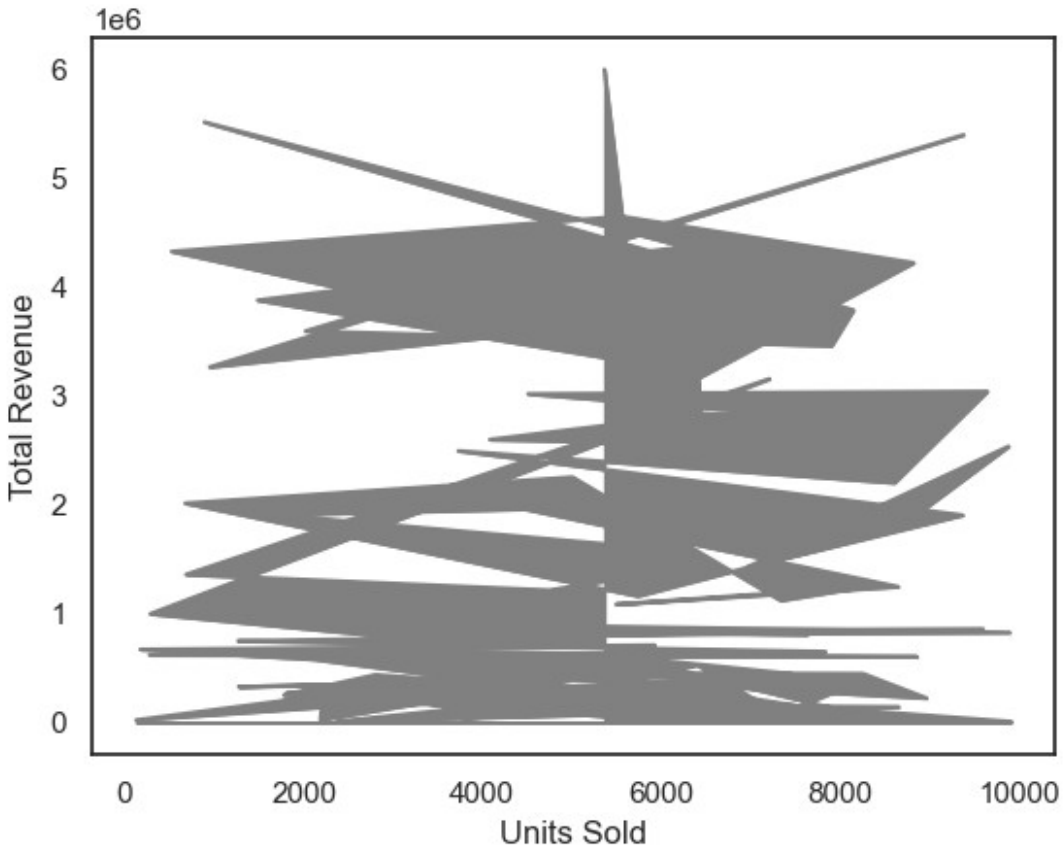



```
df.plot.area(x='Units Sold',y='Total Cost',color='aqua',legend=None)
plt.ylabel('Total Cost')

Text(0, 0.5, 'Total Cost')
```



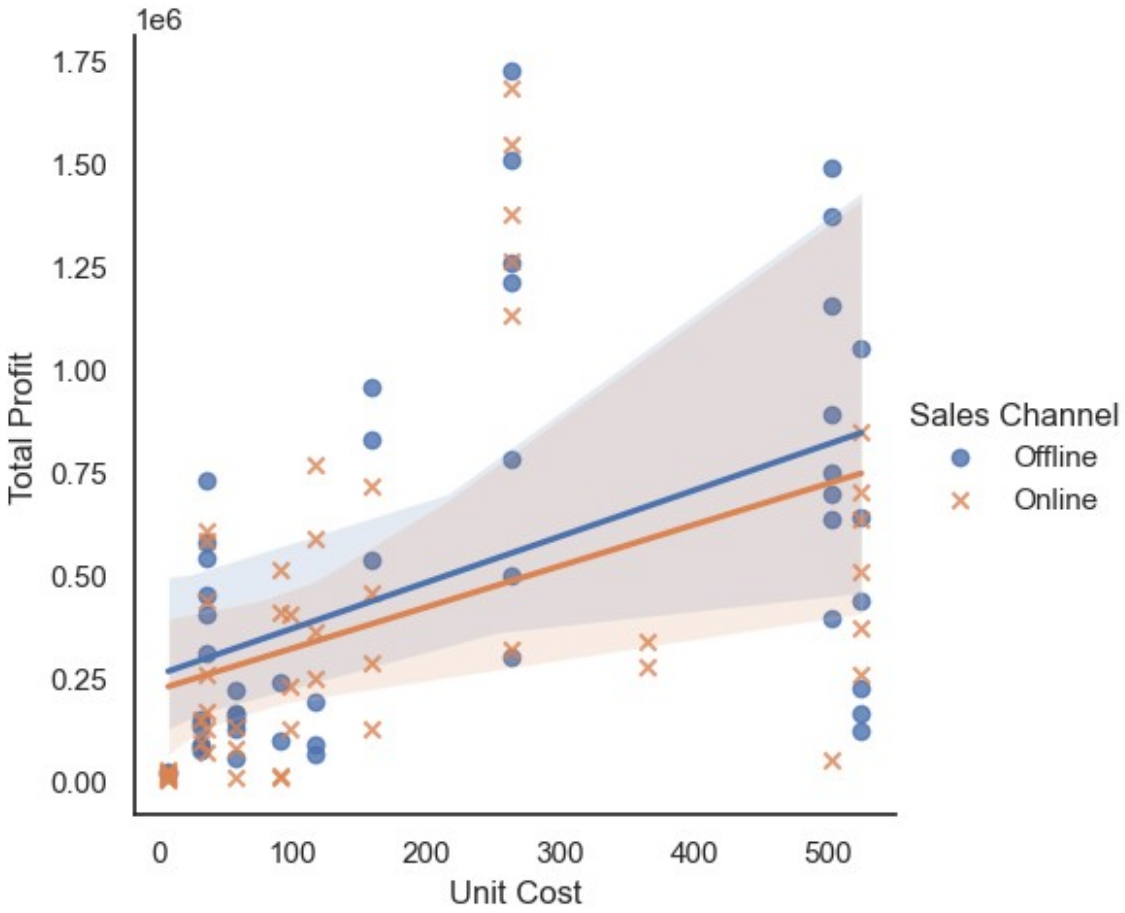

```
df.plot.area(x='Units Sold',y='Total
Revenue',color='grey',legend=None)
plt.ylabel('Total Revenue')
```

```
Text(0, 0.5, 'Total Revenue')
```

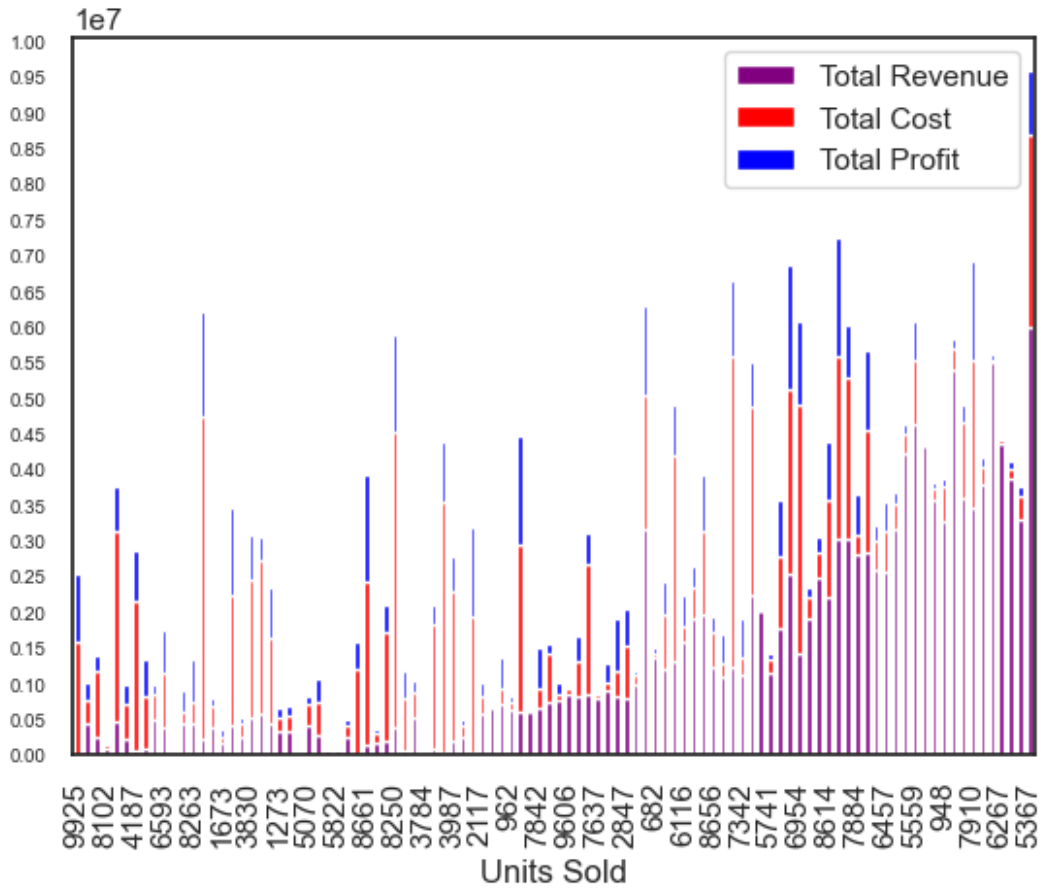



```
sns.lmplot(x='Unit Cost',y='Total
Profit',data=df,height=5,aspect=1,hue='Sales
Channel',logx=False,truncate=True,ci=100,y_jitter=2.2,scatter=True,fit
_reg=True,markers=['o','x'])
```

```
<seaborn.axisgrid.FacetGrid at 0x253f60a4f50>
```




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

bar_plot = df.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)
```




```
# Calculating the total revenue for each group with respect to Item
Type and then sorting then in descending order
revenue_by_category = df.groupby('Item Type')['Total
Revenue'].sum().sort_values(ascending=False)
revenue_by_category
```

```
Item Type
Cosmetics          21136509.24
Beverages          20551626.25
Clothes            17021358.98
Office Supplies    14403310.76
Household          12618074.65
Fruits             12599539.92
Personal Care      10576116.00
Vegetables          9713153.18
Baby Food           9062365.85
