

Analyzing Amazon Sales data

Neenu E S

PROJECT DETAILS

Project Title	Analyzing Amazon Sales data
Technologies	Data Science
Domain	E-commerce
Project Difficulty level	Advanced
Programming Language Used	Python
Tools Used	Jupyter Notebook, MS-Excel

Objective

- Extract-Transform-Load some Amazon dataset
- Find Sales-trend
 - Month-wise
 - Year-wise
 - Yearly_month-wise

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.

DATA TRANSFORMATION AND EDA

```
# Read in the input file as a Pandas DataFrame
data = pd.read_csv('Amazon Sales data.csv')
# data.head()
```

```
###Dimensions of the dataset is 100 rows and 14 columns
data.shape
```

```
(100, 14)
```

```
df = data
```

```
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
```



```
# check and drop the duplicated rows
data = data.drop_duplicates()
data.shape
```

```
(100, 14)
```

There are no duplicated rows in the data.

Checking for Missing Data

```
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
```

There is no null value present in the data.

```
## To check month wise, year wise and year_month-wise sales
# We need more column is create Month, year and year_month

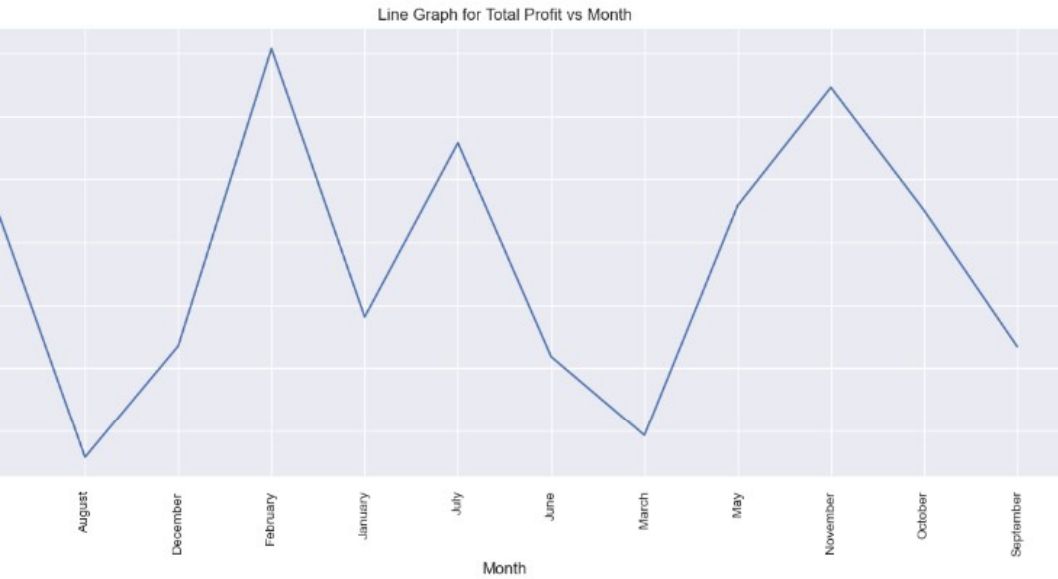
data['Year'] = pd.DatetimeIndex(data['Order Date']).year
data['Month'] = pd.DatetimeIndex(data['Order Date']).month
data['Year_month'] = pd.to_datetime(data['Order Date']).dt.strftime('%Y-%m')
# print(yearly_month_index)
```

```
data.head(3)
```

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	Year	Month	Year_month
Australia and Oceania	Tuvalu	Baby Food	Offline	High	5/28/2010	669165933	6/27/2010	9925	255.28	159.42	2533654.00	1582243.50	951410.50	2010	May	2010-05
Central America and the Caribbean	Grenada	Cereal	Online	Critical	8/22/2012	963881480	9/15/2012	2804	205.70	117.11	576782.80	328376.44	248406.36	2012	August	2012-08
Europe	Russia	Office Supplies	Offline	Low	5/2/2014	341417157	5/8/2014	1779	651.21	524.96	1158502.59	933903.84	224598.75	2014	May	2014-05

