Amazon Sales Analysis Project

Understanding the sales patterns and customer purchasing habits on the Amazon website is the aim of this project on Amazon Sales Data. This analysis will help us in better understanding of the variables that may affect sales as well as in enhancing customer satisfaction.

DEFINE

1) **Analysis 1 – Advertising**
   - DEFINE – Finding website traffic trends based on sales by month, day, and time of day is the goal of this analysis. On an e-commerce website, advertising is one of the primary methods for driving sales, so understanding when and why customers want to buy is essential. This analysis of purchasing time will also help to understand the best time to put sales and discounts on the website to attract sales.
   - KPI – "Sales" is going to be the key performance indicator in this examination. Which month, day, or hour is appropriate for advertising discounts will depend on sales.
   - CTQ – “Customers” are Critical to Quality in this analysis because it benefits the customers to offer promotions and discounts during the best times when the majority of customers are most likely to be able to make purchases. Providing discounts on the right time also makes the customers happy with increase customer retention.
2) **Analysis 2 – Warehousing (Supply Chain Management)**
   - DEFINE - The purpose of this analysis is to determine the product demand in each city. With the use of this information, we can improve the efficiency of our supply chain management, which will lower operational costs for the business and increase revenue and customer satisfaction. Supply chain management optimization supports the efficient use of a company's human and financial resources.
   - KPI – “Quantity Ordered” is the Key Performance Indicator here. The quantity ordered will let us know how many products are required in a specific city.
   - CTQ - “Customers” and “Company” are Critical to Quality in this analysis because a company may lose customers and experience low customer satisfaction if there aren't enough units available for sale in a city. Additionally, poor inventory results in the waste of human and financial resources.
3) **Analysis 3 – Recommendation Association**
   - DEFINE – The aim of this analysis is to identify the products that are usually purchased in combination. This will enable us to suggest products to customers that they can purchase in addition to the item they are presently considering buying.
   - KPI – “Quantity Ordered” is the Key Performance Indicator here. Grouped Quantity orders of the products will reveal the combinations that buyers are buying.

- CTQ – “Customers” are Critical to Quality here because a recommendation system will help customers to find more relevant products to buy and will be happier to get suggestion of the products, they may be willing to buy.

MEASURES

Data used here is the Amazon Sales Data 2019. “Asian Academy of Film and Television (AAFT)” which is an academic institute, provided the data.

The dimensions of the data were (186850, 6). We got the data in .csv format which we stored in our local systems and then uploaded in Jupyter Notebook to use it.

Following are the features of the Amazon Sales Data 2019:

- **Order ID** – The Order ID is the unique ID that is assigned to each order placed by the customer. (NOTE: If an order contains more than one item, then their order ID’s will be same)
- **Product** – This contains the name of the product ordered.
- **Quantity Ordered** – This is the total quantity of the product that was ordered in a single order.
- **Price Each** – This is the price of each product that is ordered.
- **Order Date** – This contains the date and time of the order placed.
- **Purchase Address** – This is the address where the order is to be delivered.

After cleaning the data, we added some more features to the data for detailed analysis. The columns added are:

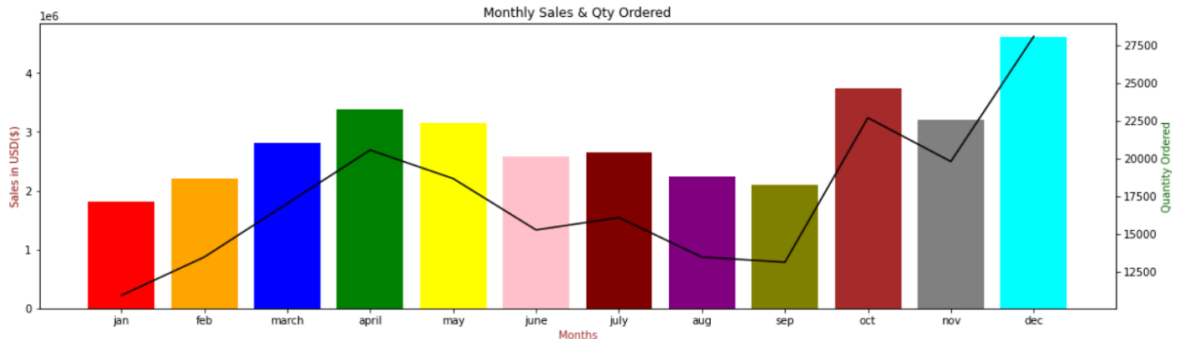
- **Sales** – This shows us the revenue that is generated from that item ordered. This column is made after multiplying “Quantity Ordered” column to “Price Each” column.
- **Month** – This column is extracted from the “Order Date” column. This shows the month in which the order is placed.
- **Day** – This column shows the Exact Date on which the order is placed. This column is also extracted from the “Order Date” column.
- **Time_hour** – This column tells the hour of the day in which the order is placed. This column is also extracted from the “Order Date” column.