Cereal              5995019.60
Snacks              3007117.98
Meat                 664575.90
Name: Total Revenue, dtype: float64
```

```
# Calculating the total profit for each group with respect to Item
Type and then sorting then in descending order
profit_by_category = df.groupby('Item Type')['Total
Profit'].sum().sort_values(ascending=False)
profit_by_category
```

```
Item Type
Cosmetics          14556048.66
Household           7412605.71
Office Supplies     5929583.75
Clothes             5233334.40
Baby Food           3886643.70
Cereal              2292443.43
Vegetables          1265819.63
Personal Care       1220622.48
Beverages            888047.28
Snacks               751944.18
Meat                 610610.00
Fruits               120495.18
Name: Total Profit, dtype: float64
```

```
# Calculating correlation of 'Total Revenue', 'Total Cost' and 'Total
Profit' columns present in dataframe
print(df[['Total Revenue', 'Total Cost', 'Total Profit']].corr())
```

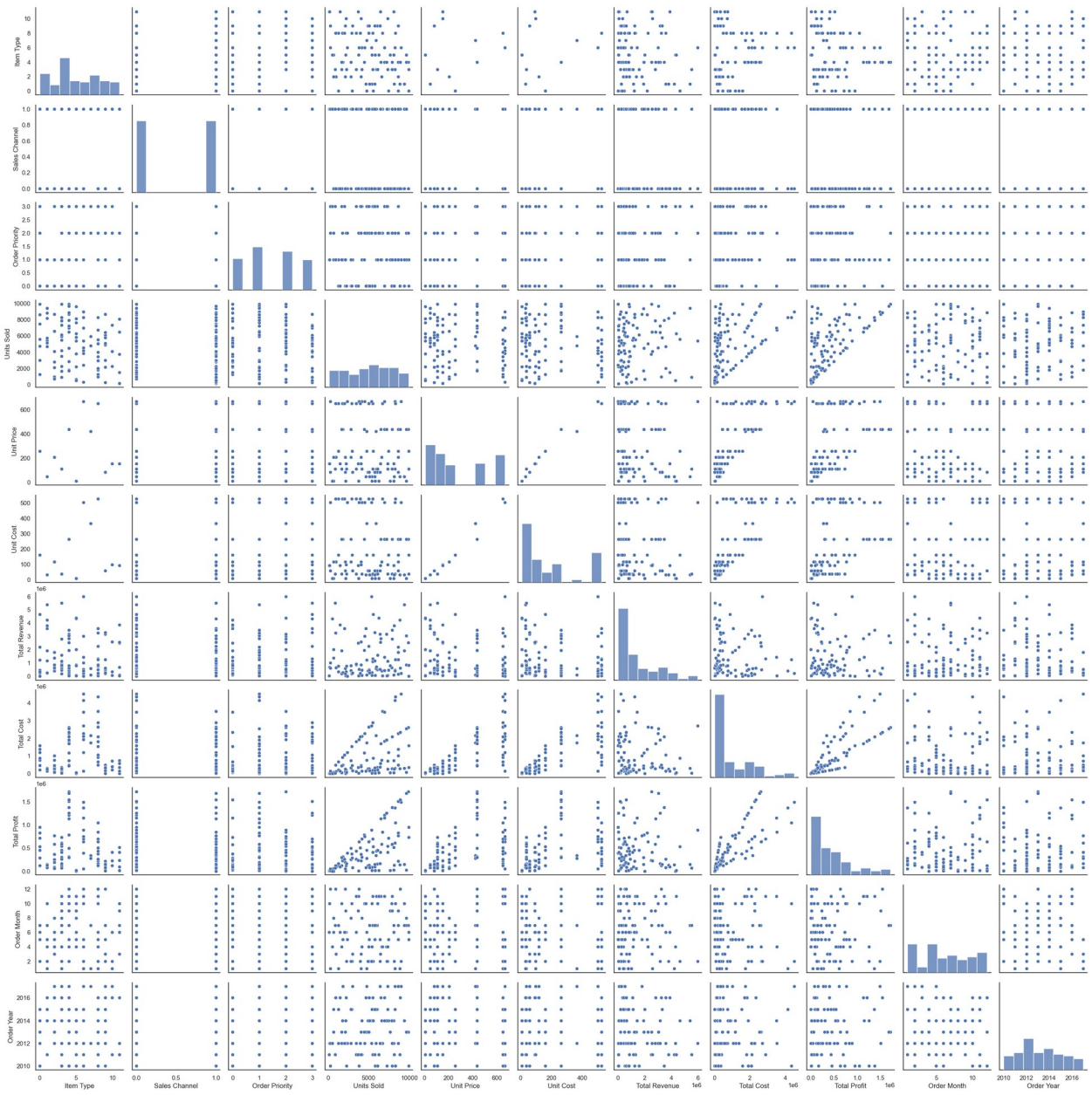
```
                Total Revenue  Total Cost  Total Profit
Total Revenue       1.000000   -0.021539      0.029844
Total Cost         -0.021539    1.000000      0.804091
Total Profit        0.029844    0.804091      1.000000
```

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

from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