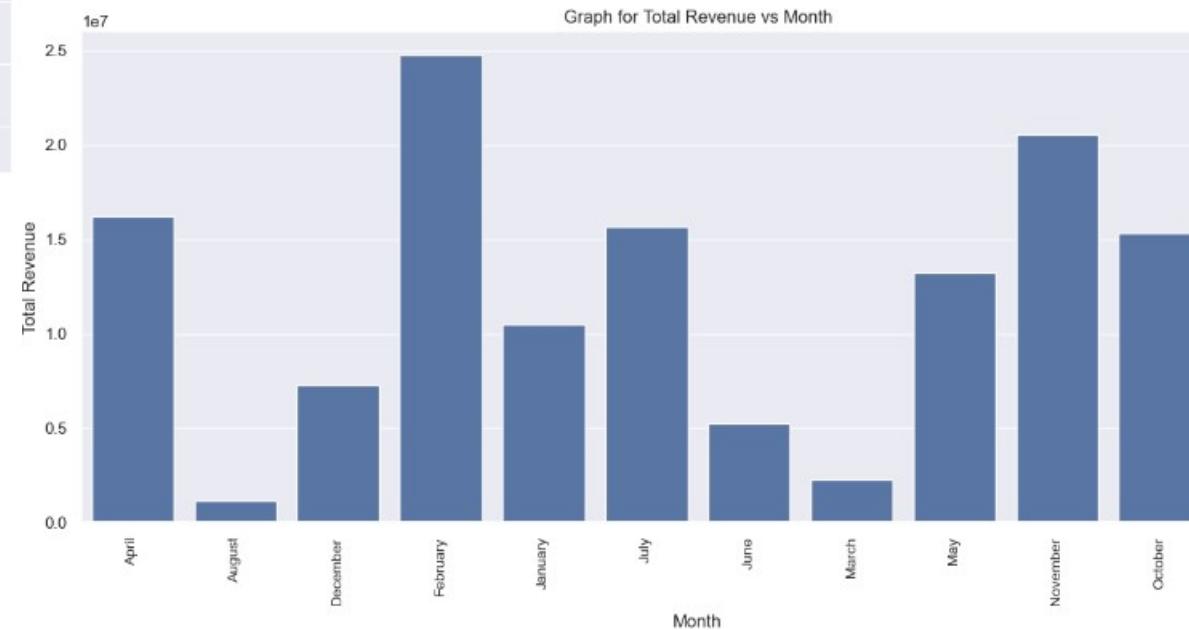
Visual Analytics And Key Insights

❖ Month - Sales



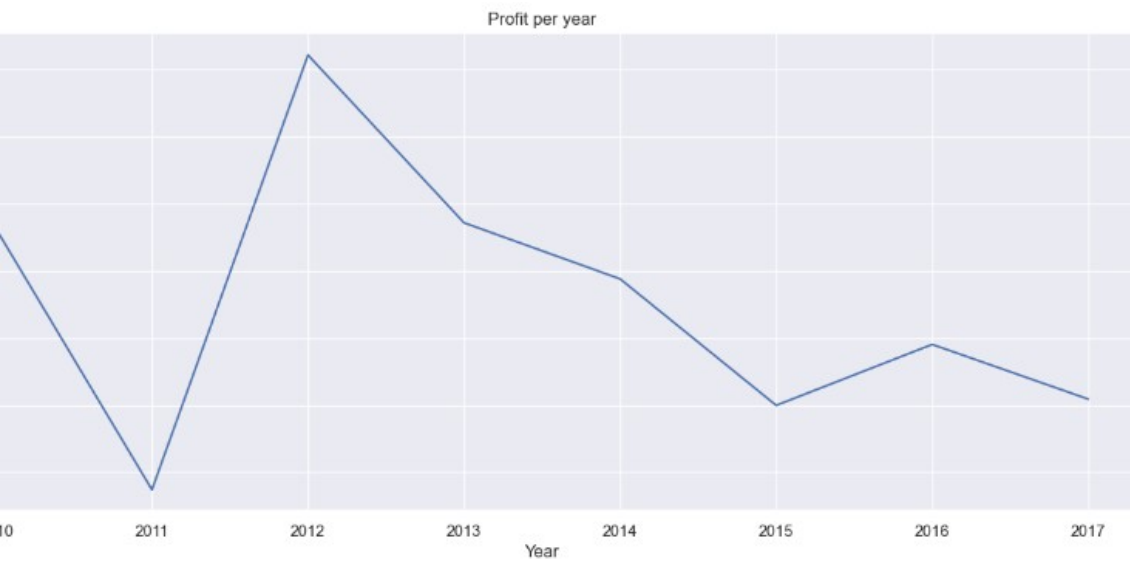
low in the month February and second in November.

