

Project -1

Amazon Sales Data

**Performing a deep analysis of the Amazon sales dataset through
Python**

Sauradip Pradhan
Date - 08/07/2024



Agenda:

- **Introduction to Amazon Sales Data**
- **Data Overview**
- **Explanation of the Sales Analysis Process**
- **Exploratory Data Analysis (EDA)**
- **Visualizations and Findings**
- **Recommendations**

Introduction to Amazon Sales Data

Amazon, a global e-commerce giant, operates in numerous countries and offers a wide range of products through various sales channels. Managing and analyzing sales data is crucial for understanding market trends, optimizing operations, and driving business growth. The Amazon Sales Data includes detailed records of transactions, capturing key attributes such as region, country, item type, sales channel, order priority, order and ship dates, units sold, unit price, unit cost, total revenue, total cost, and total profit. This comprehensive dataset provides valuable insights into sales performance and helps identify opportunities for cost reduction, profit maximization, and strategic decision-making. The aim of this analysis is to explore sales trends, uncover key metrics, and highlight meaningful relationships between different attributes to support Amazon's sales management objectives.

Data Overview

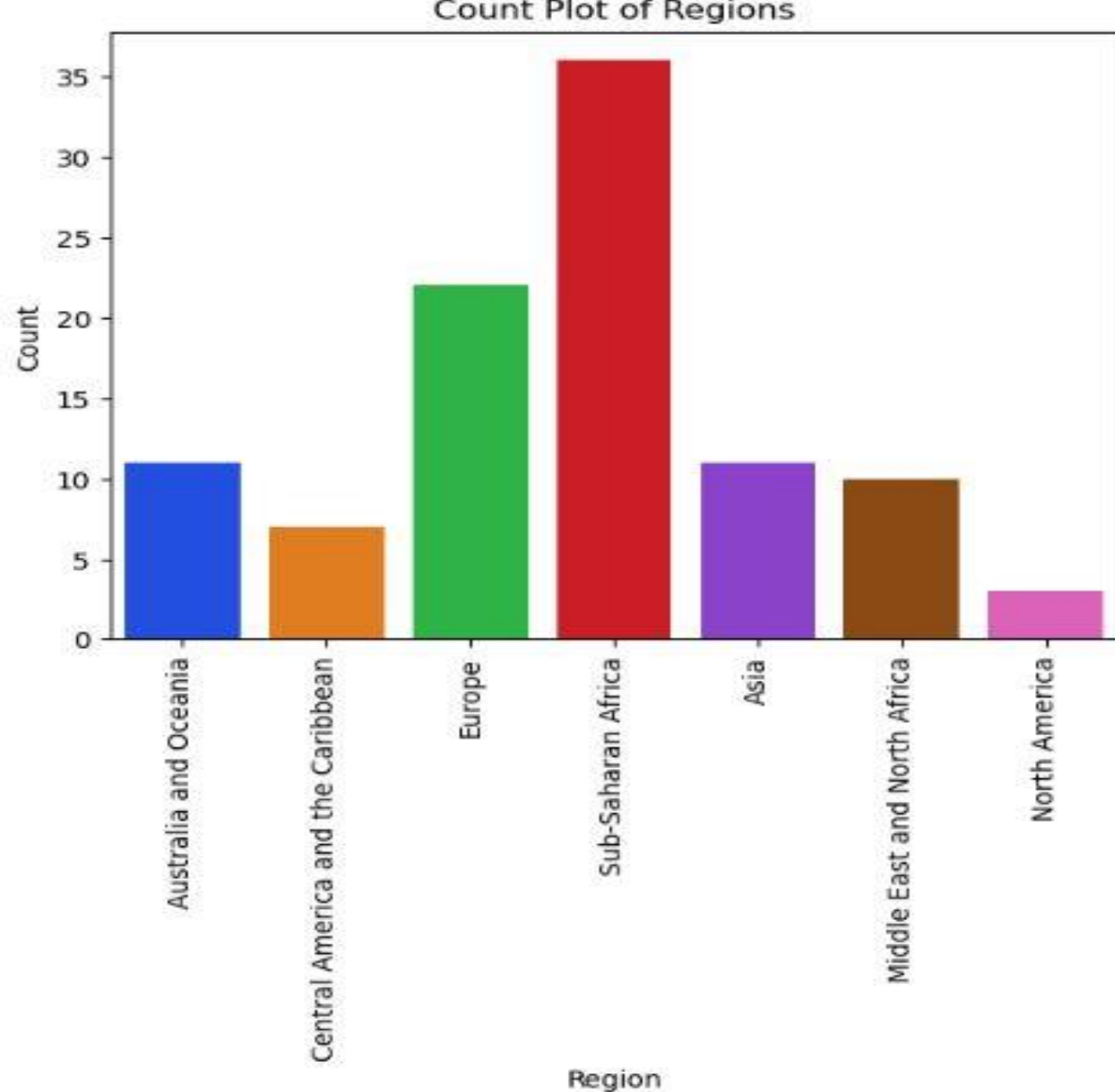
- Name of the dataset is amazon sales data.csv
- The dataset has 100 Rows and 14 columns
- There is No null values in the dataset
- 7 columns are Object data type, 2 columns are Integer data type and 5 columns are float data type

Explanation of Sales Data Analysis Process

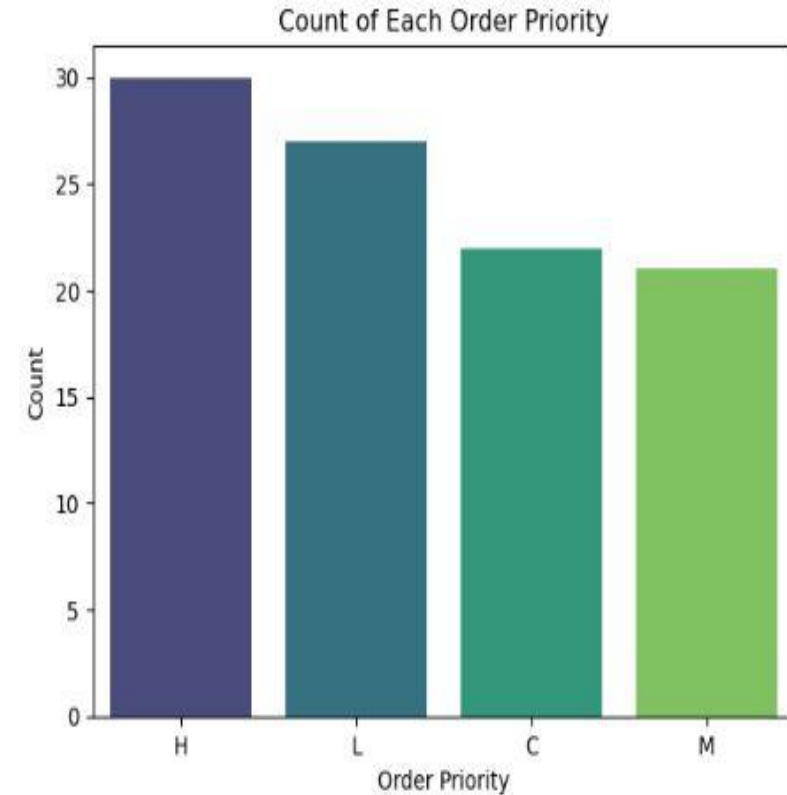
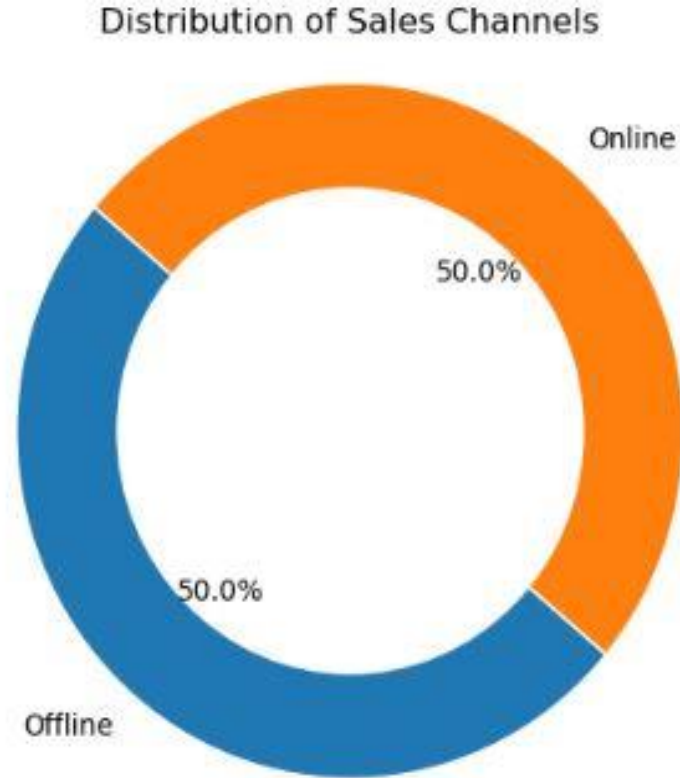
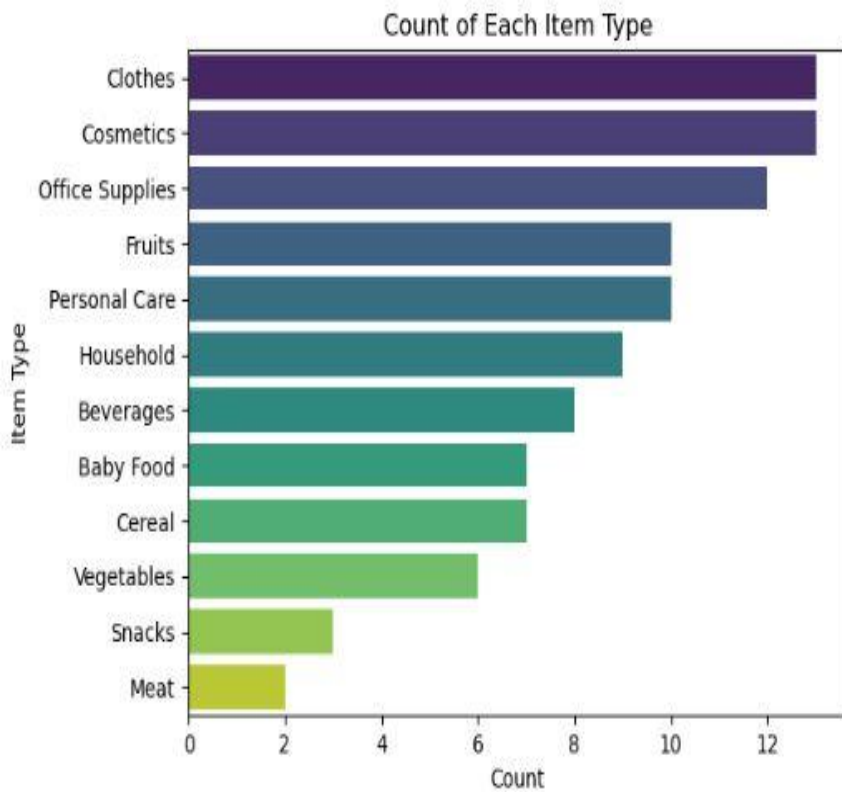
- Check the shape of the dataset, identify any null values, and verify the data types of each column.
- Drop the Order ID column.
- Convert the Order Date and Ship Date columns from object type to datetime type.
- Create a new column, Delivery Duration, by subtracting the Order Date from the Ship Date, and extract the Year, Month, and Day from the Order Date.
- Perform Univariate and Multivariate Analysis of the dataset.
- Note: All operations are conducted in Python using Jupyter Notebook.

Exploratory Data Analysis

- In EDA we create Multiple Graphs and Charts.
- To import the Dataset and for Mathematical calculation we use Pandas and numpy
- To create Graphs and Charts we use seaborn and matplotlib

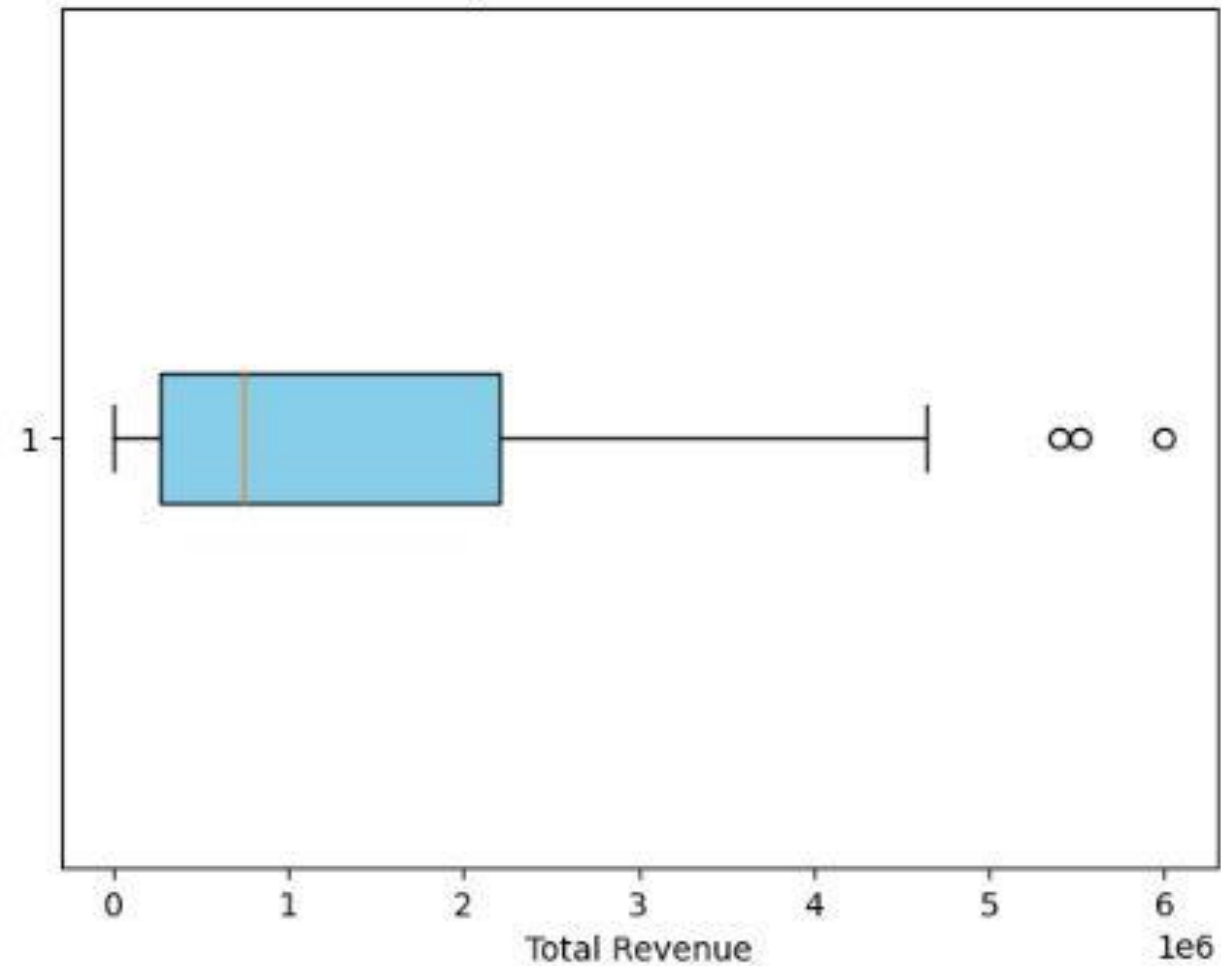


- Most of the Amazon sales occur in Sub-Saharan Africa, followed by Europe. Sales in Australia and Oceania, Asia, the Middle East, and Central America and the Caribbean are moderate. North America has the minimum sales.
- Amazon is most popular in The Gambia. São Tomé and Príncipe, Mexico, Australia, Djibouti, and Sierra Leone have moderate popularity. Bulgaria, Mali, Libya, and Azerbaijan are in the lowest group for popularity.