df["Item Type"] = le.fit_transform(df["Item Type"])
df["Sales Channel"] = le.fit_transform(df["Sales Channel"])
df["Order Priority"] = le.fit_transform(df["Order Priority"])

# Drop columns Region, Country, Order Date MonthYear, Order ID and
Ship Date
df = df.drop("Region", axis=1)
df = df.drop("Country", axis=1)
df = df.drop("Order Date MonthYear", axis=1)
df = df.drop("Order ID", axis=1)
df = df.drop("Ship Date", axis=1)

df.head()
```

```
   Item Type  Sales Channel  Order Priority  Units Sold  Unit Price  \
0          0              0               1        9925      255.28
```

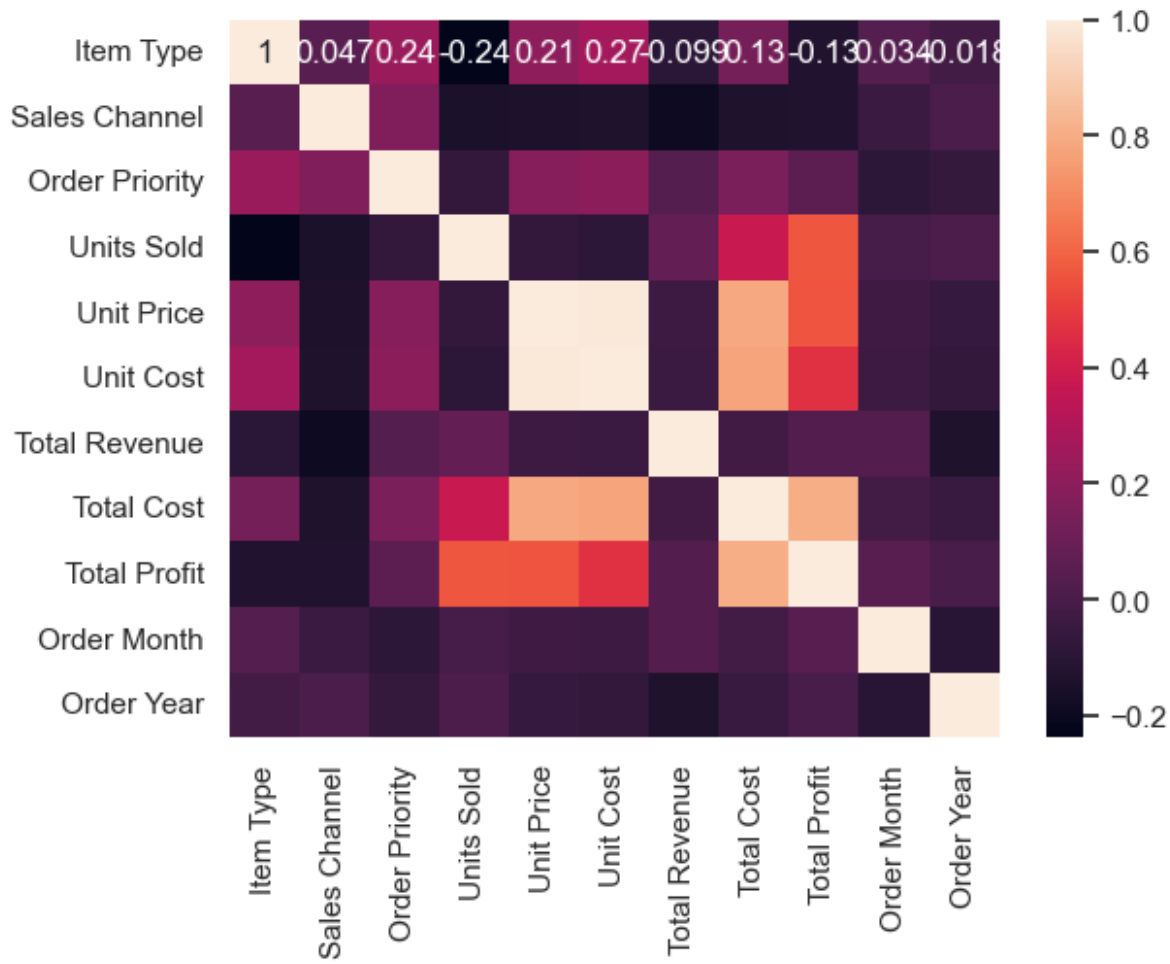
```
1          2              1               0        2804      205.70
2          8              0               2        1779      651.21
3          5              1               0        8102        9.33
4          8              0               2        5062      651.21

   Unit Cost  Total Revenue  Total Cost  Total Profit  Order Month
Order Year
0     159.42        4870.26  1582243.50     951410.50            5
2010
1     117.11      435466.90   328376.44     248406.36            8
2012
2     524.96      247956.32   933903.84     224598.75            5
2014
3       6.92       75591.66    56065.84      19525.82            6
2014
4     524.96      471336.91  2657347.52     639077.50            2