the **February** shows the highest revenue and profit



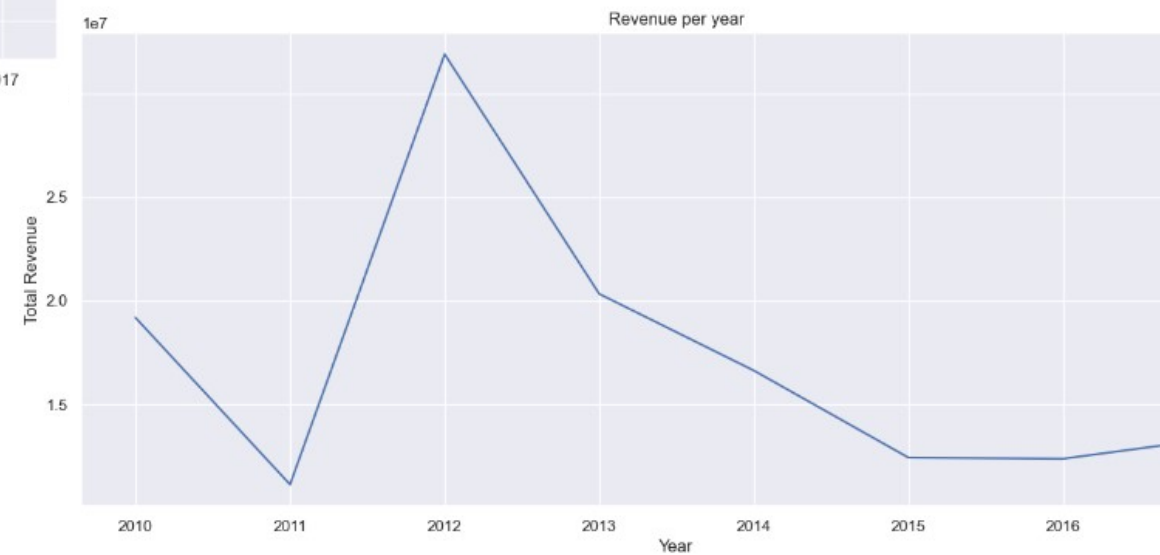
Revenue shows a greater peak in February than in November.