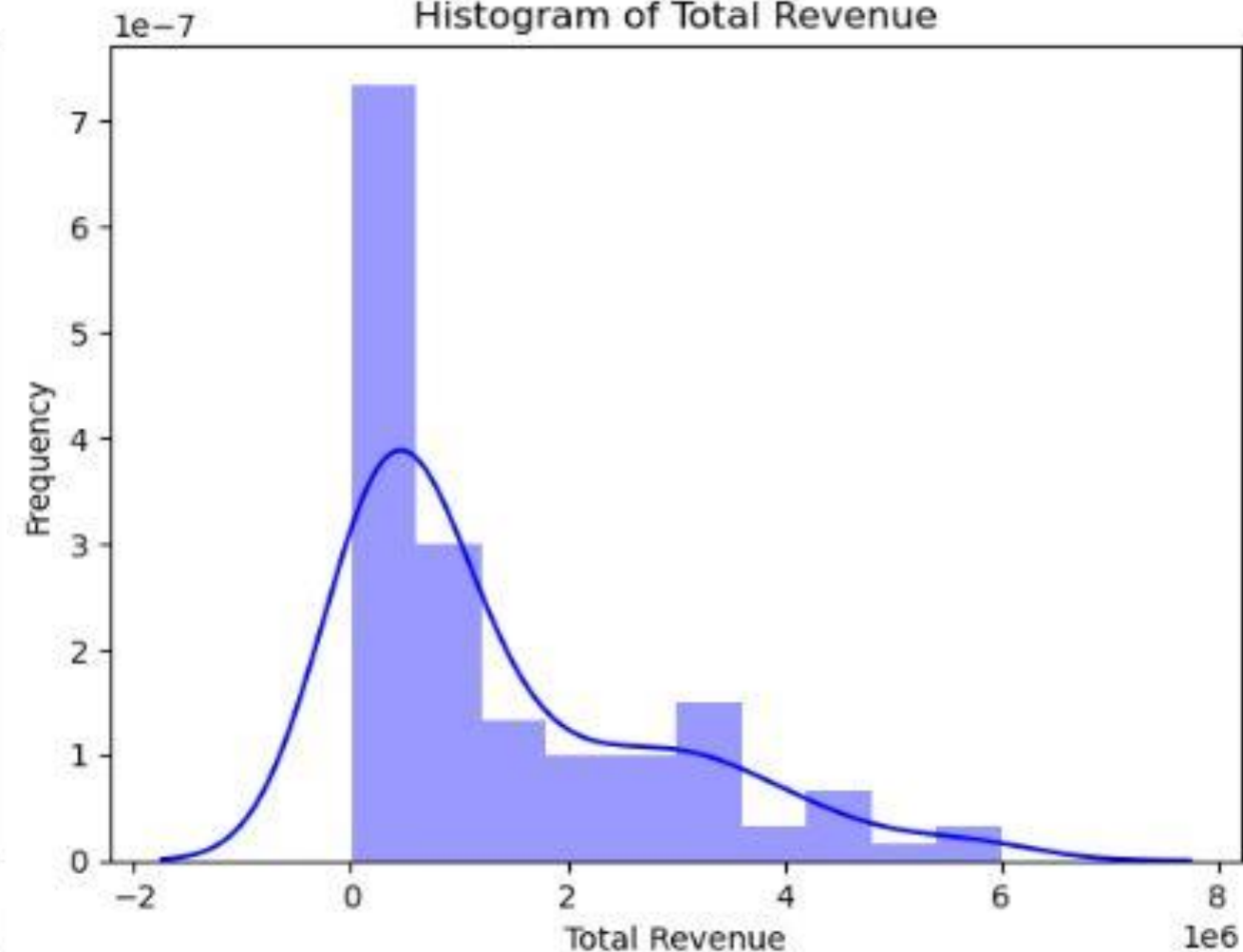


- **Clothes and cosmetics are the most popular items on Amazon. The second most popular category is office supplies. Fruits, personal care, household items, and beverages are the third highest in popularity. Following these, baby food, cereal, and vegetables come next. The least popular items are snacks and meat.**
 - **Amazon's sales channels are equally distributed between offline and online.**
 - **In this dataset, Order Priority is given as H, L, C, and M, which mean : H = High, L = Low, C = Critical, M = Moderate**
- The data shows that High Order Priority is the most frequent, followed by Low, Critical, and Moderate.**

Boxplot of Total Revenue

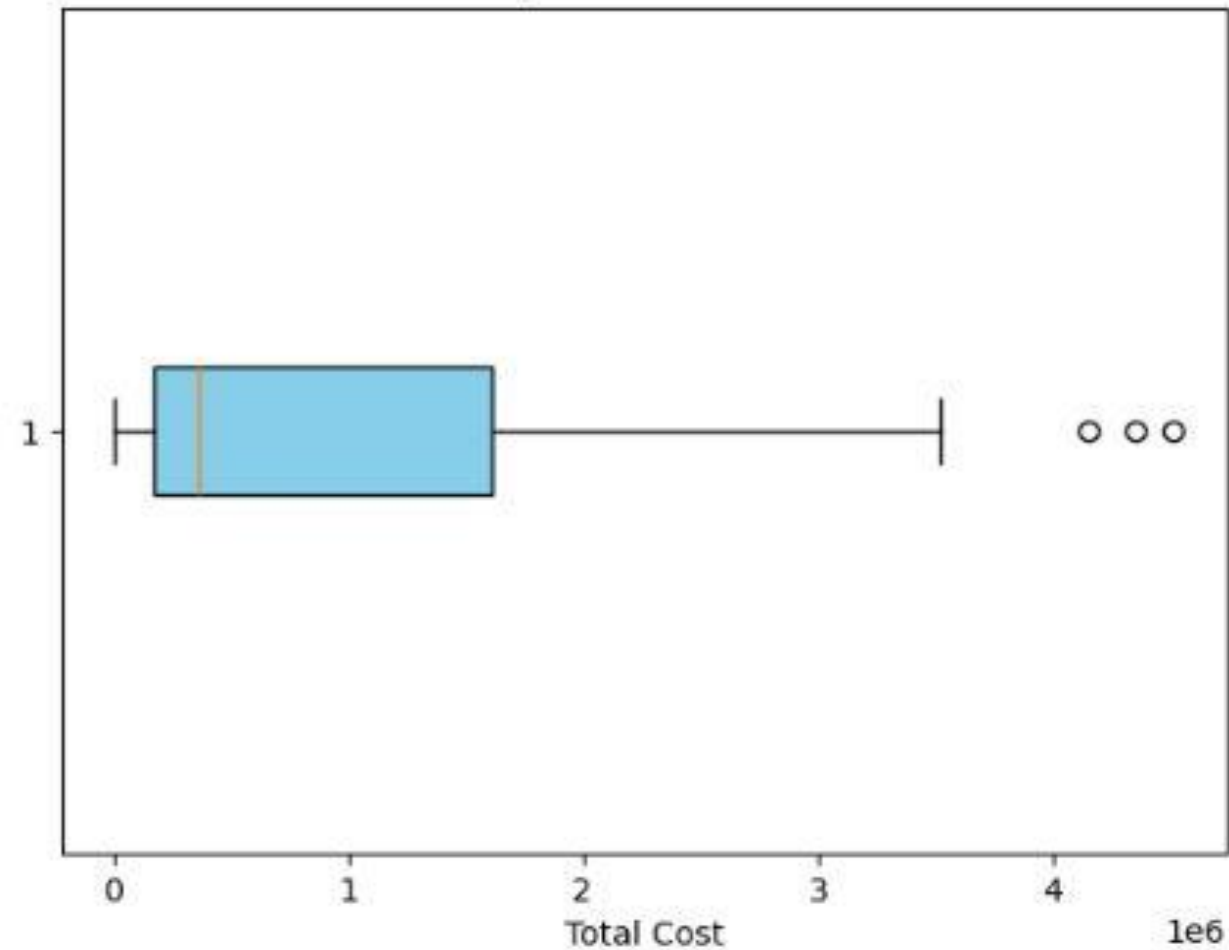


Histogram of Total Revenue

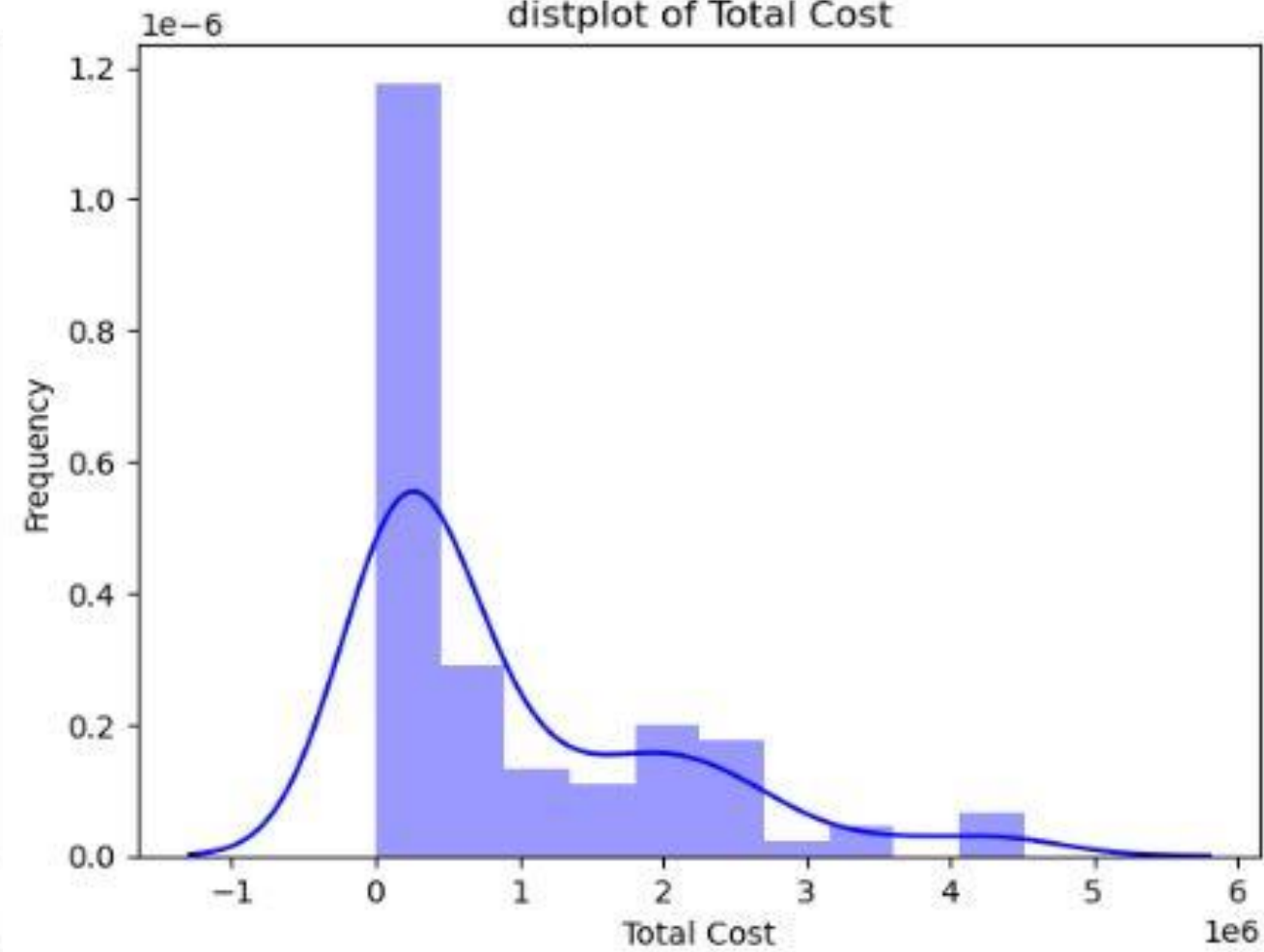


- Here we see the boxplot of Total Revenue. The median is between 0 to 1e6. There are also three outliers visible in the data.
- From the histogram, we observe that Total Revenue is at its maximum between 0 and 1e6. The frequency decreases between 1e6 and 2e6, and it becomes very minimal at 6e6. There is no increasing trend after 2e6.

Boxplot of Total Cost

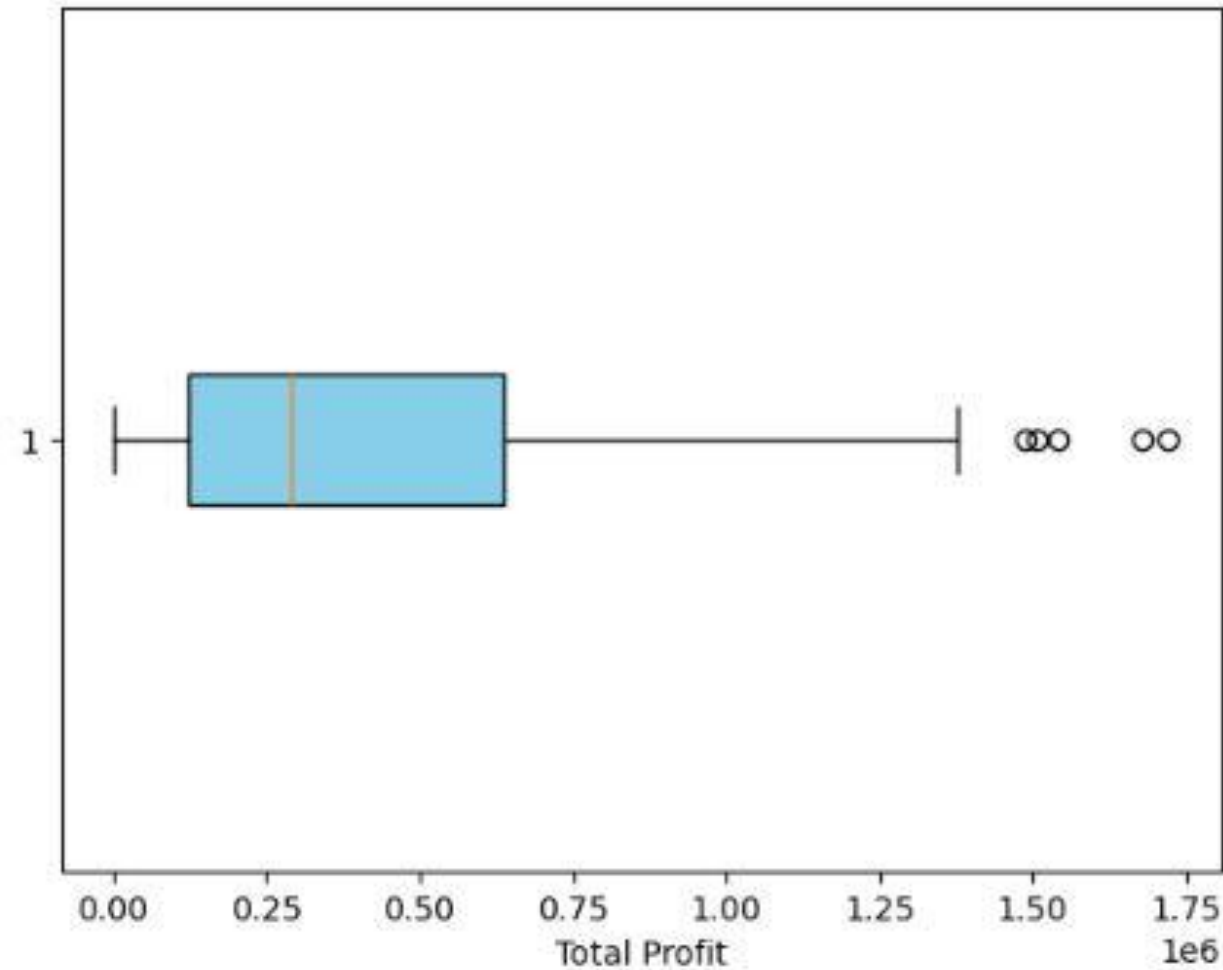


distplot of Total Cost

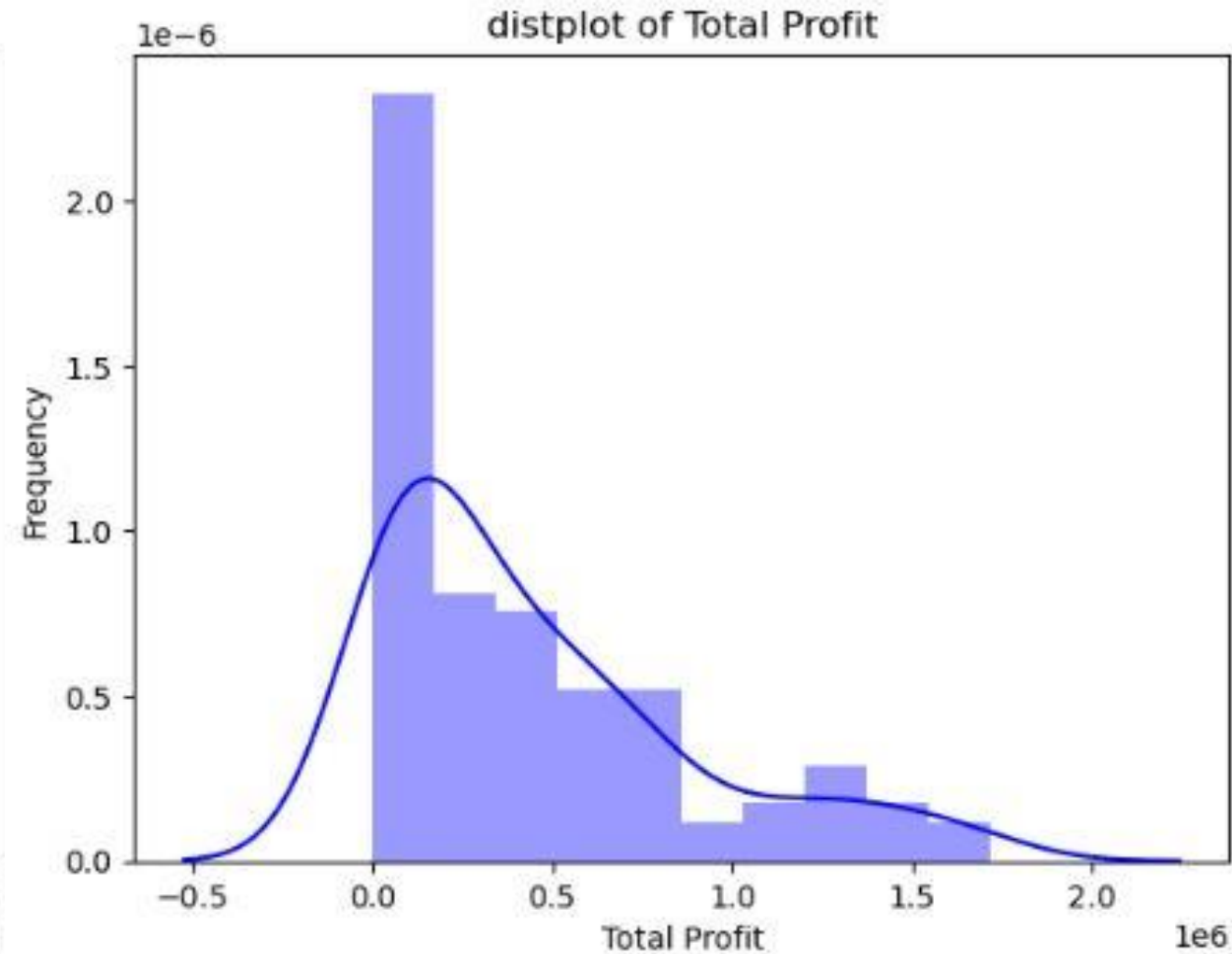


- From the box plot of Total Cost, we observe three outliers, and the median is between 0 and $1e6$.
- The maximum Cost is in the 0 to $1e6$. and then the graph start to decrease. and it is minimum at $4e6$