The data was now ready for analysis after all these changes. The next stage was to do the analysis and determine the findings.

ANALYSIS

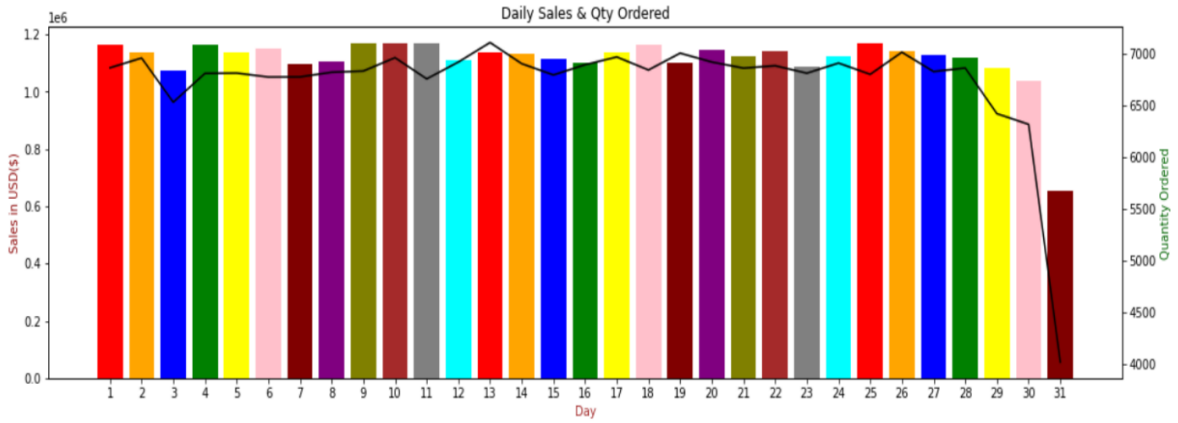
1) Analysis 1 – Advertising

- Here, for an annual sale campaign, we looked at which month produced the most sales over the course of the year. To do this, we first created a new data frame by grouping orders and sales by month. The data is then plotted on a graph.
  The analysis reads as follows:



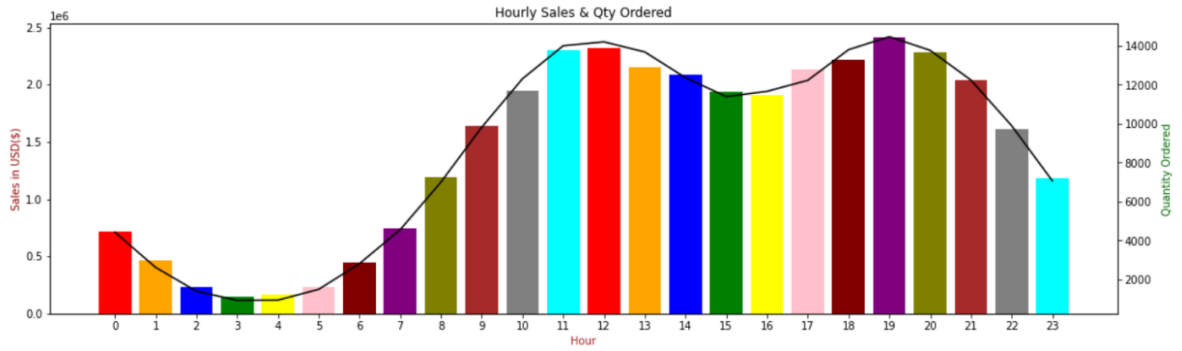
  Here the bar represents the sales of each month and Quantity Ordered is presented by line.
  Looking at the graph, it is clear that December is the ideal month for campaigns because it has the best sales graph.

- After choosing the month, we moved on to the second step, which involved choosing the best three days of the month for a sale promotion. We started by separating the sales and quantity ordered into different data frames and grouping them according to days. Next, we plot the data on a graph.
  The graph came like this:



  Here the bar represents the sales of each day and Quantity Ordered is presented by line.
  By examining the graph, we can clearly say that the 9th, 10th and 11th are the best 3 days for the campaign because it has the best sales graph. But 25th also have one of the highest bars in the graph and we know that 25th of December is “Christmas” so it is a more important day of the month to add into the campaign. So, it is decided to start sales from the 23rd of December to 25th of December.