❖ Year - Sales



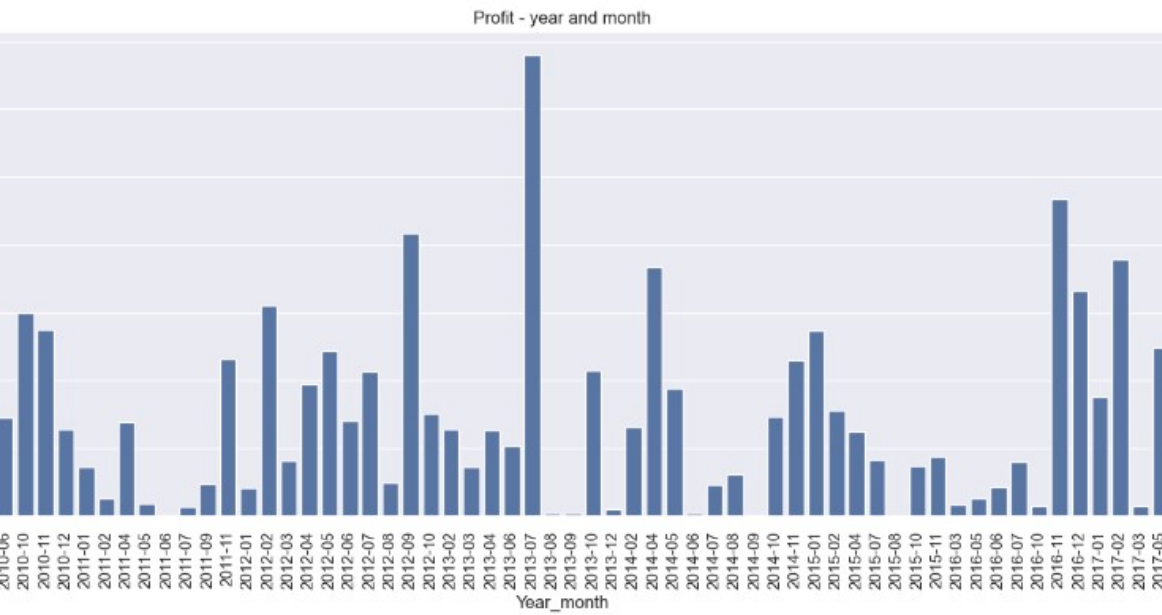
was a sharp peak in profit.

In **2012** shows the highest revenue and profit



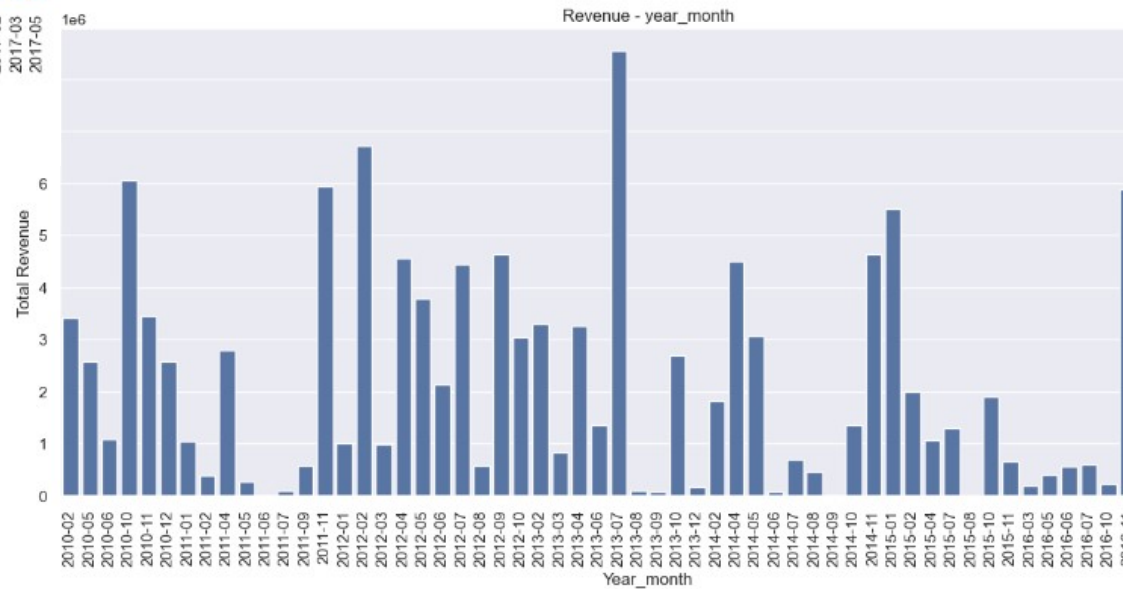
In total revenue, it also shows a sharp peak in 2012.