Boxplot of Total Profit

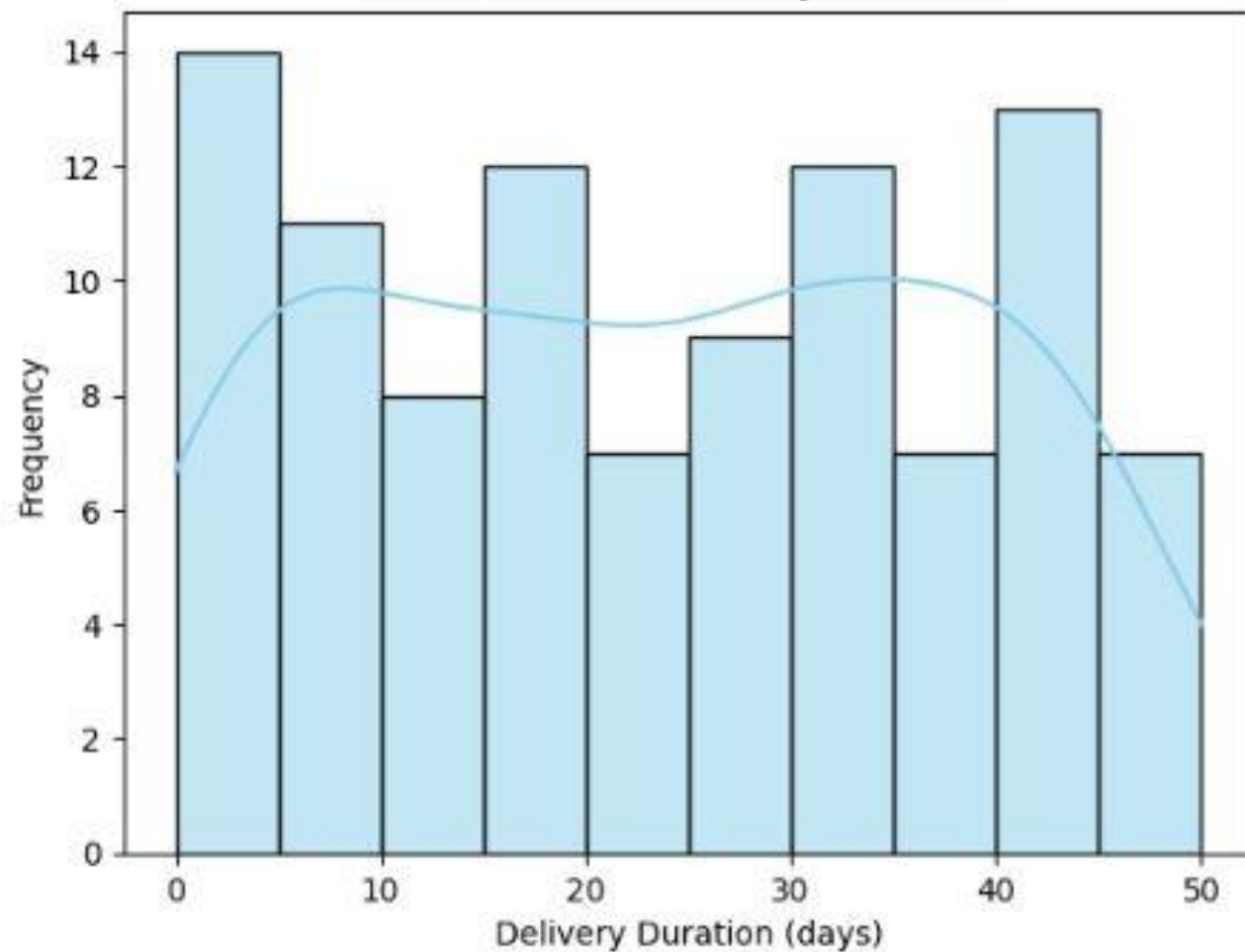


distplot of Total Profit

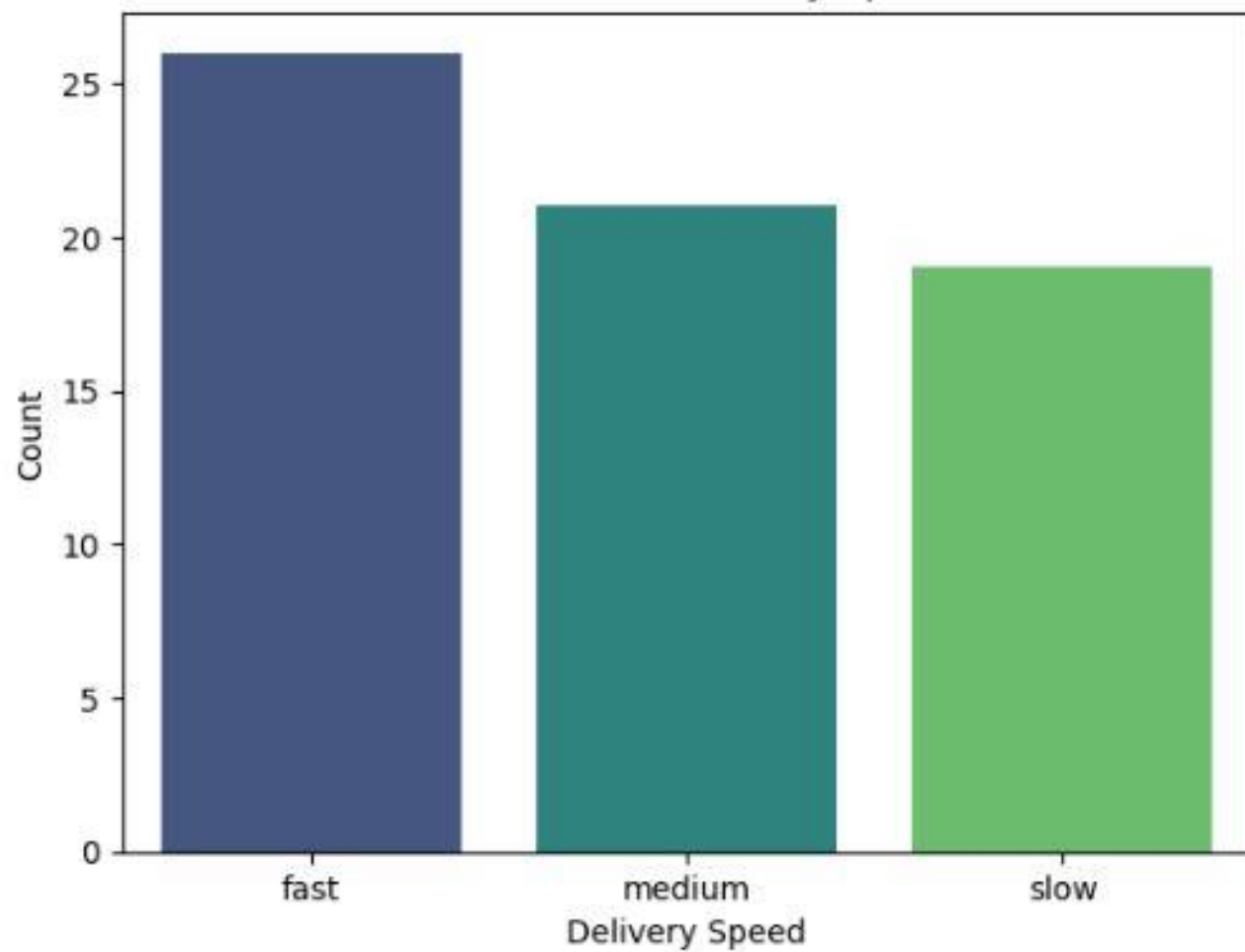


- We see that there is five outliers in the boxplot of the total profit boxplot. The mean of the total profit is between in the 0.25 $1e6$ to 0.50 $1e6$
- The total profit is highest between 0 and 0.5 million ($1e6$). After 0.5 million, the graph decreases, and it remains low around 1.5 million ($1.5e6$)

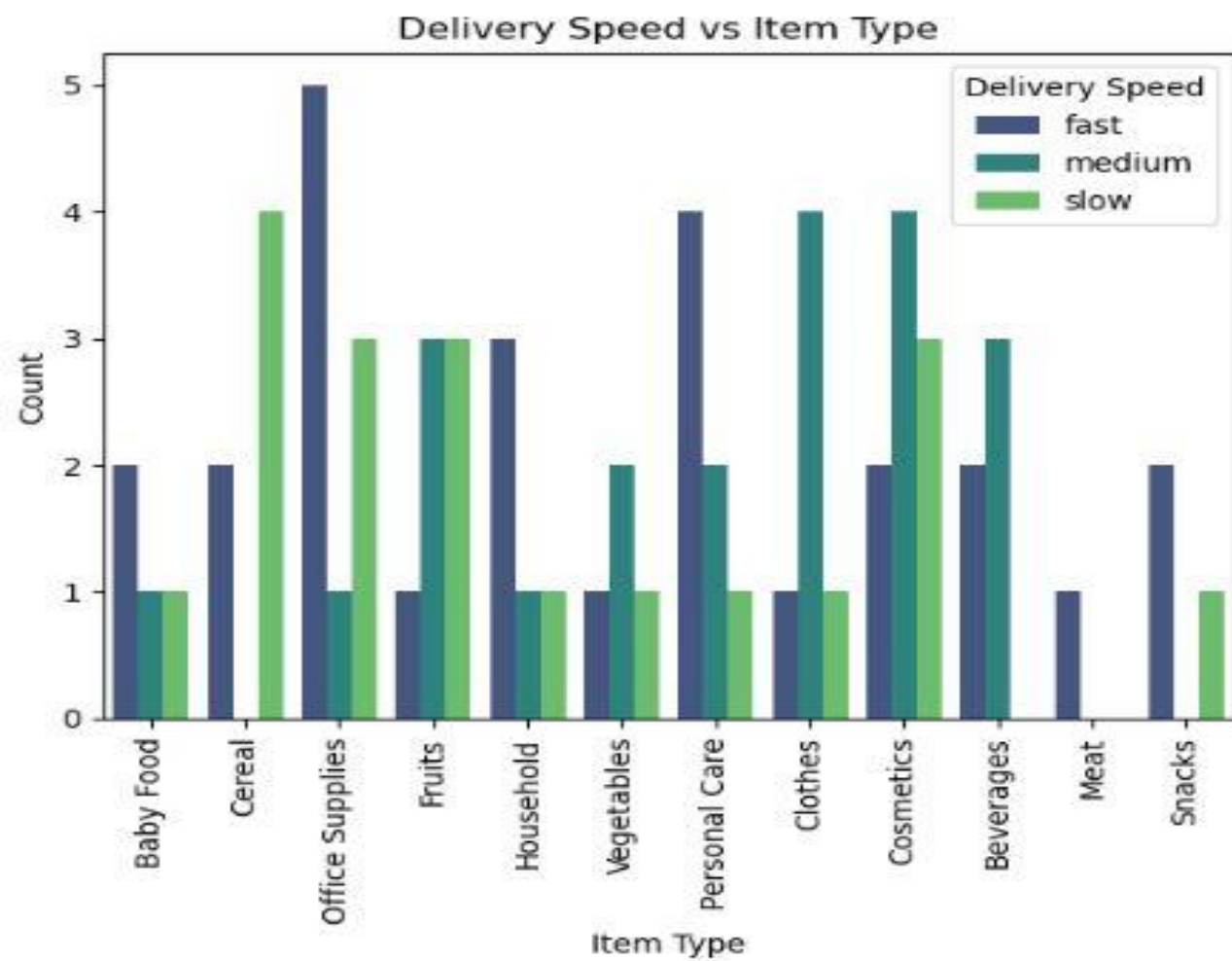
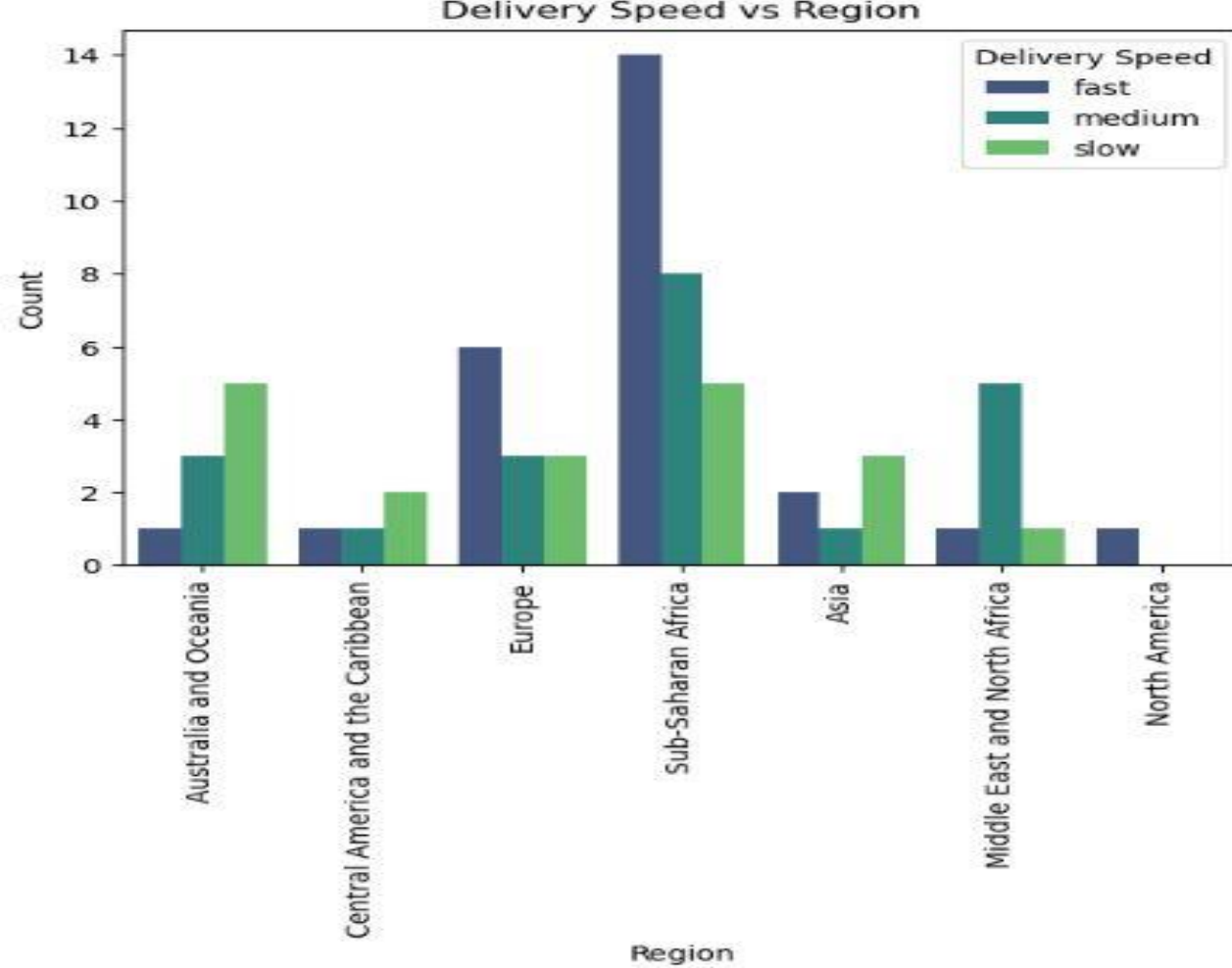
Distribution of Delivery Duration



Count Plot of Delivery Speed

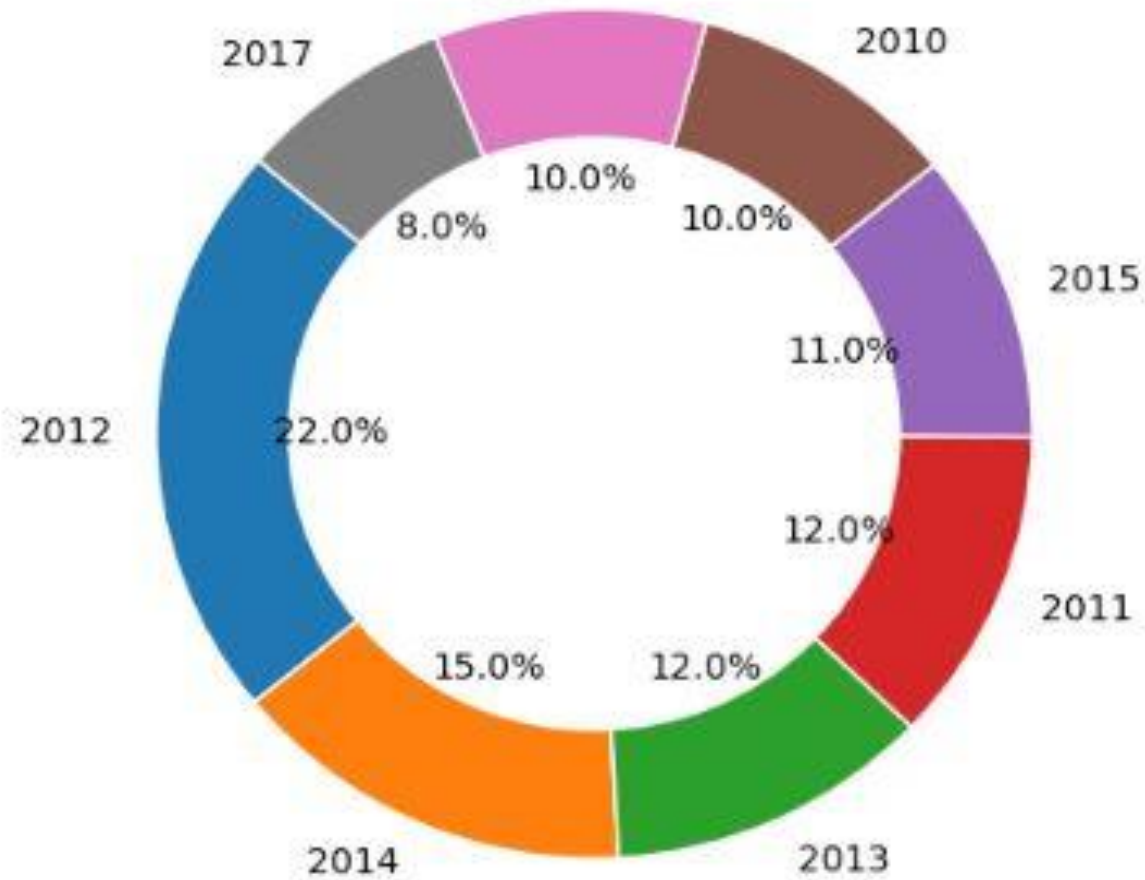


- The delivery duration is highest between 0 to 5 days and also between 40 to 45 days. It is lowest between 20 to 30 days.
- Amazon Delivery is Fast maximum time.

