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

df = pd.read_csv("C:\\Users\\Sarthak Tyagi\\Downloads\\Amazon Sales
data (1).csv")

df
```

```
                                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 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
..           ...            ...         ...        ...         ...
95        Online              M   7/26/2011  512878119    9/3/2011
96       Offline              L  11/11/2011  810711038  12/28/2011
97       Offline              C    6/1/2016  728815257   6/29/2016
98       Offline              M   7/30/2015  559427106    8/8/2015
99       Offline              L   2/10/2012  665095412   2/15/2012

    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
..         ...         ...        ...            ...         ...
...
95         888      109.28      35.84       97040.64    31825.92
65214.72
96        6267        9.33       6.92       58471.11    43367.64
15103.47
97        1485      154.06      90.93      228779.10   135031.05
93748.05
98        5767       81.73      56.67      471336.91   326815.89
144521.02
99        5367      668.27     502.54     3586605.09  2697132.18
889472.91

[100 rows x 14 columns]
```

```
#checking for null values
df.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
```

```
#getting basic info of dataset
df.describe()
df.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
```

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

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

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

   Unit Price  Unit Cost  Total Revenue  Total Cost  Total Profit
0      255.28     159.42     2533654.00  1582243.50     951410.50
```

```
1       205.70     117.11       576782.80   328376.44     248406.36
2       651.21     524.96      1158502.59   933903.84     224598.75
3         9.33       6.92        75591.66    56065.84      19525.82
4       651.21     524.96      3296425.02  2657347.52     639077.50
```

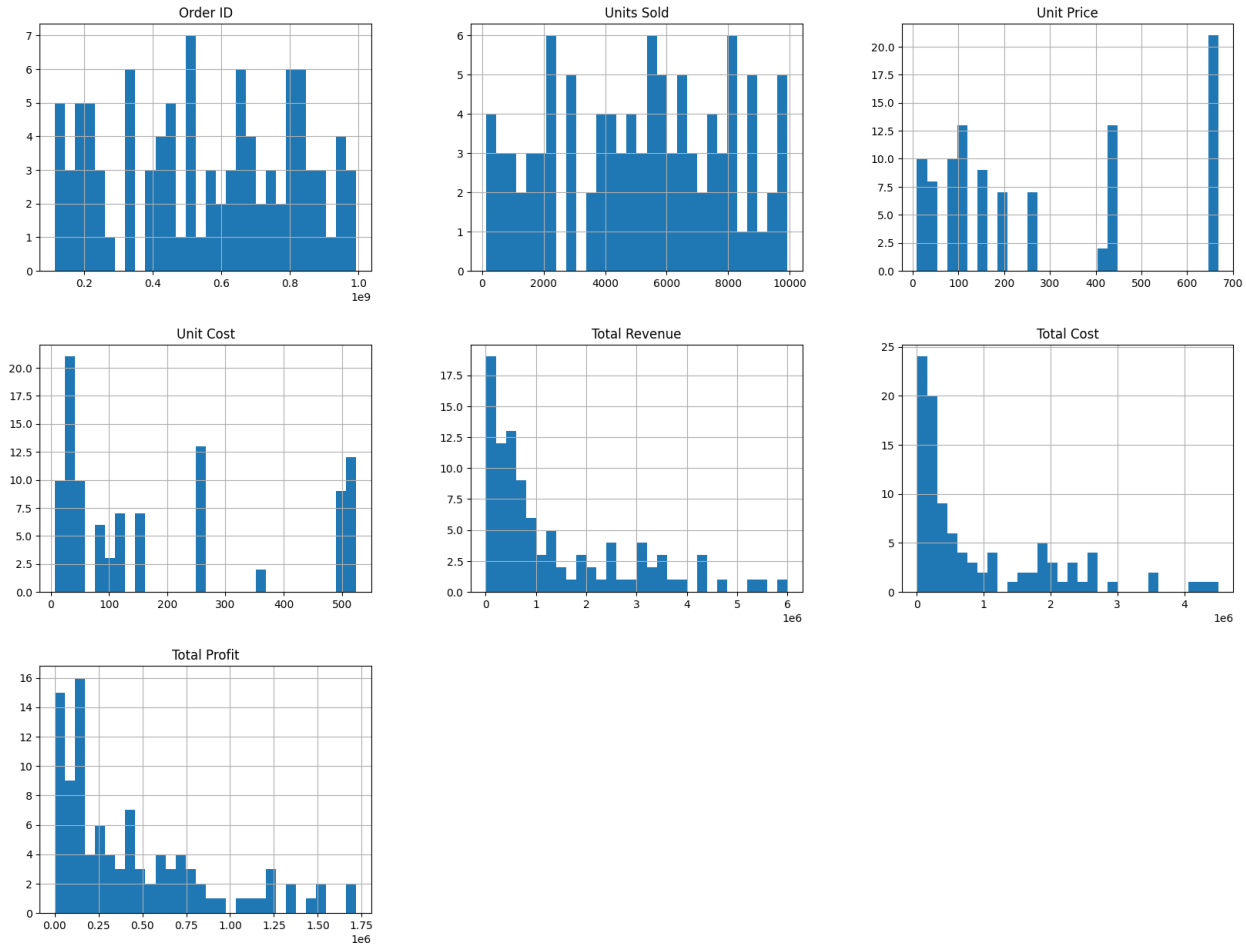
```
df.columns
```

```
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')
```

```
print(f"the number of rows are:{df.shape[0]}, and the number of
columns are: {df.shape[1]}")
```

```
the number of rows are:100, and the number of columns are: 14
```

```
df.hist(bins=30, figsize=(20, 15))
plt.show()
```



```
print(df['Total Revenue'].describe())
```

```
count    1.000000e+02
mean     1.373488e+06
std      1.460029e+06
min      4.870260e+03
25%      2.687212e+05
50%      7.523144e+05
75%      2.212045e+06
max      5.997055e+06
Name: Total Revenue, dtype: float64
```

```
#convert date and extract month and year from 'date'
df['date']=pd.to_datetime(df['Order Date'])
df['month']=df['date'].dt.month
df['year']=df['date'].dt.year

#groupby month and sum the profits
monthly_sales = df.groupby(['year', 'month'])['Total
Profit'].sum().reset_index()

#monthly profits
print(monthly_sales)
```