❖ Yearly month - Sales



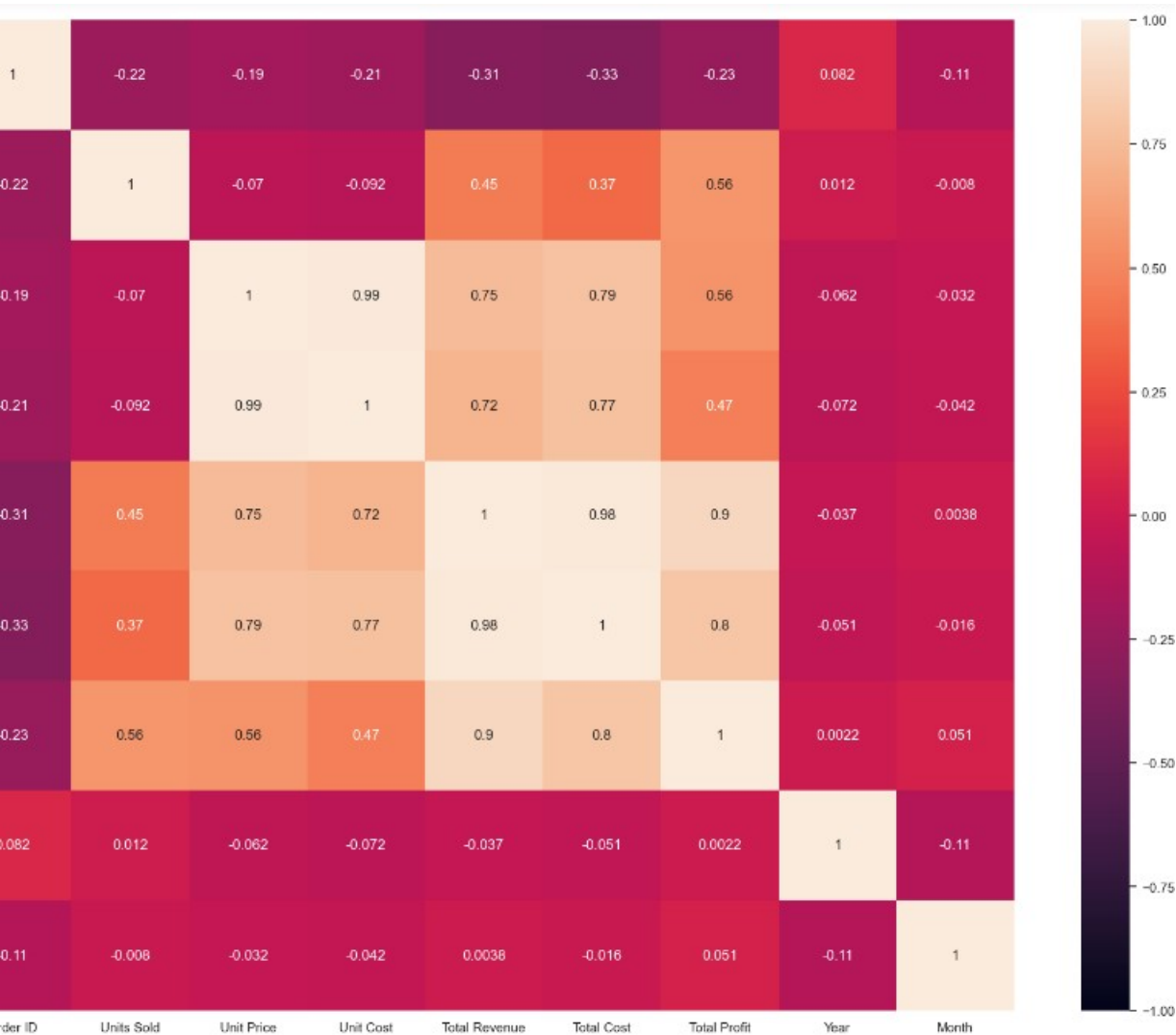
2013 – 07 shows the highest profit.