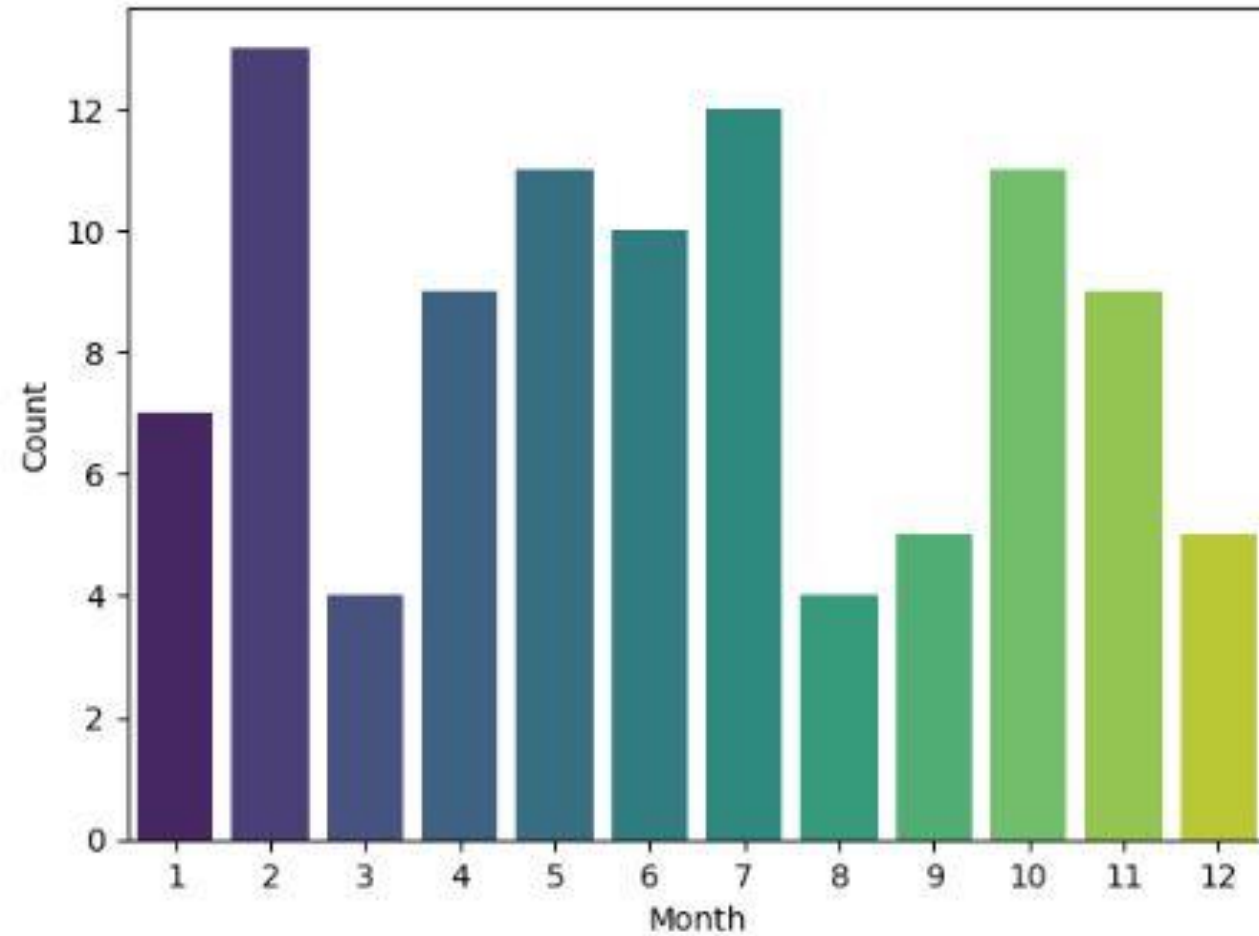


- In sub-saharn Africa, Europe, North america the amazon delivery is mostly fast.
- In Austraila and Oceania, Asia,Central amrecia and the caribbean the amazon delivery is Slow
- In the Middle East and North Africa the amazon delivery is medium.
- Office Supplies and personal care products are deliverd with in few days of order placed
- cloths, cosmetics, vegetables, fruits and beverages are deliverd medium delivery speed
- cereal, fruits are deliverd very slow speed. its taking time

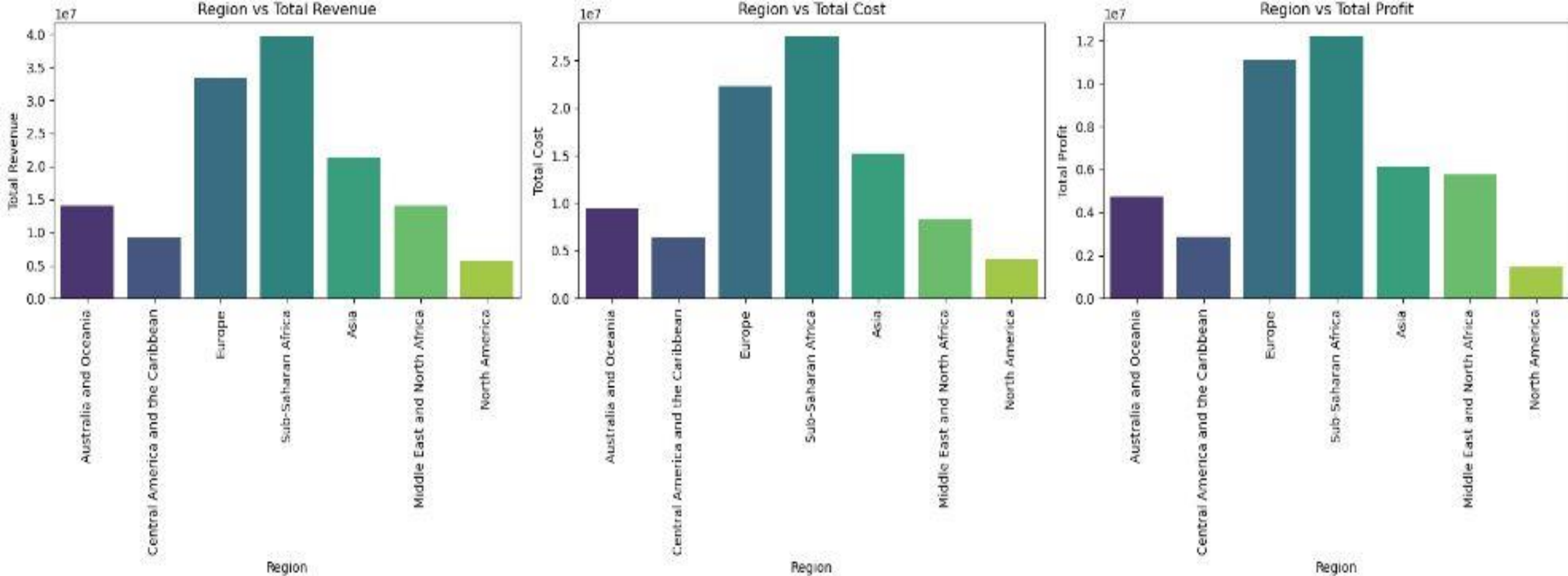
Distribution of Orders by Year
2016



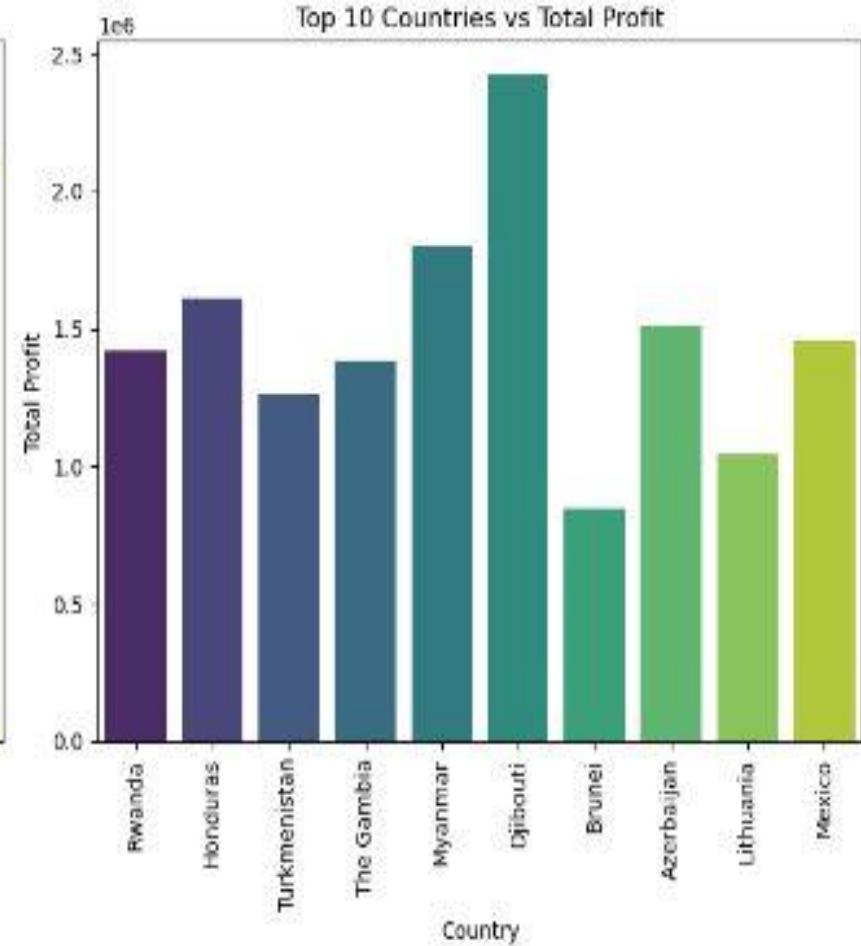
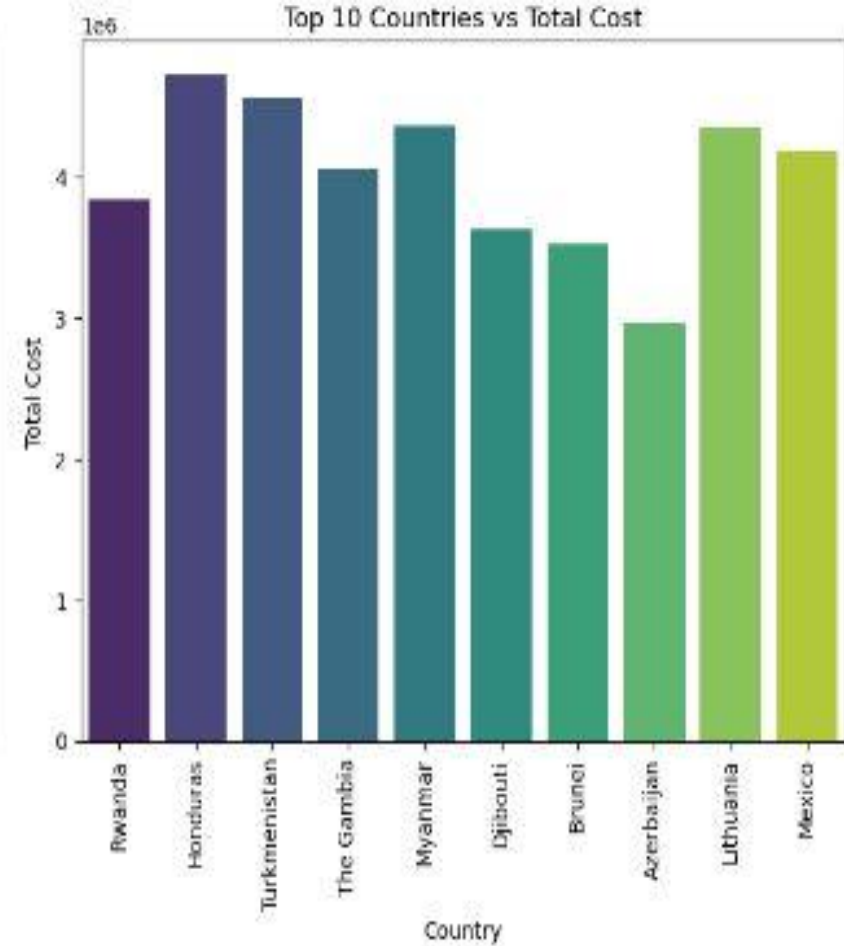
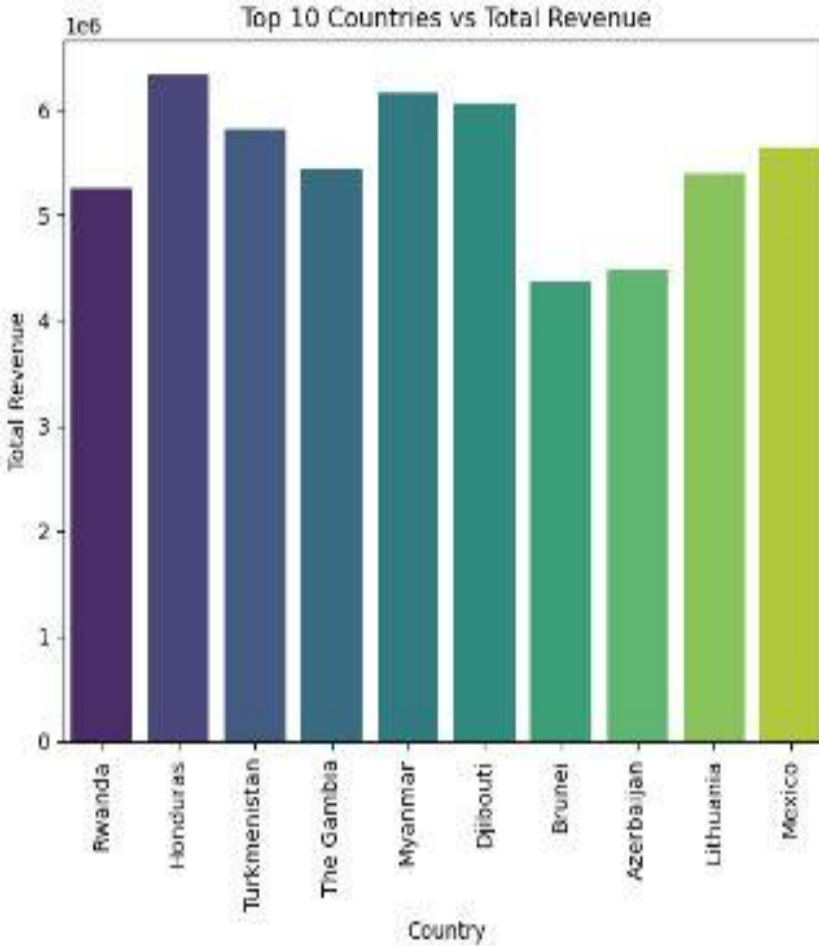
Count of Orders by Month



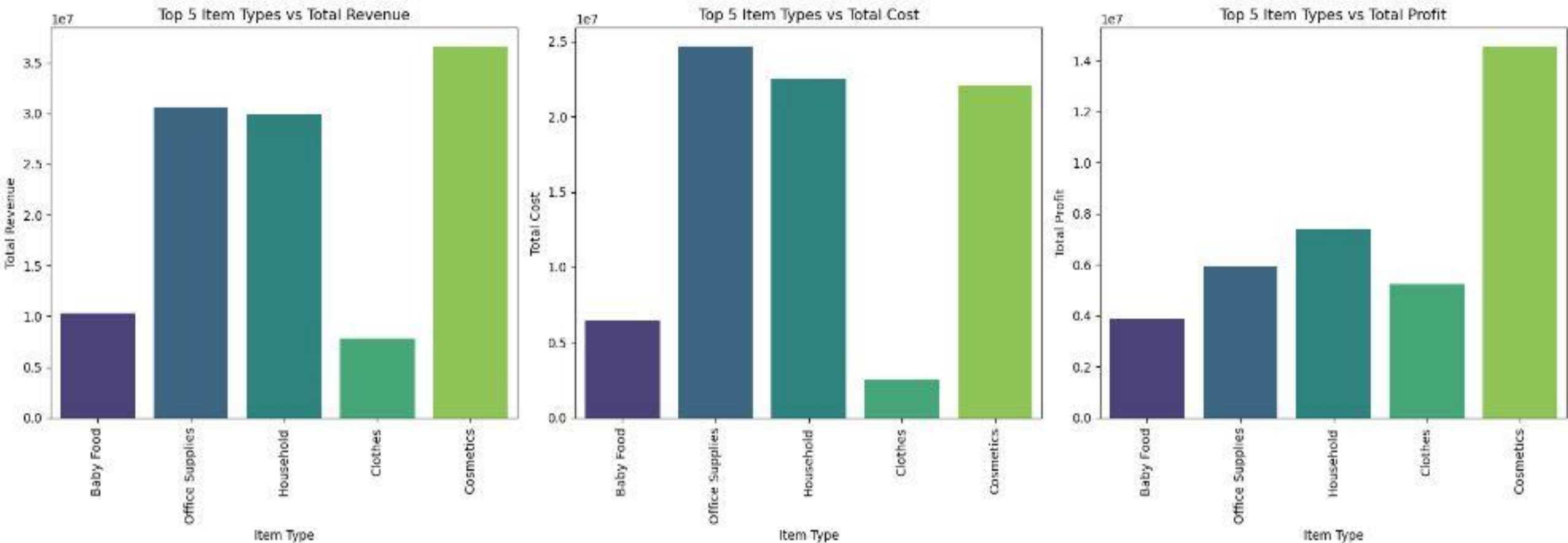
- We observe that the percentage of orders placed increased from 2010 onwards. In 2010, it was 10%. It gradually increased to 12% in 2011 and 22% in 2012, reaching its peak. However, there was a sudden decrease afterwards. In 2013, it was 12%, and in 2014, it was 15%. By 2016 and 2017, the percentage of orders placed dropped to 10% and 8%, respectively, marking the lowest points in the trend.
- We observe that in February and July, the orders placed are the highest. In May and October, they are the second highest. In January, April, June, and November, the orders are moderate. Orders are limited in the other months.