2013

df.cov()

                     Item Type  Sales Channel  Order Priority    Units
Sold  \
Item Type           10.154646       0.075758        0.796869 -
2.119174e+03
Sales Channel        0.075758       0.252525        0.085859 -
2.055202e+02
Order Priority       0.796869       0.085859        1.120303 -
2.167714e+02
Units Sold       -2119.174040    -205.520202     -216.771414
7.809144e+06
Unit Price         155.090233     -17.151162       44.692413 -
4.640481e+04
Unit Cost          161.485010     -13.017677       38.830141 -
4.850918e+04
Total Revenue  -462240.691114 -142763.883283    52903.698730
3.274260e+08
Total Cost      454217.176179  -76604.015909   169300.764165
1.135124e+09
Total Profit   -181491.066687  -28652.802727    27176.072242
6.918495e+08
Order Month          0.359798      -0.080808       -0.325455 -
7.492384e+01
Order Year          -0.121111       0.005051       -0.149596
7.268354e+01

                   Unit Price     Unit Cost  Total Revenue    Total
Cost  \
Item Type        1.550902e+02  1.614850e+02  -4.622407e+05
4.542172e+05
Sales Channel  -1.715116e+01 -1.301768e+01  -1.427639e+05 -
```

```
7.660402e+04
Order Priority  4.469241e+01  3.883014e+01   5.290370e+04
1.693008e+05
Units Sold     -4.640481e+04 -4.850918e+04   3.274260e+08
1.135124e+09