	year	month	Total Profit
0	2010	2	1424410.94
1	2010	5	965441.52
2	2010	6	727423.20
3	2010	10	1495392.79
4	2010	11	1375311.70
5	2010	12	641587.28
6	2011	1	363509.04
7	2011	2	127722.96
8	2011	4	693911.51
9	2011	5	89904.06
10	2011	6	7828.12
11	2011	7	65214.72
12	2011	9	235601.16
13	2011	11	1157316.66
14	2012	1	206568.36
15	2012	2	1553766.98
16	2012	3	407630.41
17	2012	4	971008.14
18	2012	5	1218518.14
19	2012	6	698414.36
20	2012	7	1065073.62
21	2012	8	248406.36
22	2012	9	2084889.03
23	2012	10	758734.72
24	2013	2	639077.50

```
25  2013      3     359941.17
26  2013      4     632512.50
27  2013      6     515753.38
28  2013      7    3398463.02
29  2013      8      23150.46
30  2013      9      18405.17
31  2013     10    1074864.34
32  2013     12      53252.50
33  2014      2     655704.80
34  2014      4    1838545.92
35  2014      5     938755.75
36  2014      6      19525.82
37  2014      7     227273.58
38  2014      8     306097.92
39  2014      9       5270.67
40  2014     10     735800.80
41  2014     11    1152486.42
42  2015      1    1367272.50
43  2015      2     780095.53
44  2015      4     624230.28
45  2015      7     418665.00
46  2015      8       1621.93
47  2015     10     369155.00
48  2015     11     435499.20
49  2016      3      85223.58
50  2016      5     127054.20
51  2016      6     216434.55
52  2016      7     403773.12
53  2016     10      72975.60
54  2016     11    2336986.67
55  2016     12    1661390.29
56  2017      1     879507.12
57  2017      2    1891271.80
58  2017      3      75555.90
59  2017      5    1243018.63
```