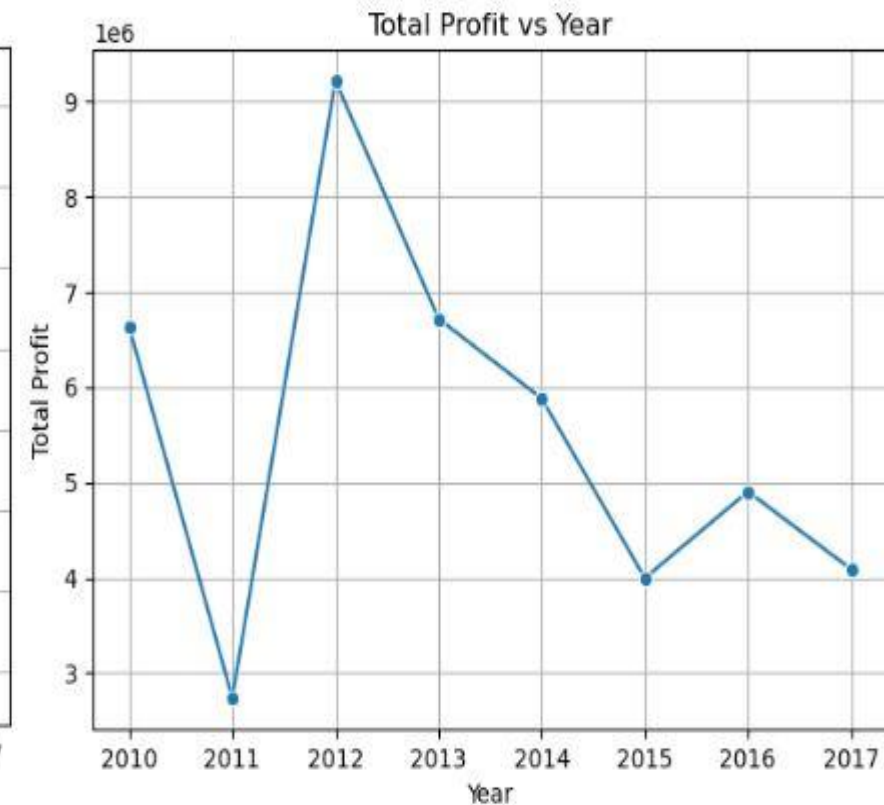
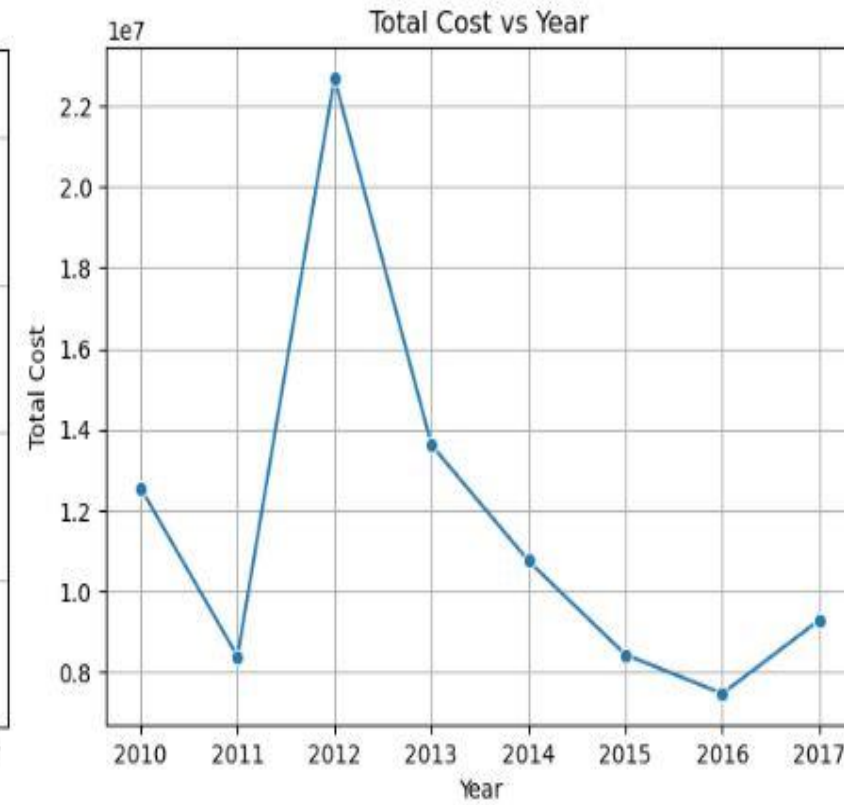
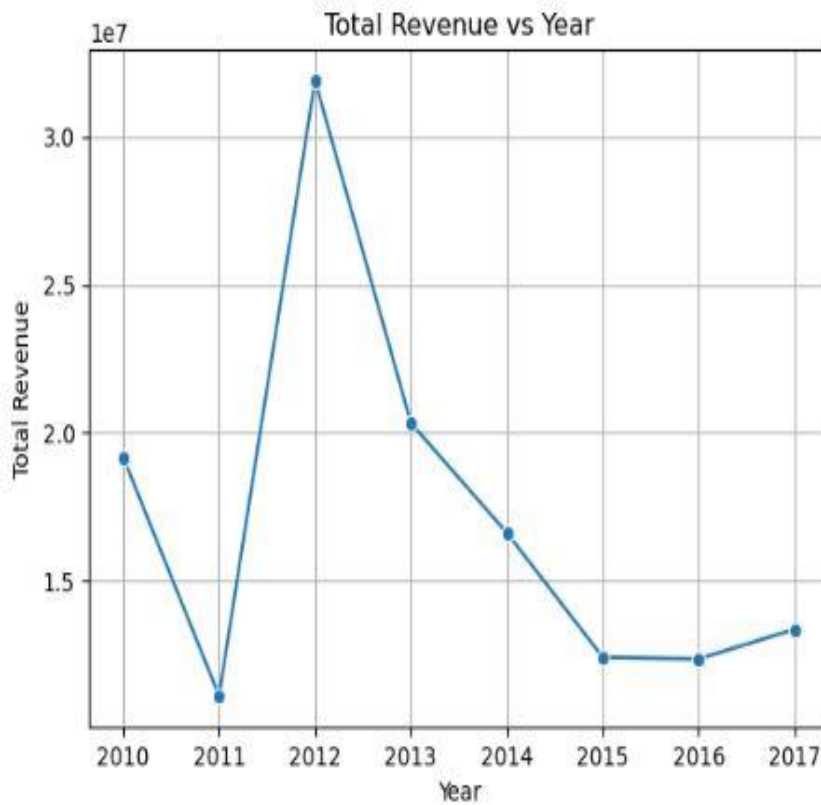
Here we see that Total Revenue, Total Profit, and Total Cost are highest in Sub-Saharan Africa and Europe. Australia and Oceania, Asia, and the Middle East and North Africa are moderate in these three fields. Central America and the Caribbean, and North America have the lowest values. and The Total Revenue ,Total Profit and Total Cost are distributed equally.



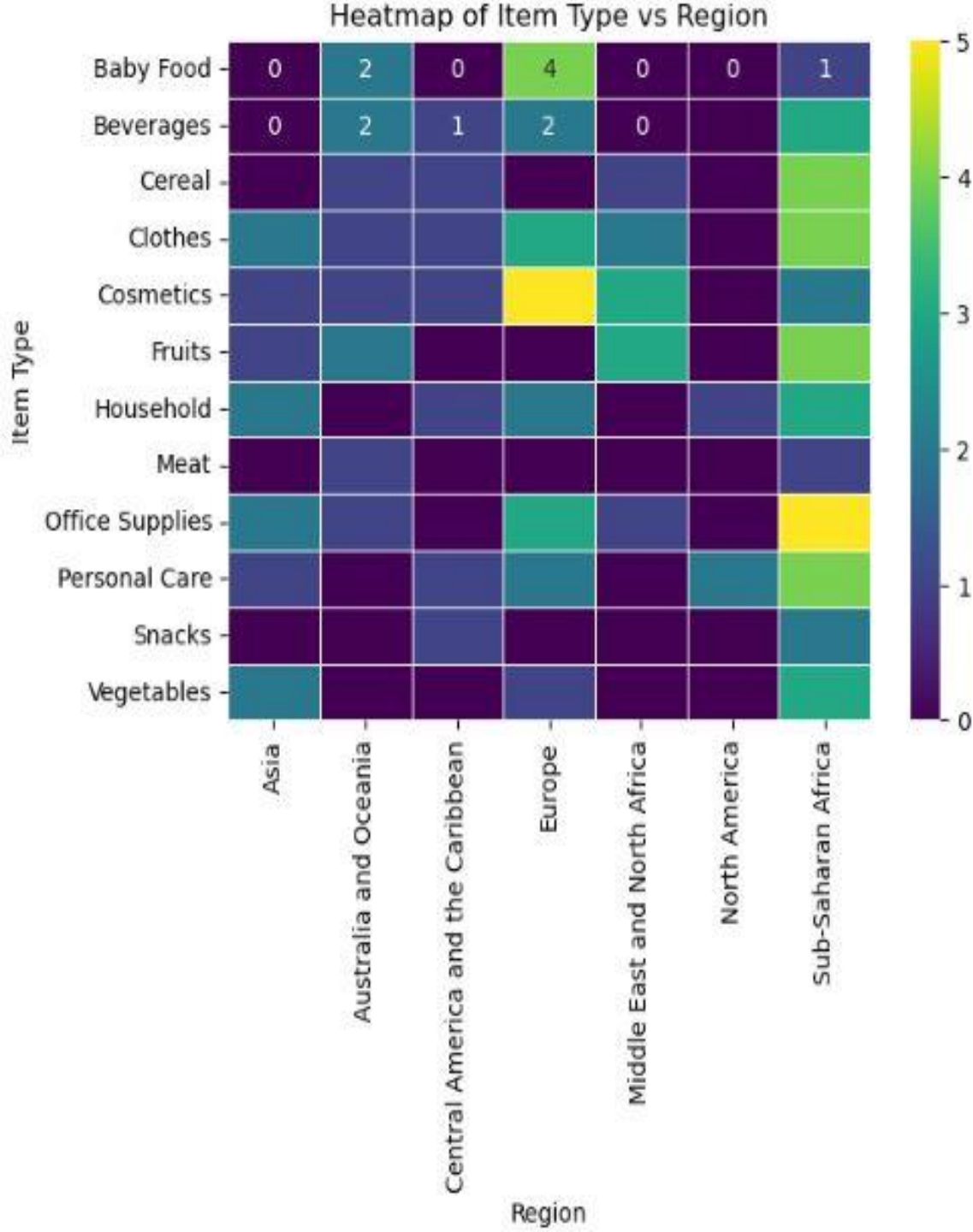
- In case of Total revenue and Total cost The top countries are Honduras, Turkmenistan, Myanmar, Djibouti, Lithuania, and Mexico.
- In Case of Total Profit the top Countries are Djibouti, Myanmar, Honduras and Azarbaizan



- The Total Revenue of cosmetics is high, and the Total Cost is low compared to the Total Revenue, making it the second highest. Total Profit is high compared to other categories but relatively low compared to the Total Revenue.
- The Total Revenue for Other Suppliers is the second highest, and the Total Cost is the highest. The Total Profit for Other Suppliers is not very good.
- The same pattern is observed for household products.
- Revenue, Cost, and Profit are equally distributed for baby food.
- The Total Profit for clothes is high compared to its Total Cost and Revenue, but the Total Revenue for clothes is minimal.

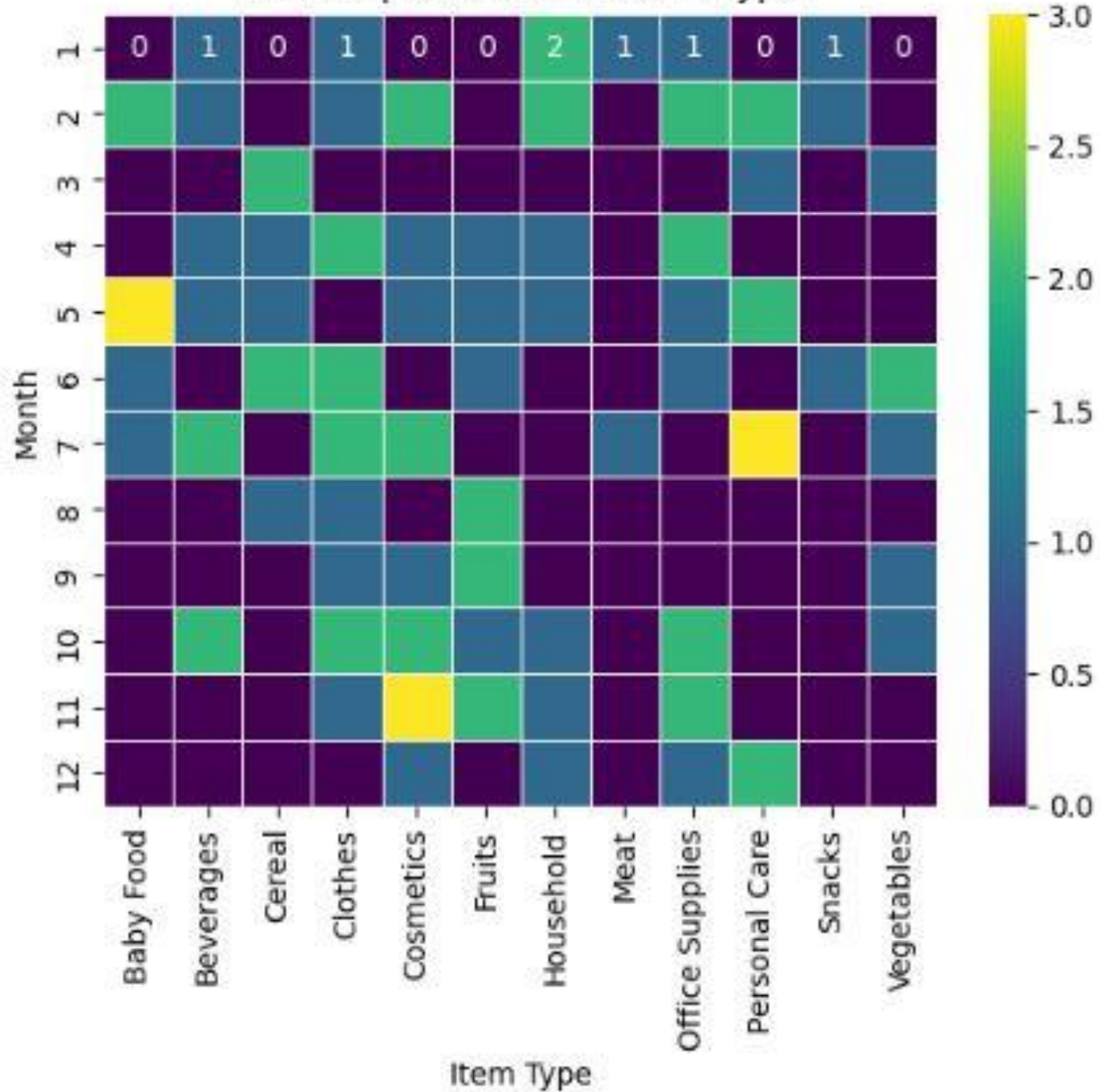


- From 2010 Total Revenue decrease and in 2011 it is very low. In 2012 at its peak. After that it started to decrease up to 2015. 2015 and 2016 has continuous graph. from 2016 there is a slight increase trend to 2017.
- Total cost is also decrease from 2010 to 2011. and go to its peak in 2012. after that it also started to decrease up to 2016. and then there is a clear increase line towards 2017
- Total Profit is also decrease in 2011 from 2010. Maximum profit is in 2012. after that it also started to fall up to 2017



- In Asia, clothes, household items, office supplies, and vegetables are moderately demanded items.
- In Australia and Oceania, baby food, beverages, and fruits are moderately demanded items.
- There are no items with high demand in Central America and the Caribbean.
- In Europe, cosmetics are highly demanded, followed by clothes, baby food, and office supplies, which are moderately demanded.
- In the Middle East and North Africa, cosmetics and fruits are moderately demanded items.
- There are no items with high demand in North America.
- In Sub-Saharan Africa, office supplies are highly demanded, while cereal, clothes, fruits, and personal care items are moderately demanded.

Heatmap of Month vs Item Type



- The demand for baby food is high in May.
- The demand for cosmetics is high in November.
- The demand for personal care items is high in July.

Recommendation :

Regional Analysis

1.The company should increase its business in Asia, the Middle East and North Africa, Australia and Oceania, and especially in North America. They need to conduct more research to understand why Amazon is not as popular in these specific regions.