2013 – 07 shows the highest revenue and profit



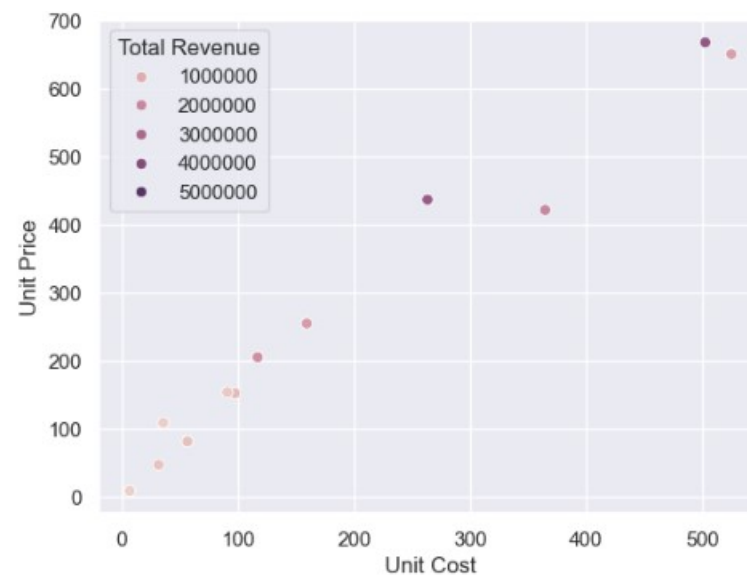
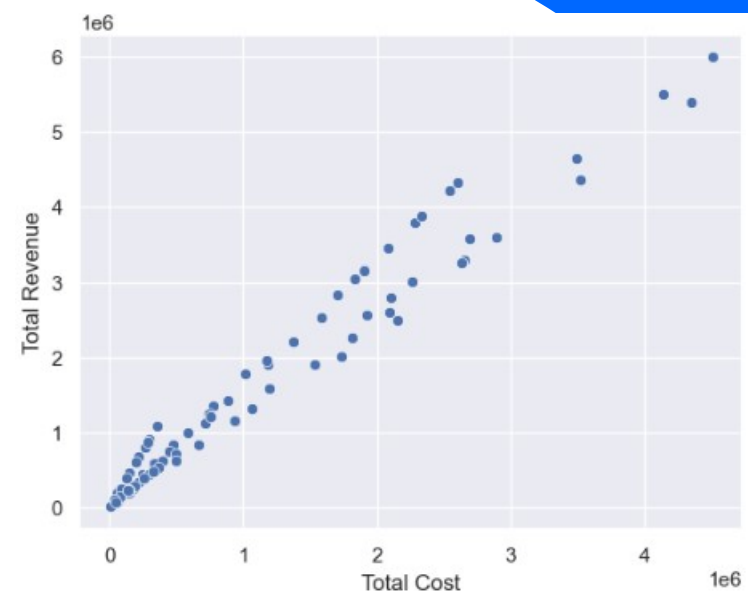
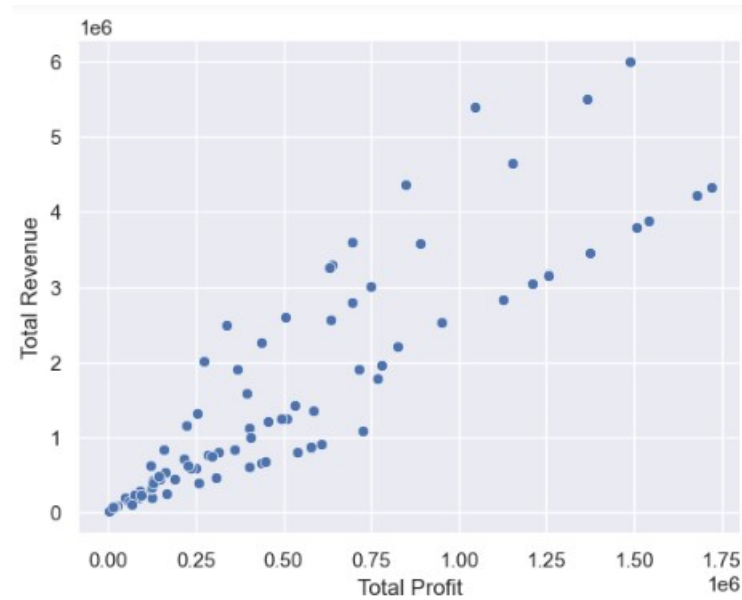
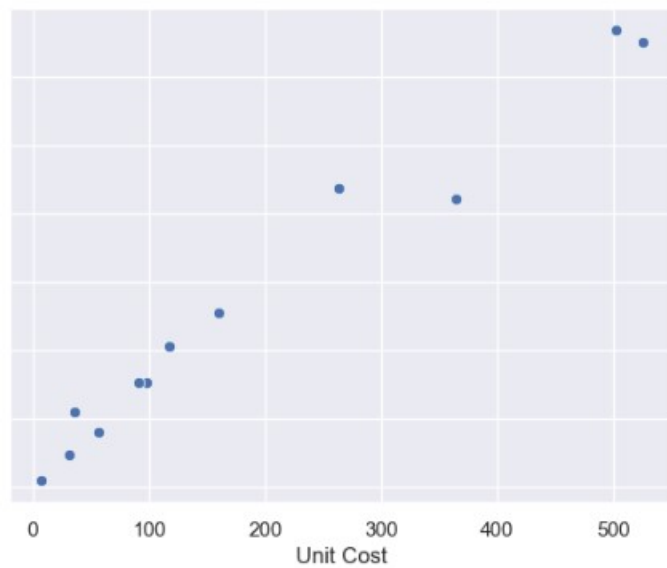
Here too, 2013–07 show the highest revenue.