Unit Price      5.550370e+04  4.377593e+04  -1.374388e+07
2.012054e+08
Unit Cost       4.377593e+04  3.542232e+04  -1.269463e+07
1.580833e+08
Total Revenue  -1.374388e+07 -1.269463e+07   2.131684e+12 -
3.408793e+10
Total Cost      2.012054e+08  1.580833e+08  -3.408793e+10
1.174922e+12
Total Profit    5.758482e+07  3.856216e+07   1.910834e+10
3.822231e+11
Order Month    -2.521499e+01 -2.651735e+01   1.592179e+05 -
5.676315e+04
Order Year     -3.039949e+01 -2.812741e+01  -4.176989e+05 -
1.152107e+05

                 Total Profit    Order Month     Order Year
Item Type      -1.814911e+05       0.359798      -0.121111
Sales Channel  -2.865280e+04      -0.080808       0.005051
Order Priority  2.717607e+04      -0.325455      -0.149596
Units Sold      6.918495e+08     -74.923838      72.683535
Unit Price      5.758482e+07     -25.214988     -30.399494
Unit Cost       3.856216e+07     -26.517354     -28.127414
Total Revenue   1.910834e+10  159217.856863 -417698.919508
Total Cost      3.822231e+11  -56763.151481 -115210.724134
Total Profit    1.923155e+11   75537.477333    2010.645333
Order Month     7.553748e+04      11.244848      -0.747273
Order Year      2.010645e+03      -0.747273       4.360707

sns.pairplot(df)

<seaborn.axisgrid.PairGrid at 0x253f9594150>
```




```
sns.heatmap(df.corr(),annot=True)
```

```
<Axes: >
```




```
#prediction...