Delivery Duration

2. Amazon's delivery duration is slow in Australia and Oceania, Asia, Central America, and the Caribbean. The delivery duration is moderate in the Middle East and North Africa. This slower delivery is one reason affecting Amazon's popularity in these areas. Therefore, Amazon should focus on improving delivery times in these regions.

3. From the data, we see that clothes and cosmetics are the most popular items among customers. Unfortunately, the delivery timing for these two items is moderate. It should be faster to maintain the company's goodwill and help the business grow.

Order Distribution

4. We observe that the percentage of orders placed increased from 2010 onwards. It was 10% in 2010, gradually increasing to 12% in 2011 and peaking at 22% in 2012. However, there was a sudden decrease afterward, with 12% in 2013, 15% in 2014, and dropping to 10% in 2016 and 8% in 2017. The company should analyze the sales data of 2013, 2014, 2015, 2016, and 2017 to identify the problems affecting the company's profit growth.

5. We observe that the highest number of orders are placed in February and July. May and October have the second highest. January, April, June, and November have moderate order volumes. Orders are limited in the other months. The company should focus more on February and July and research why the number of orders falls in other months.

Profit and Loss

6. Although Amazon is not very popular in the Middle East and North Africa, it generates higher profit margins compared to revenue in these areas. Therefore, Amazon should focus more on this region.

7. Amazon's profit margin is low in the top 10 countries. The company should research this issue.

8. Profit is higher for cosmetics compared to other items.

9. We see that Amazon had the maximum profit in 2012. After that, there was no significant profit, and revenue and cost also declined. The company should research the 2012 sales data to gather information on why that particular year had maximum profit. One reason might be that Amazon invested maximum money into their business that year.

Popular Items

10. Amazon should increase the sales of cosmetics in Europe and office supplies in Sub-Saharan Africa.

11. The demand for baby food is high in May, the demand for cosmetics is high in November, and the demand for personal care items is high in July. This information can help the company understand seasonal demand trends.

CROP PRODUCTION IN INDIA

Performing a deep analysis of Crop Production Details through Python

Sauradip Pradhan
Date - 27/07/2024

Agenda:

- **Introduction to Crop Production Data**
- **Data Overview**
- **Explanation of the Sales Analysis Process**
- **Exploratory Data Analysis (EDA)**
- **Visualizations and Findings**
- **Summary**

Introduction to Crop Production Data

The Agriculture business domain, as a vital part of the overall supply chain, is expected to highly evolve in the upcoming years via the developments, which are taking place on the side of the Future Internet. This paper presents a novel Business-to-Business collaboration platform from the agri-food sector perspective, which aims to facilitate the collaboration of numerous stakeholders belonging to associated business domains, in an effective and flexible manner. This dataset provides a huge amount of information on crop production in India ranging from several years. Based on the Information the ultimate goal would be to predict crop production and find important insights highlighting key indicators and metrics that influence crop production.

Data Overview

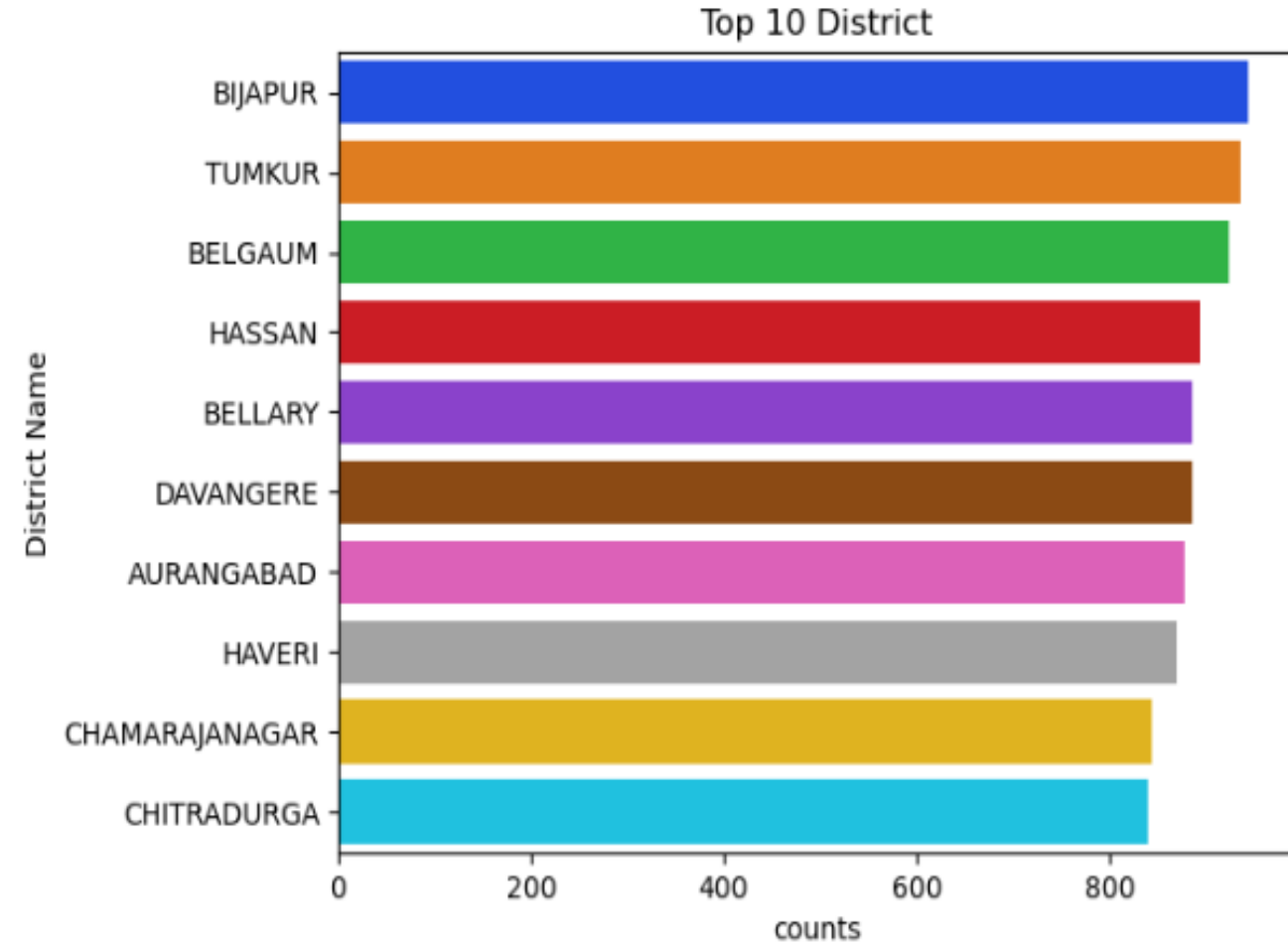
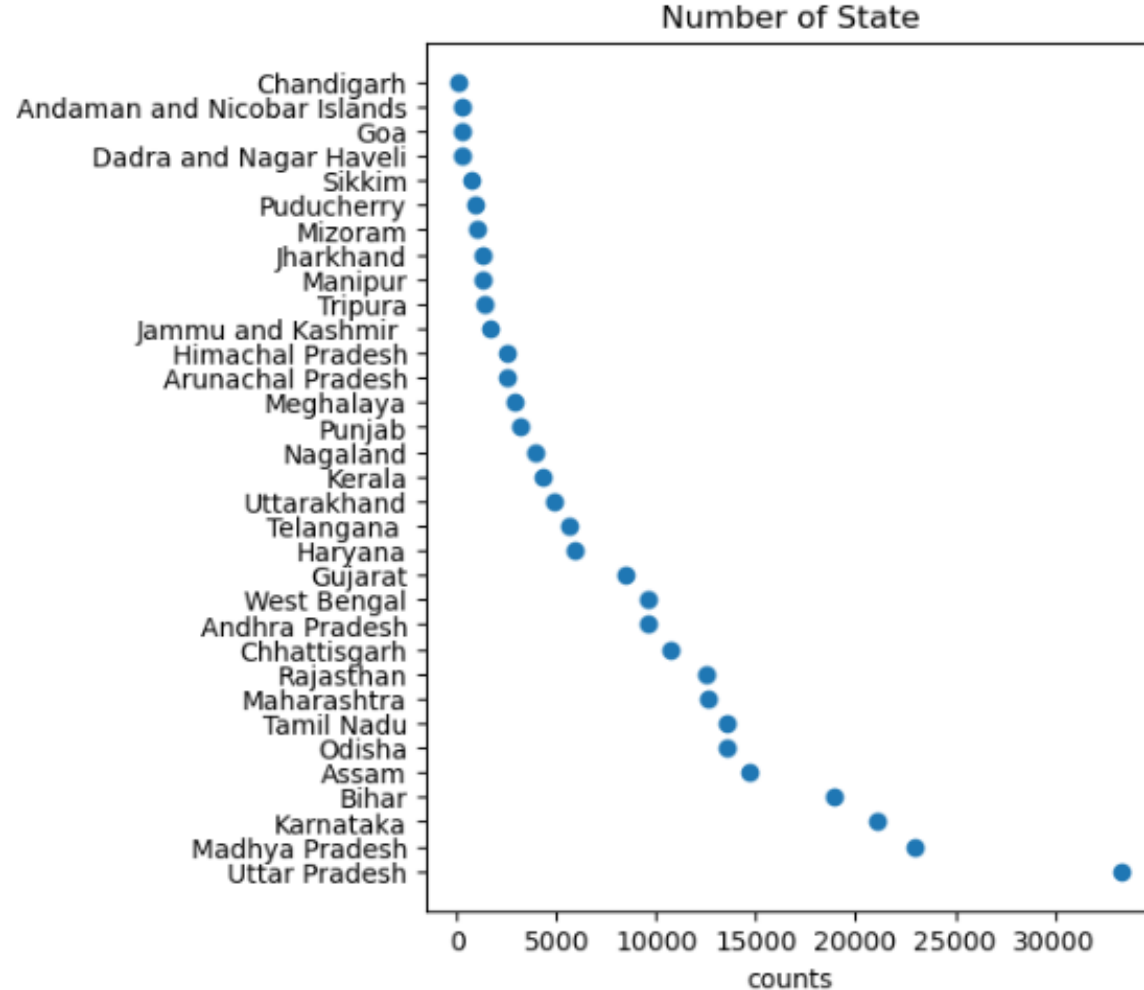
- Name of the dataset is Crop Production data.csv
- The dataset has 246091 rows and 7 columns
- There is some null values in 'Production' column of the dataset
- 4 columns are Object data type, 1 columns are Integer data type and 2 columns are float data type

Explanation of Sales Data Analysis Process

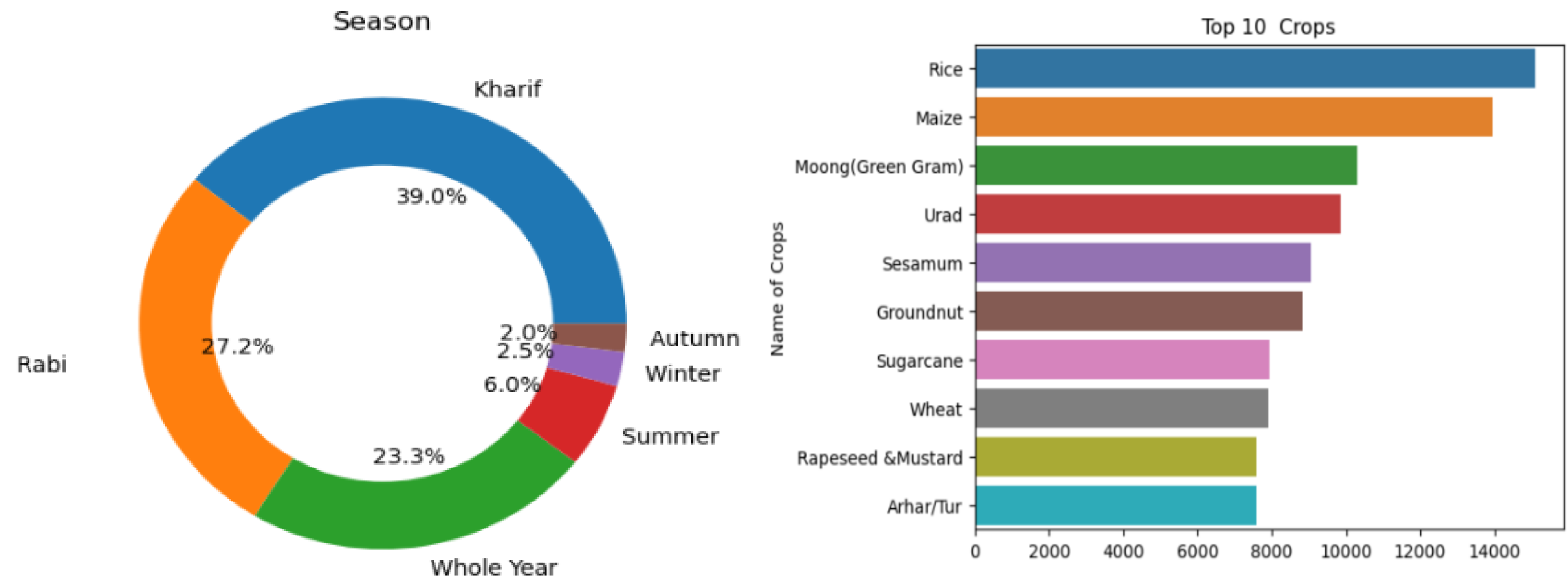
- Check the shape of the dataset, identify any null values, and verify the data types of each column.
- Fill the null values by the mean values .(Null values are in the numerical column)
- Create two new columns one is 'Category of Area' and other one is 'Production Category'
- Perform Univariate, Bi variate and Multivariate Analysis of the dataset.
- Note: All operations are conducted in Python using Jupyter Notebook.

Exploratory Data Analysis

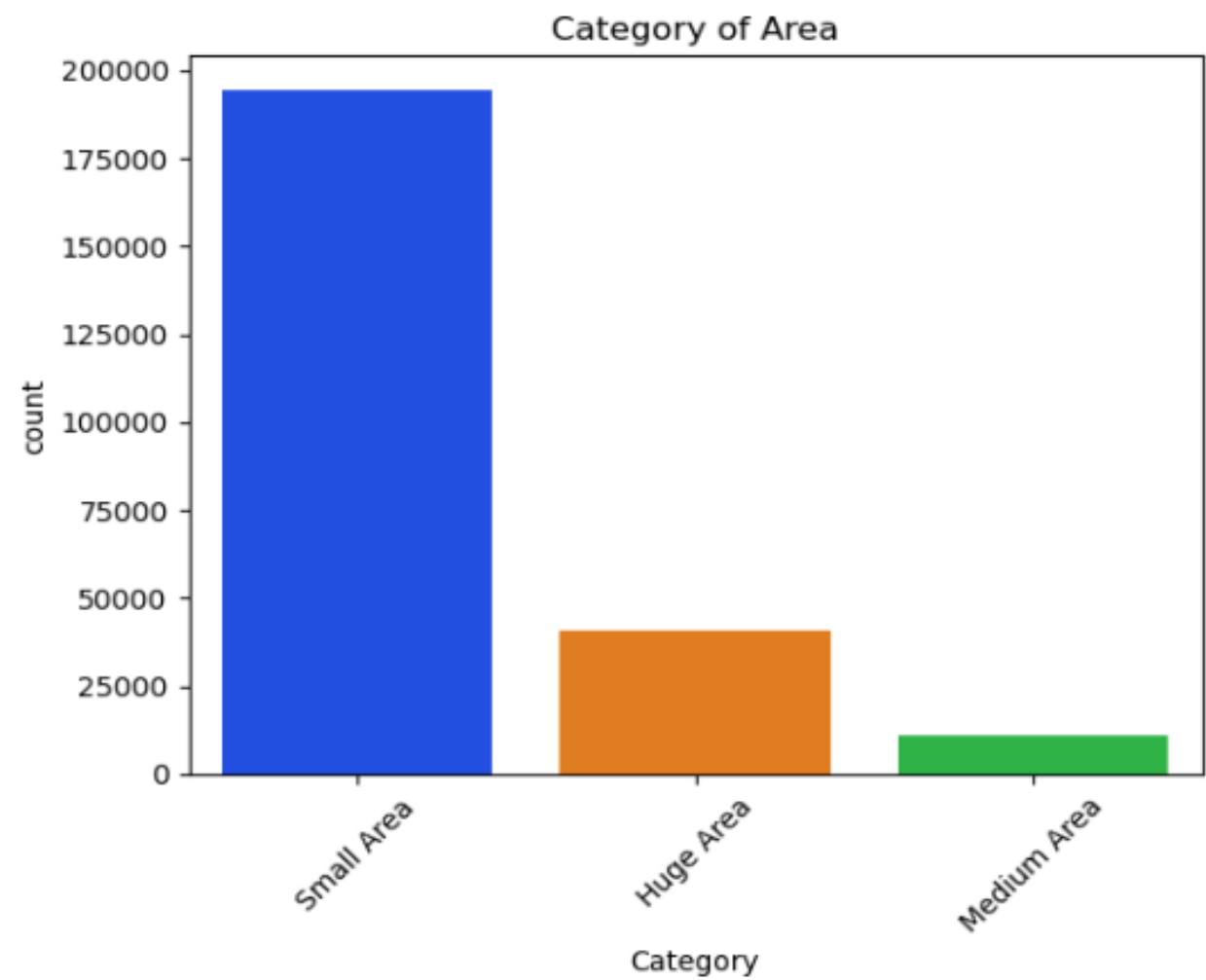
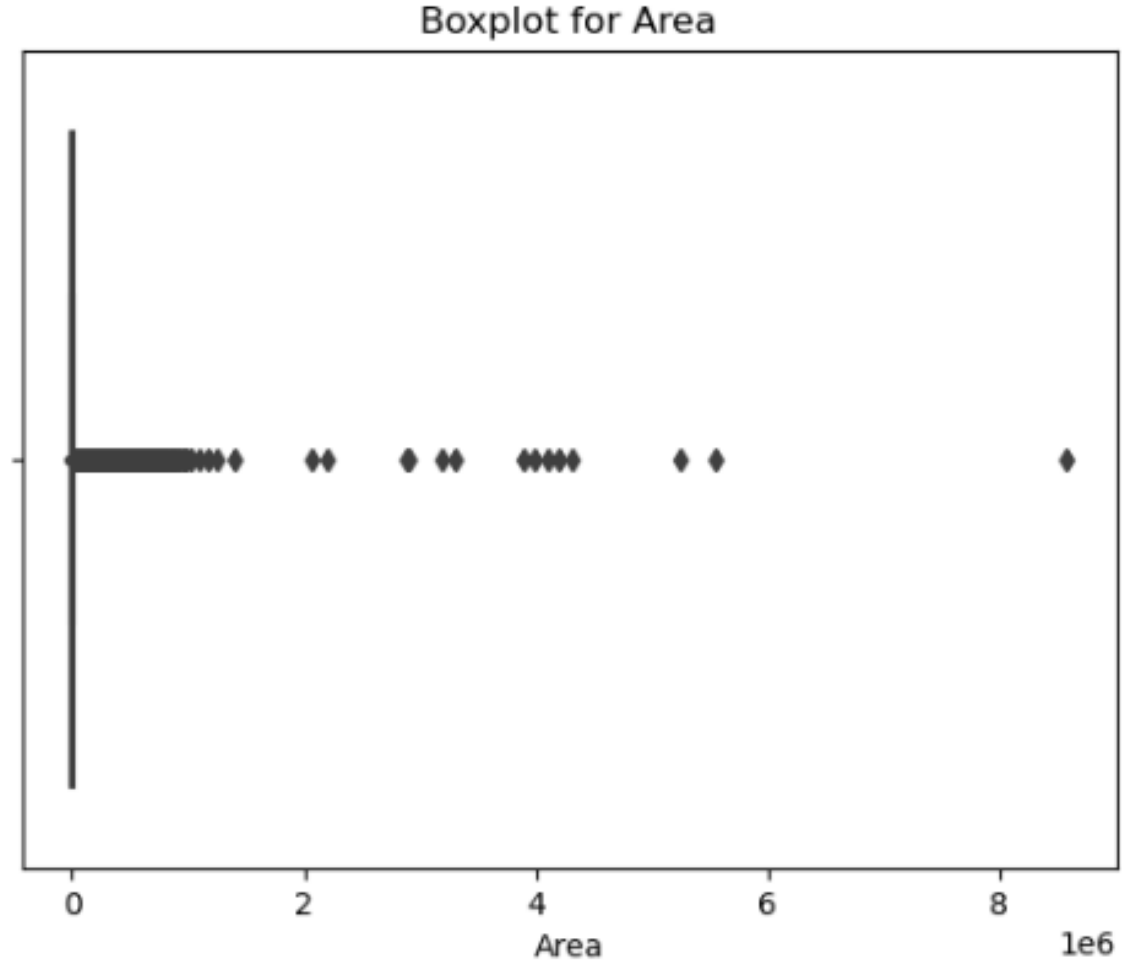
- In EDA we create Multiple Graphs and Charts.
- To import the Dataset and for Mathematical calculation we use Pandas and numpy
- To create Graphs and Charts we use seaborn and matplotlib