Relation between the variables



Observation :-

- The relationship between total sales/revenue and total cost and p approximately 0.98 and 0.9, respectively.
- Greater profit will be made if revenue is strong, and vice versa.
- With a correlation of 0.99, unit cost and unit price are strongly co
- There is a moderate relationship between Total Profit, Total Cost Total Revenue and Units Sold, Unit Price, and Unit Cost.
- Total profit is 0.8 times connected to total cost.
- The correlation coefficient between units sold and unit cost is neg which suggests that the relationship between product quantity and inverse.
- This also applies to the case of sold units and unit price.
- The price of a product increases with the number of units offered



Conclusion

- Data analysis helps to find the basic structure of the dataset.
- Perform EDA on the datasets to explore the data and extract all possible insights, which can help in model building and better decision-making.
- In the month of **February**, it showed the highest profit and revenue. The highest number of units sold was in July, followed by February.
- In **2012**, there was a sharp peak in profit, revenue, and units sold.
- **2013–07** showed the highest profit and revenue. The highest number of units sold was in 2012–07 and 2014–10, followed by 2013–07.
- Cosmetics items show the highest profit, revenue, and number of units sold.
- Units Sold, Unit Price, and Unit Cost are related to Total Profit, Total Cost and Total Revenue.
- As the values of unit price, unit cost, and unit sold increase, the value of total profit and total revenue increase.
- The lower the number of units of a product available, the higher the price.
- The Asia region shows the highest revenue compared to other regions; next comes North America.
- Lithuania shows the most revenue among the countries, and the lowest will be New Zealand, Kyrgyzstan, Slovakia, and Kuwait.

Thank you