X = df[['Item Type', 'Sales Channel', 'Order Priority', 'Units Sold',
'Unit Price', 'Unit Cost', 'Total Revenue', 'Total Cost', 'Order
Month', 'Order Year']]
y = df['Total Profit']

# Split the data into training and testing sets
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.3, random_state=42)

# Standardizing the dataset
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()

# Performing fit transform on X_train dataframe
X_train = scaler.fit_transform(X_train)

# Performing transform on X_test dataframe
X_test = scaler.transform(X_test)
```

```
# Applying Linear Regression on X_train and y_train
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import cross_val_score
regression = LinearRegression()
regression.fit(X_train,y_train)
```

```
LinearRegression()
```

```
# Calculating mean squared error
mse =
cross_val_score(regression,X_train,y_train,scoring="neg_mean_squared_e
rror",cv=5)

np.mean(mse)
```

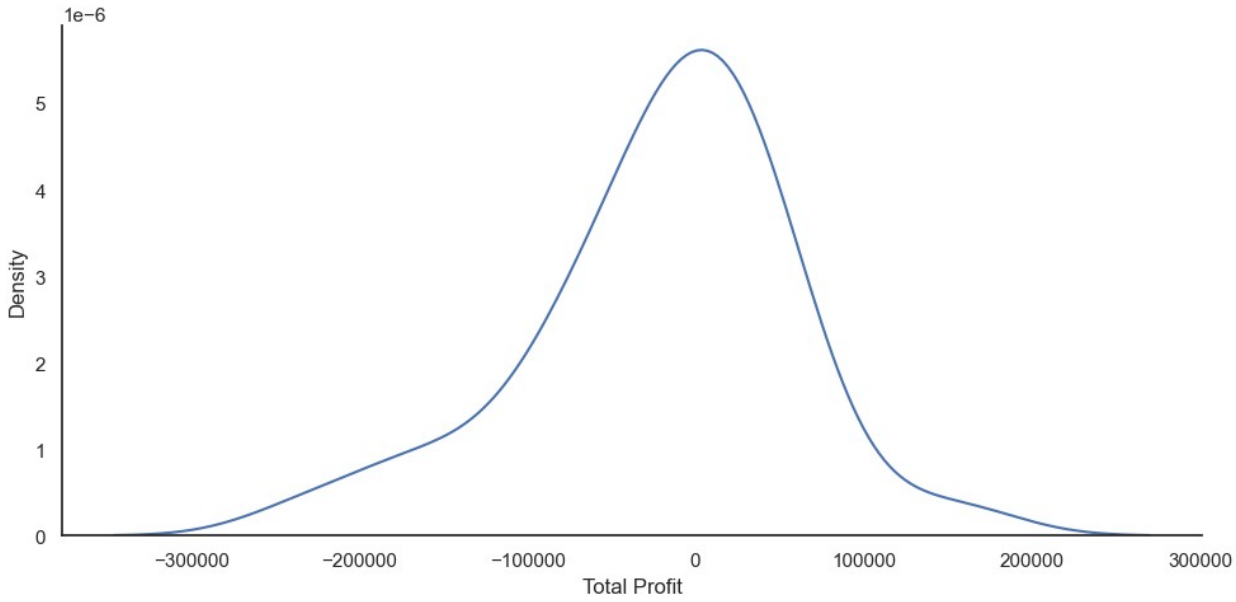
```
-15255186176.670282
```

```
## prediction
reg_pred = regression.predict(X_test)

reg_pred
```

```
array([ 208192.60250249,  392508.92543095,  667219.435758  ,
        110391.9834907 ,  457740.66219825,  469772.04977741,
        102412.71409508,  800928.99608871,    4847.23276108,
        871802.56757238,  686704.66185942, 1339282.25483319,