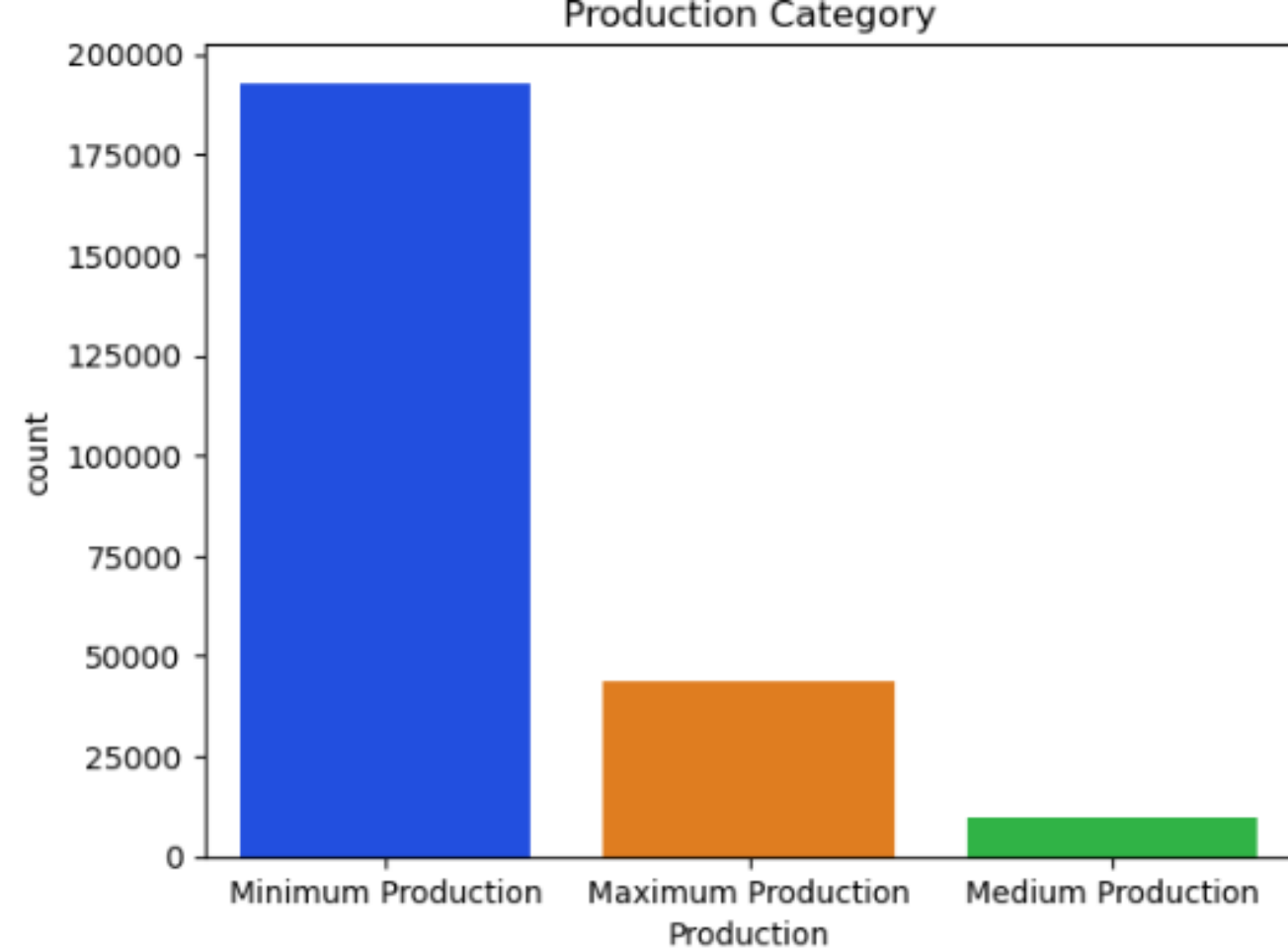
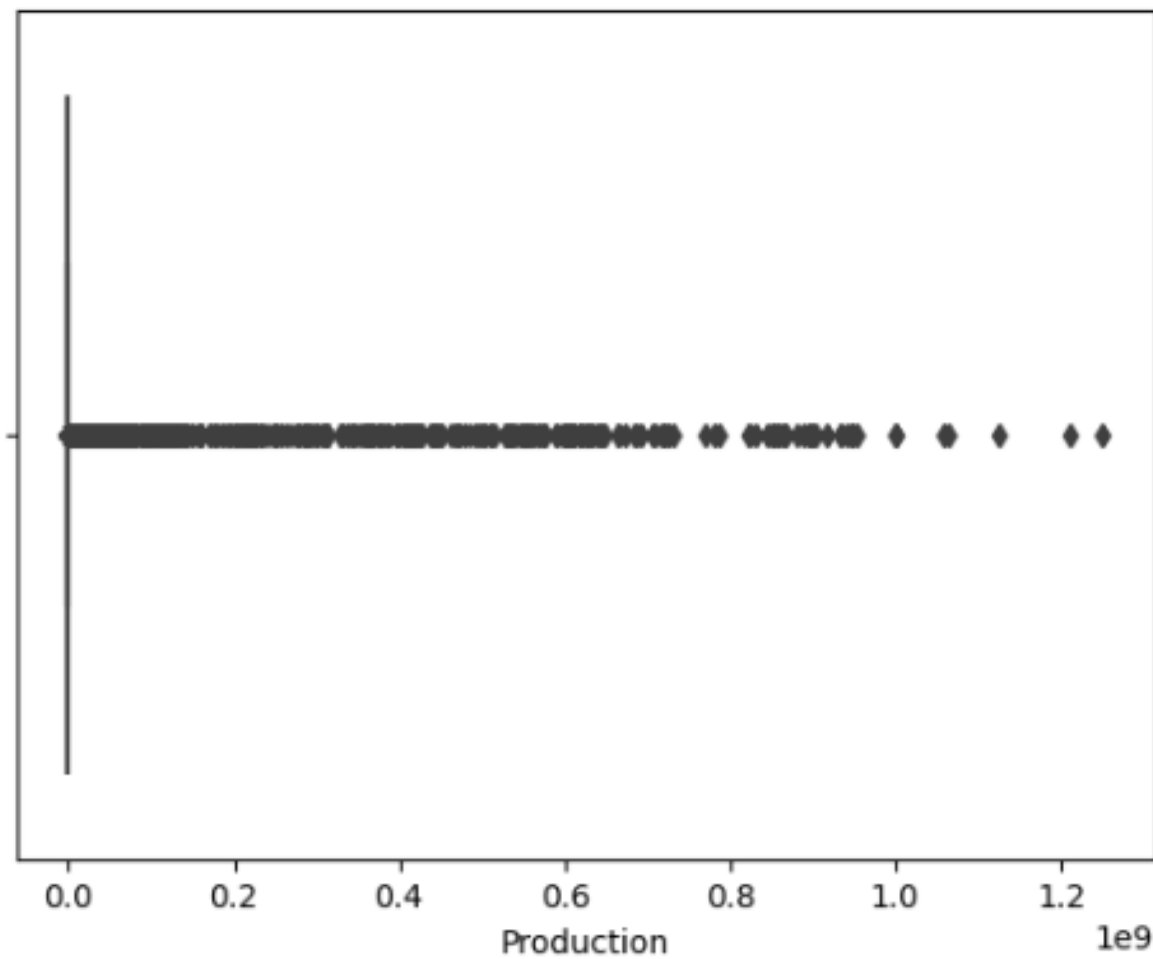
- Uttar Pradesh is overrepresented in the dataset, as evidenced by the univariate analysis graph showing its higher frequency of occurrences. Madhya Pradesh, Karnataka, and Bihar follow in terms of the number of entries. This suggests that, overall, the number of crops recorded is higher in these states.
- We see that among the top 10 districts, Bijapur has the highest number of entries. It is followed by Tumkur, Belgaum, and Hassan.



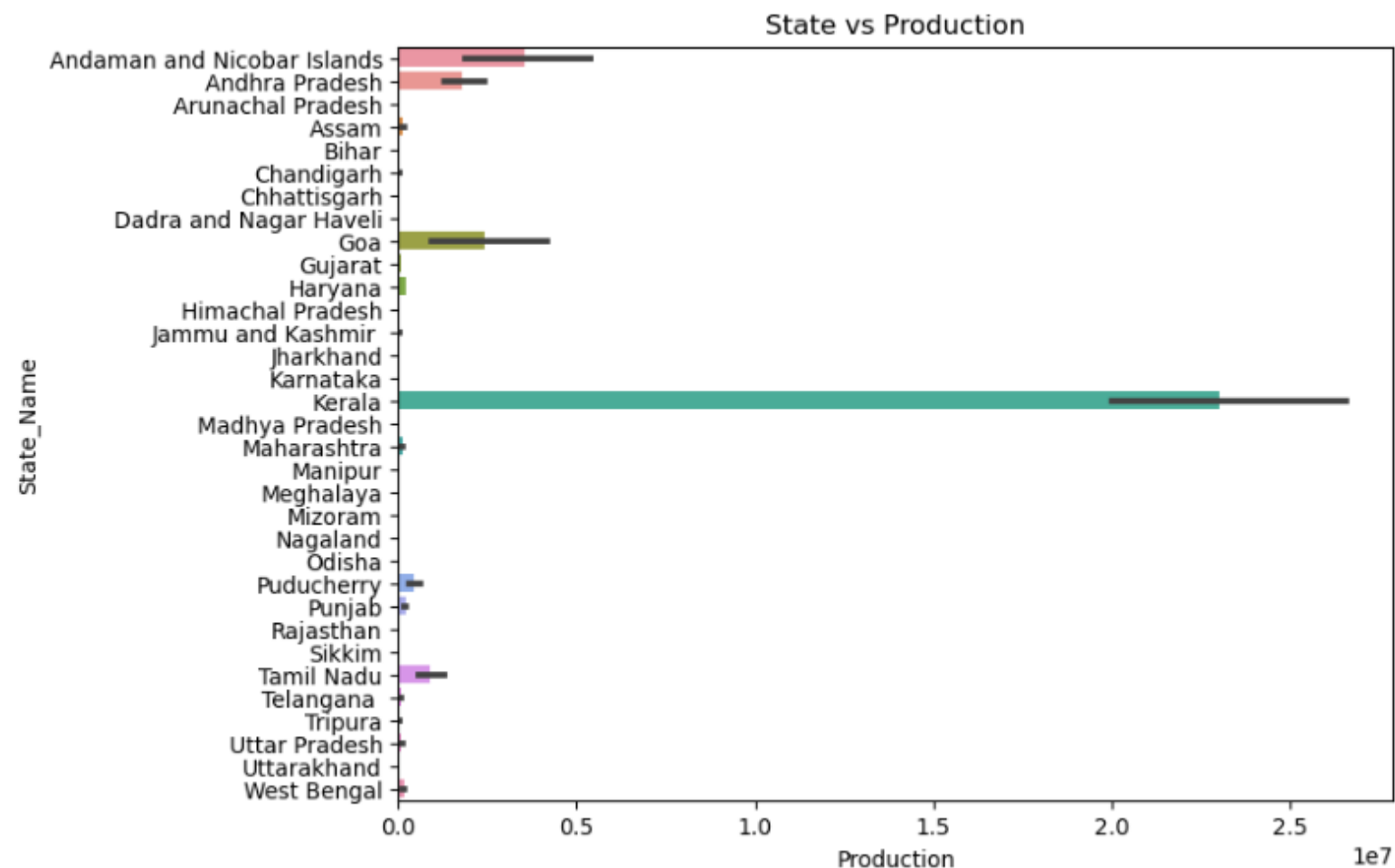
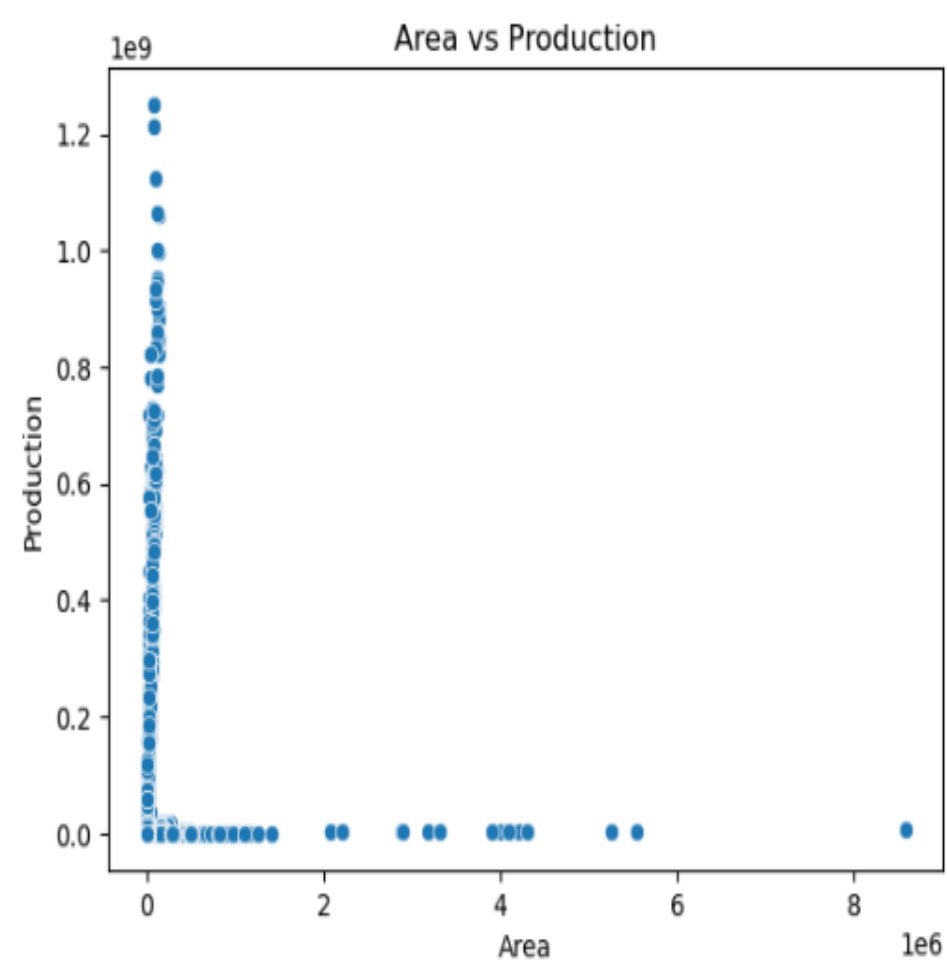
- We observe that the number of crops in the Kharif season is the highest, followed by the Rabi season. The number of crops in the Autumn season is the lowest
- The number of Rice entries is the highest, followed by Maize. Moong (Green Gram), Urad, Sesame, Groundnut, and Sugarcane follow in descending order.