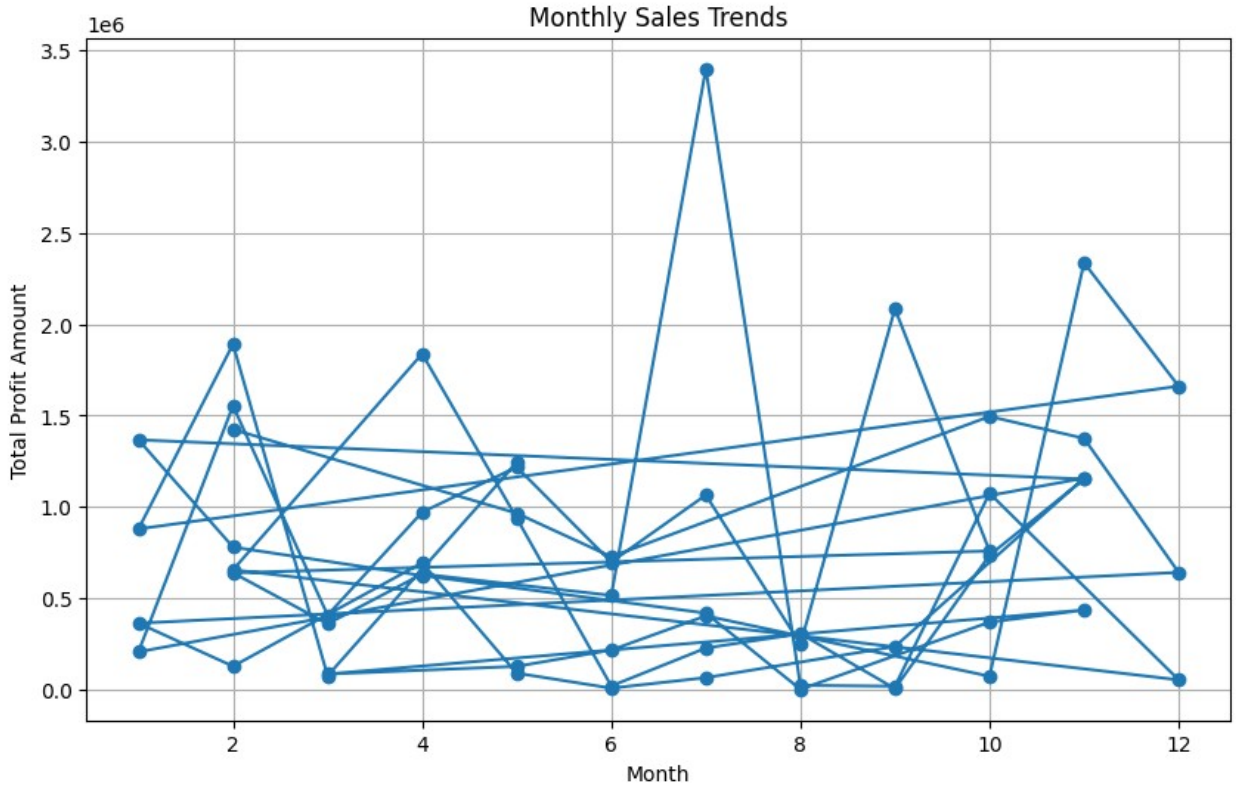
```
max_sales_month = monthly_sales.loc[monthly_sales['Total
Profit'].idxmax()]

#maximum sales month
print(max_sales_month)
```

```
year               2013.00
month                 7.00
Total Profit    3398463.02
Name: 28, dtype: float64
```

```
plt.figure(figsize=(10, 6))
plt.plot(monthly_sales['month'], monthly_sales['Total Profit'],
marker='o')
```

```
plt.title('Monthly Sales Trends')
plt.xlabel('Month')
plt.ylabel('Total Profit Amount')
plt.grid(True)
plt.show()
```




```
#knownig top products
top_products = df.groupby('Item Type')['Units
Sold'].sum().reset_index()

# Sort the products based on total revenue
top_products = top_products.sort_values(by='Units Sold',
ascending=False)

# Select the top 10 products
top_10_products = top_products.head(10)

# Print the results
print(top_10_products)
```

	Item Type	Units Sold
4	Cosmetics	83718
3	Clothes	71260
1	Beverages	56708
5	Fruits	49998

```
9     Personal Care       48708
8   Office Supplies       46967
6         Household       44727
0         Baby Food       40545
2            Cereal       25877
11       Vegetables       20051
```

```
df
```

```
                                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 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
..           ...            ...         ...        ...         ...
95        Online              M   7/26/2011  512878119    9/3/2011
96       Offline              L  11/11/2011  810711038  12/28/2011
97       Offline              C    6/1/2016  728815257   6/29/2016
98       Offline              M   7/30/2015  559427106    8/8/2015
99       Offline              L   2/10/2012  665095412   2/15/2012

    Units Sold  Unit Price  Unit Cost  Total Revenue  Total Cost  \
0         9925      255.28     159.42     2533654.00  1582243.50
1         2804      205.70     117.11      576782.80   328376.44
```

```
2          1779      651.21     524.96     1158502.59   933903.84
3          8102        9.33       6.92       75591.66    56065.84
4          5062      651.21     524.96     3296425.02  2657347.52
..          ...         ...        ...            ...         ...
95          888      109.28      35.84       97040.64    31825.92
96         6267        9.33       6.92       58471.11    43367.64
97         1485      154.06      90.93      228779.10   135031.05
98         5767       81.73      56.67      471336.91   326815.89
99         5367      668.27     502.54     3586605.09  2697132.18

    Total Profit       date  month  year
0      951410.50 2010-05-28      5  2010
1      248406.36 2012-08-22      8  2012
2      224598.75 2014-05-02      5  2014
3       19525.82 2014-06-20      6  2014
4      639077.50 2013-02-01      2  2013
..           ...        ...    ...   ...
95      65214.72 2011-07-26      7  2011
96      15103.47 2011-11-11     11  2011
97      93748.05 2016-06-01      6  2016
98     144521.02 2015-07-30      7  2015
99     889472.91 2012-02-10      2  2012

[100 rows x 17 columns]
```