        732219.80482019, 1388215.51108883,  -43607.75256582,
        627513.46052791,  299412.45257405,  238954.5985345 ,
        536668.70015791,  -56077.18844879,  293247.90484298,
       -226891.37235381,   26408.43870545,  209550.29848811,
        471023.68852941,  191445.43257938,  -12531.17279198,
         46626.42923937,  566803.98478096,   74464.93743837])
```

```
# Creating kernel density estimate plot
import seaborn as sns
sns.displot(reg_pred - y_test,kind='kde', height=5, aspect=2)
```

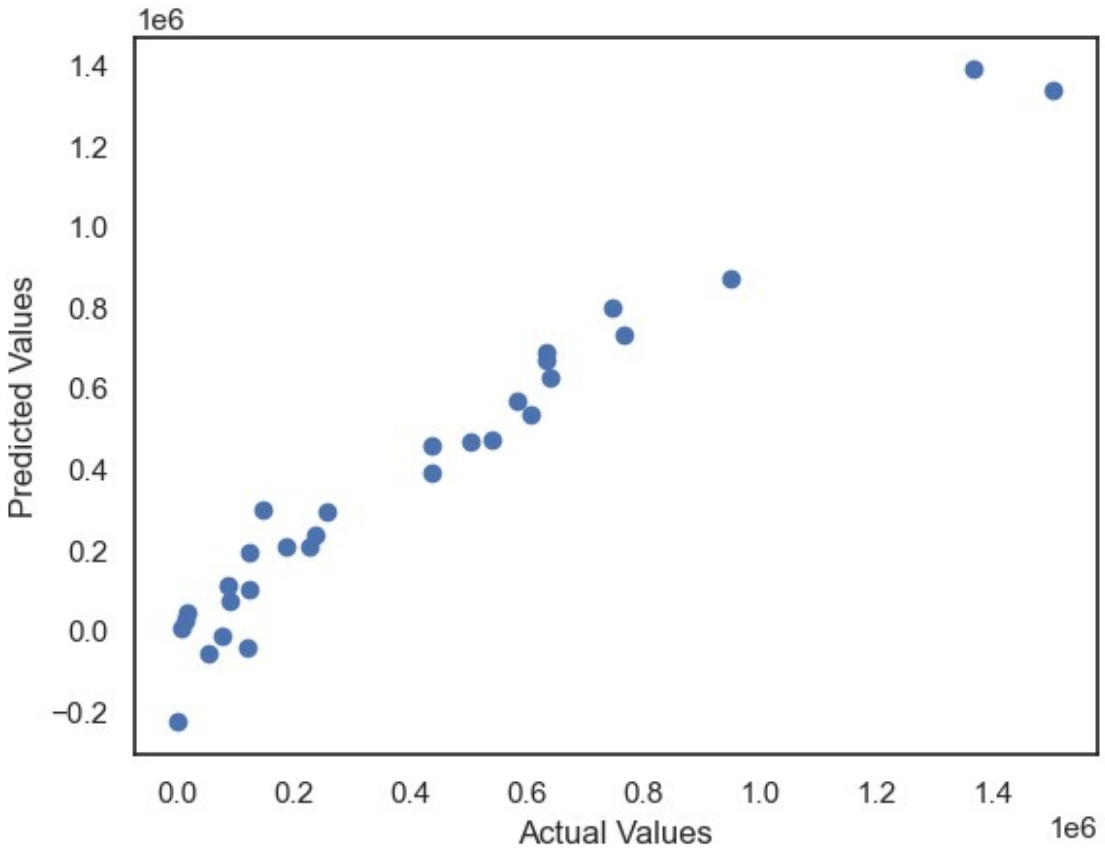
```
<seaborn.axisgrid.FacetGrid at 0x253ede52f50>
```




```
# Finding Accuracy percentage on the bases of r2 score
from sklearn.metrics import r2_score
score = r2_score(reg_pred,y_test)

# Calculate the percentage of accuracy
accuracy_pct = score * 100

print("Accuracy: {:.2f}%".format(accuracy_pct))
```

```
Accuracy: 95.79%
```

```
# Plot the predicted values against the actual values to visualize how
well the model is fitting the data.
import matplotlib.pyplot as plt

plt.scatter(y_test, reg_pred)
plt.xlabel('Actual Values')
plt.ylabel('Predicted Values')
plt.show()
```




```
# Calculate the mean squared error (MSE) or root mean squared error
(RMSE) to quantify the model's performance.

from sklearn.metrics import mean_squared_error
import numpy as np

mse = mean_squared_error(y_test, reg_pred)