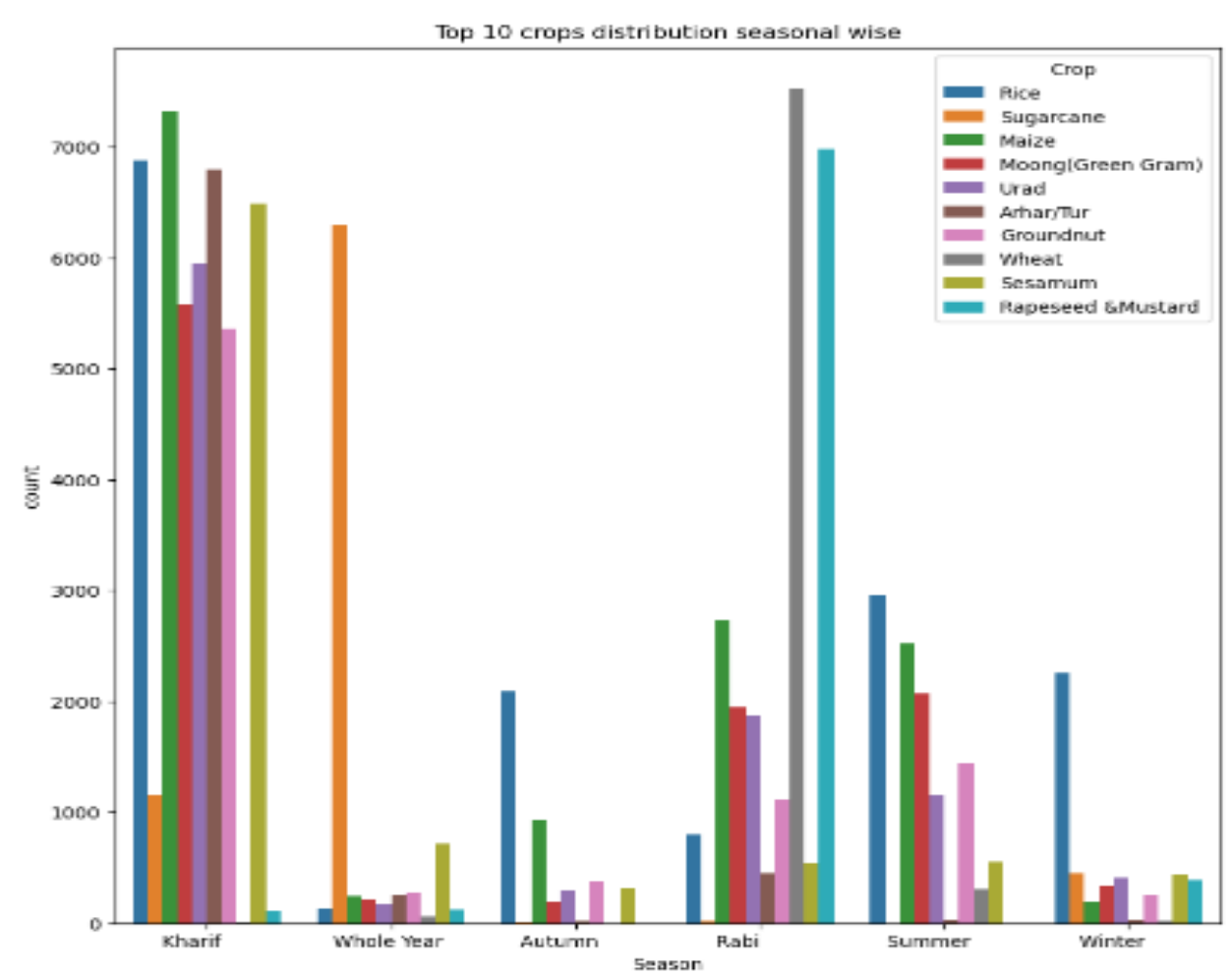
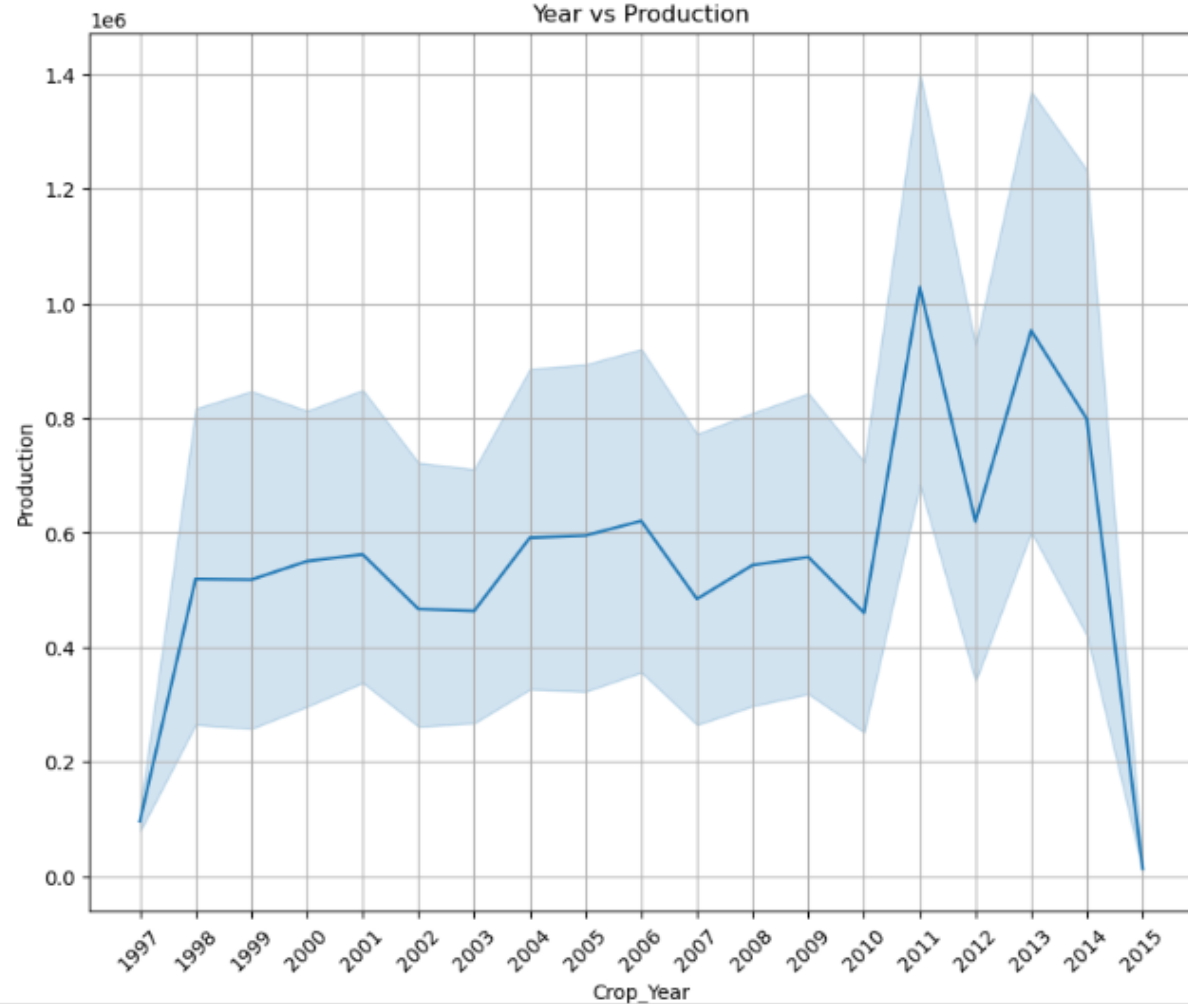
- There is too much outliers in the Area boxplot. So we make a new column for Area
- Number of small areas is high than Huge type Area and Medium Type Area



- There is so many outliers in the Production column also so we also make new column for the prouction
- Production rate is Minimum in India. Rate of maximum Production is very low.

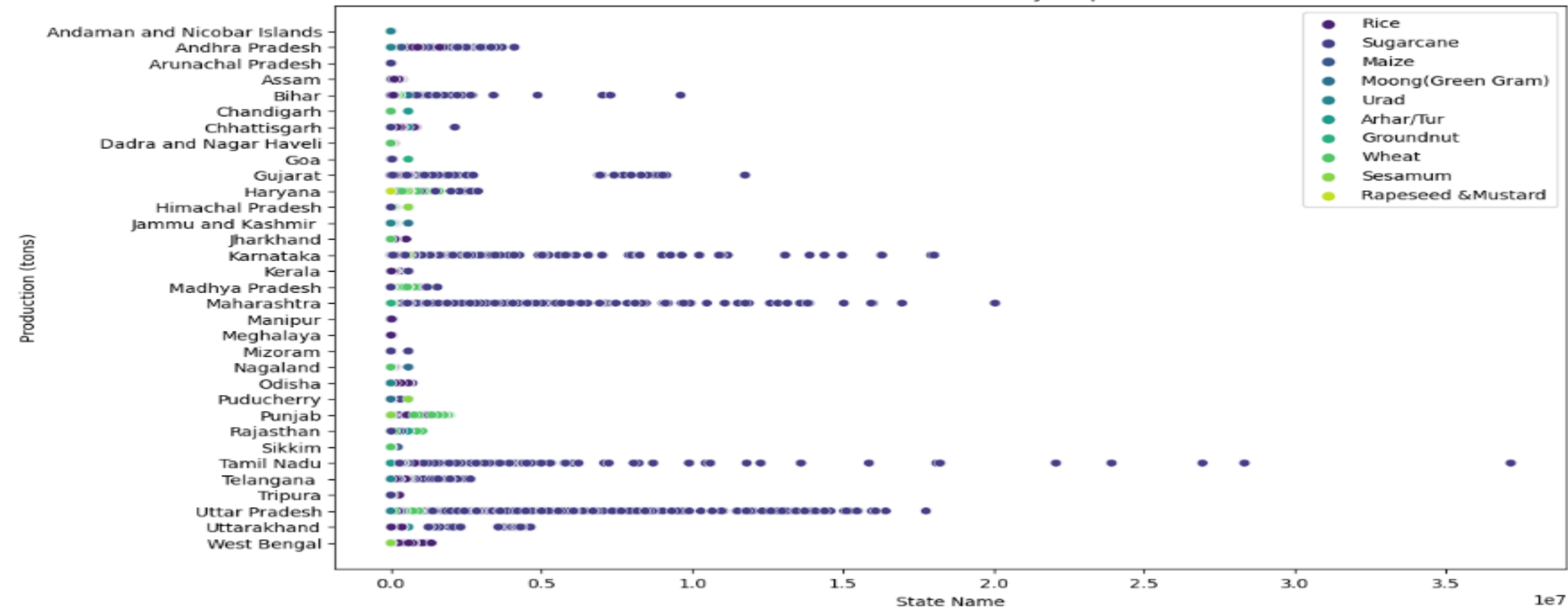


- Maximum number of production is happen in Minimum Area. (around 0-1 le6)
- The highest production is in the state of Kerala in India. Following Kerala are the Andaman and Nicobar Islands, Goa, and Andhra Pradesh.

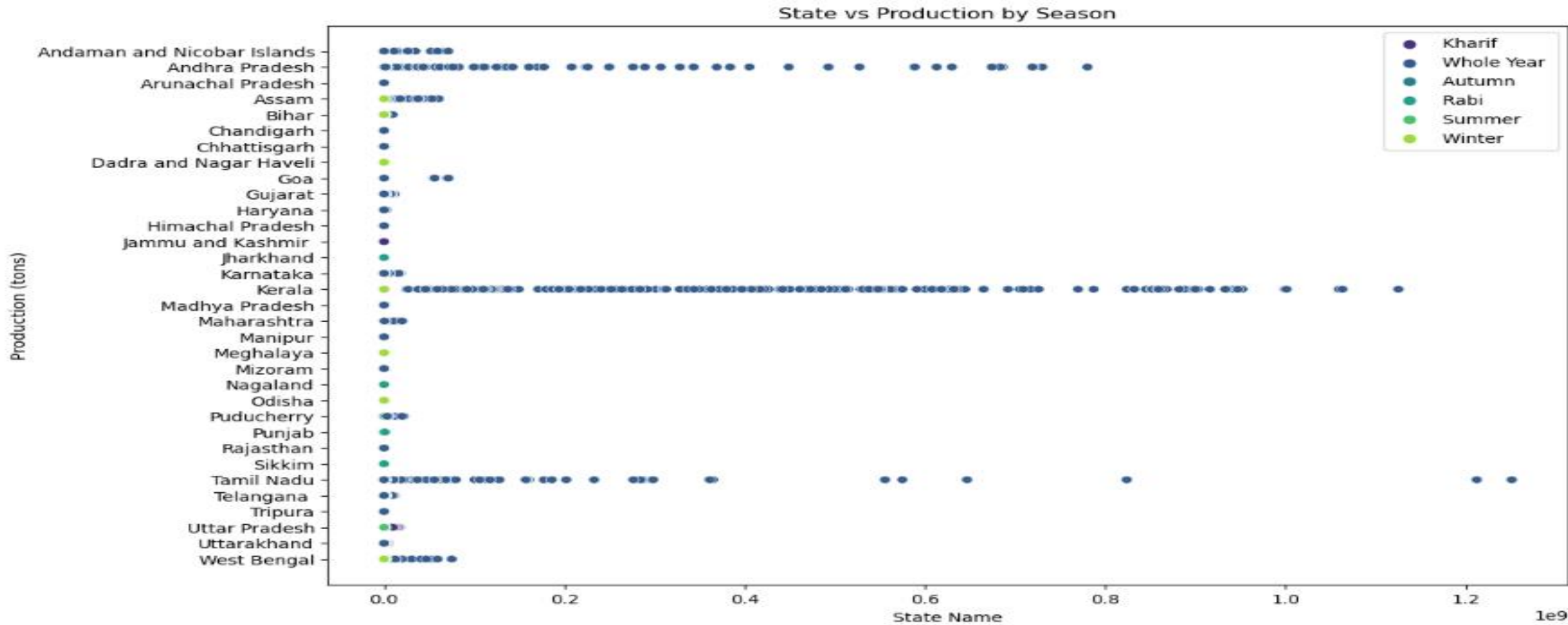


- The production rate was high in the year 2011. It declined in 2012, started to increase again in 2013, and reached a very low point in 2015.
- In the Kharif season, the production rates of Maize, Rice, Arhar/Tur, and Sesame are very high, while the production of Rapeseed and Mustard is very low.
- Rice production is high in the Autumn and Winter seasons, but not as high as in the Kharif and Summer seasons." - "Wheat and Rapeseed and Mustard production are high in the Rabi season.

State vs Production by Crop



- We see that the maximum number of states produce rice. Tamil Nadu, Maharashtra, Karnataka, Uttar Pradesh, and Gujarat are the states where rice is produced in large amounts.
- In Punjab and Haryana, the production of wheat, sesame, and rapeseed & mustard is popular.



- In Kerala, Andhra Pradesh, and Tamil Nadu, Kharif season crops are produced in large quantities. Overall, Kharif season crops are very popular in India.
- However, crops are produced in different seasons across various states. For example, in Odisha, winter crops are produced. In West Bengal, both winter and Kharif season crops are produced.
- States like Assam, Bihar, Mizoram, and Meghalaya produce both winter and summer crops.

Summary

State Representation:

- Uttar Pradesh is overrepresented in the dataset, followed by Madhya Pradesh, Karnataka, and Bihar. These states have a higher number of crop entries.

Top Districts: Among the top 10 districts, Bijapur has the highest number of entries, followed by Tumkur, Belgaum, and Hassan.

Seasonal Crop Distribution:

- The number of crops in the Kharif season is the highest, followed by the Rabi season. The Autumn season has the lowest number of crops.
- In Kerala, Andhra Pradesh, and Tamil Nadu, Kharif season crops are produced in large quantities. Overall, Kharif season crops are very popular in India.
- In Odisha, winter crops are produced, while West Bengal produces both winter and Kharif season crops.
- Assam, Bihar, Mizoram, and Meghalaya produce both winter and summer crops.

Crop Types:

- The number of Rice entries is the highest, followed by Maize. Moong (Green Gram), Urad, Sesame, Groundnut, and Sugarcane follow in descending order.
- In the Kharif season, the production rates of Maize, Rice, Arhar/Tur, and Sesame are very high, while the production of Rapeseed and Mustard is very low.
- Rice production is high in the Autumn and Winter seasons, but not as high as in the Kharif and Summer seasons.
- Wheat and Rapeseed and Mustard production are high in the Rabi season.
- Punjab and Haryana are known for the production of wheat, sesame, and rapeseed & mustard.

Production Analysis:

- There are many outliers in the Area and Production columns. New columns were created to categorize the areas and production values.
- The number of small areas is higher than huge and medium-type areas.
- The production rate is minimal across India, with the maximum production rate being very low.
- Maximum production occurs in minimal areas (around 0-1 le6).
- The highest production is in Kerala, followed by the Andaman and Nicobar Islands, Goa, and Andhra Pradesh.

Yearly Production Trends:

- The production rate was high in 2011, declined in 2012, increased again in 2013, and reached a very low point in 2015.

Acknowledgement

I would like to express my sincere gratitude to **UNIFIED MENTOR** for providing me with the opportunity to undertake this one-month project from July 1st to August 1st. During this period, I had the chance to work on two significant projects: Amazon Sales Data Analysis and Crop Production Analysis.

I am immensely grateful for the invaluable guidance and support provided by the team. Their expertise and insights were crucial in completing these industrial-based projects. Thank you for this enriching experience and for helping me grow professionally.

Thank You