rmse = np.sqrt(mse)
print('MSE:', mse)
print('RMSE:', rmse)
```

```
MSE: 6235319368.743605
RMSE: 78964.03845259945
```

```
import plotly.graph_objs as go

reg = LinearRegression()

# Fit the model to the training data
reg.fit(X_train, y_train)

# Make predictions on the test set
y_pred = reg.predict(X_test)
```

```
# Calculate the R-squared score
r2 = reg.score(X_test, y_test)

# Create a copy of the X_train numpy array with modified column names
X_train_df = pd.DataFrame(X_train, columns=['Item Type', 'Sales
Channel', 'Order Priority', 'Units Sold', 'Unit Price', 'Unit Cost',
'Total Revenue', 'Total Cost', 'Order Month', 'Order Year'])

coef = reg.coef_

# Create a DataFrame with the coefficients and feature names
feature_importances = pd.DataFrame({'Feature': X_train_df.columns,
'Importance': coef})

# Sort the DataFrame by importance
feature_importances = feature_importances.sort_values('Importance',
ascending=False)

# Create a bar chart of the feature importances using Plotly
fig = go.Figure()
fig.add_trace(go.Bar(x=feature_importances['Feature'],
y=feature_importances['Importance']))
fig.update_layout(title='Feature Importance (R-squared =
{:.2f})'.format(r2),
                  xaxis_title='Feature',
                  yaxis_title='Importance',
                  xaxis_tickangle=-45)
fig.show()
```