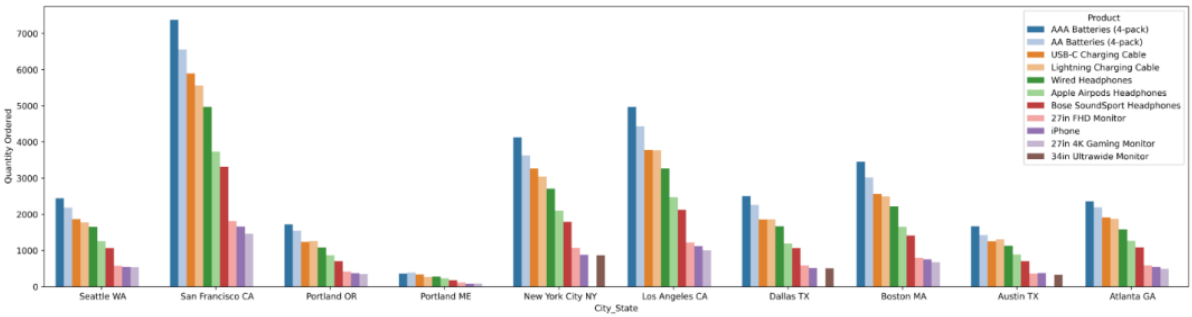
- After this we examined the sales of the 25th of December. We again made another data by grouping Sales and Quantity Ordered on the basis of Hours and then made a graph on the data.
  The graph came like this:



Here the bar represents the sales of each day and Quantity Ordered is presented by line.
After examining the graph, we came to think about the two most suitable time periods of the day to start the sale on. First time period is 9 AM to 1 PM and second is 5 PM to 9 PM. Both time periods have huge amount of traffic on the website and sales too. But we choose the second time period which is 5 PM to 9 PM because it has more traffic and in evening more people are likely to have free time to order things online.

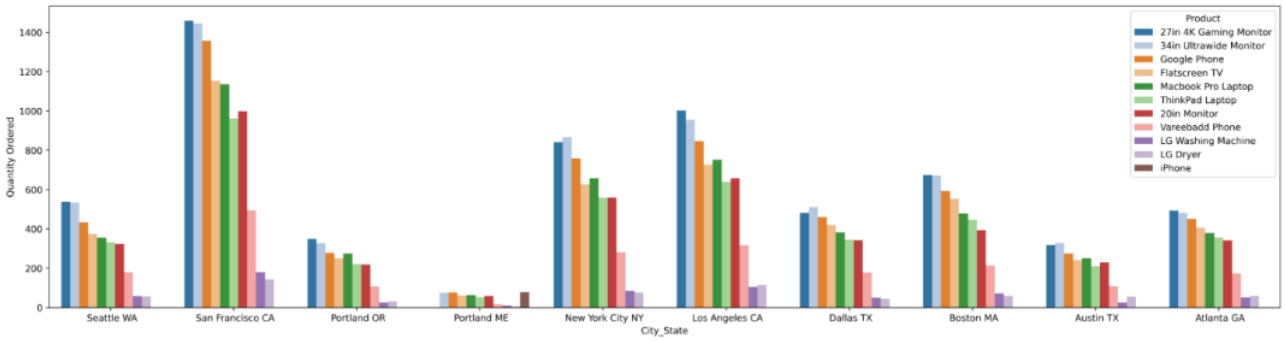
2) Analysis 2 – Warehousing (Supply Chain Management)

- The purpose of this analysis was to determine the product demand in each city and optimise Supply Chain Management. For this we made a different data frame by grouping Product and Quantity Ordered on the basis of City. The we made a graph of Quantity Ordered and City with products.
  Graph came like this:



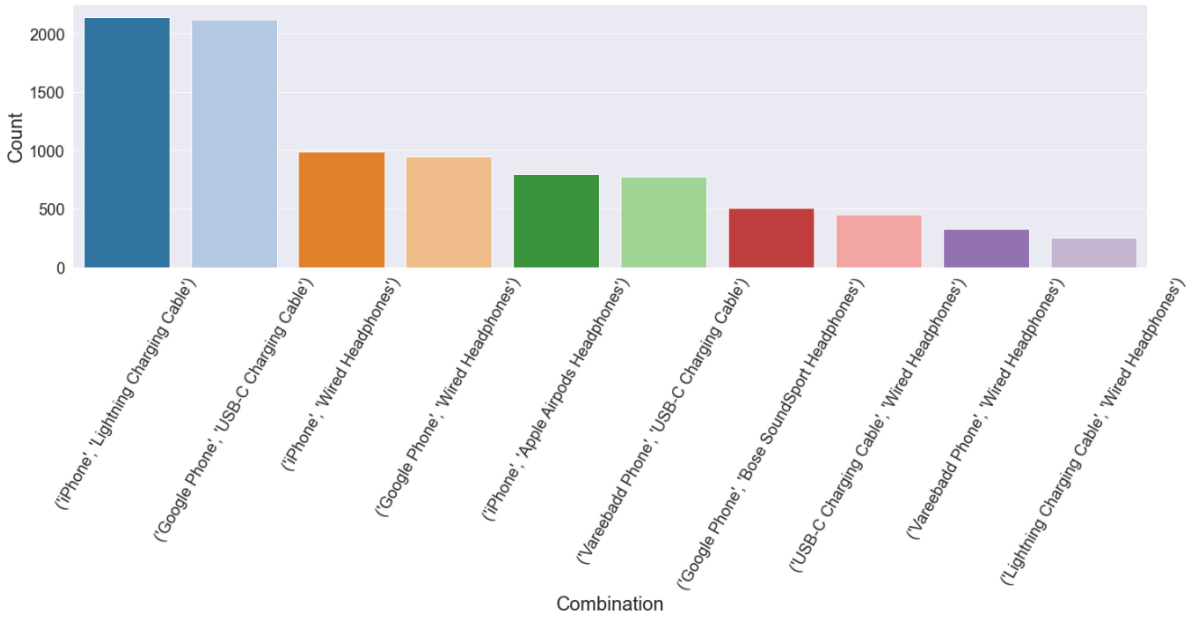
Here the City is on the x-axis and Quantity Ordered is on y-axis and bars represents the products that are ordered.

- We also made a graph of least ordered items in each state to know which city have less demand of which product.
  Graph:



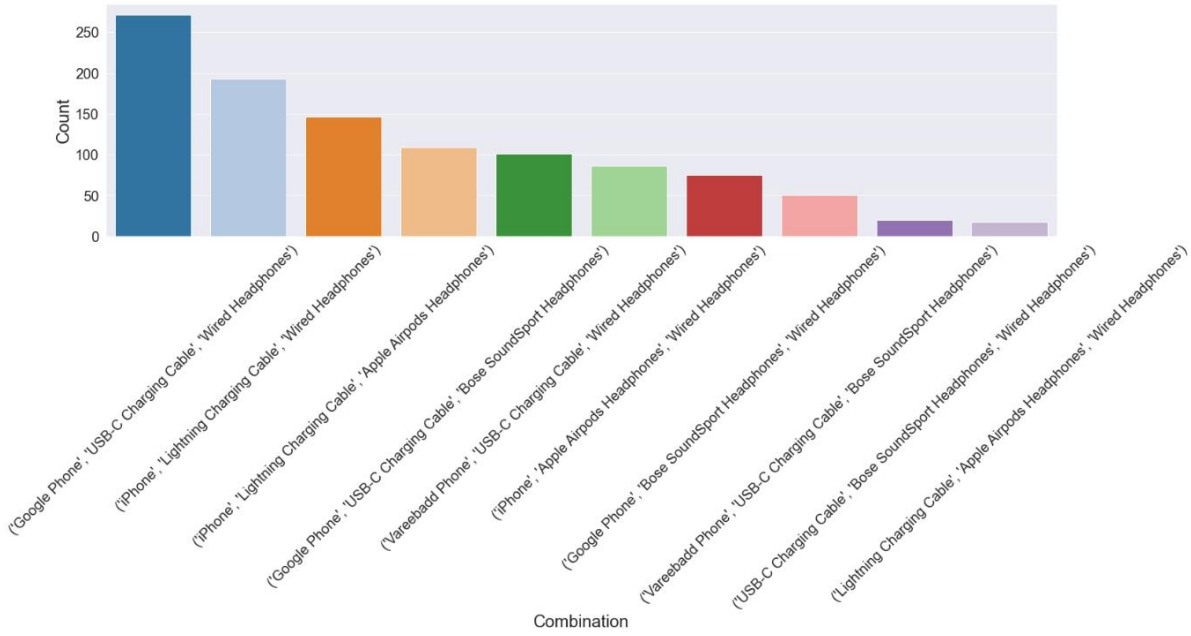
3) Analysis 3 – Recommendation association

- The aim of this analysis was to identify the products that are usually purchased in combination. For this we made a data frame by grouping Products on the basis of same Order ID. Then we moved on to plot a graph on the data.



Here we can see the top combinations of two products that are bought together mostly.

- We also made a graph of three products that are being ordered in combinations.



IMPROVE

According to the analysis we made, here are some points which can be done to improve sales of the company and improve customer satisfaction and retention:

1) According to Analysis 1, we can start the Sale Campaign from 23rd to 25th of December. This will help the company to acquire maximum number of purchases and will help the company to gain many more new customers also.
2) On the basis of Analysis 2 we can optimise the inventory of the warehouses bases on the demand and supply of the products. After optimising, warehouses will have the right number of units of any product which they should have according to the location of the warehouse. This will decrease the expenses of the company and will also help in delivering the orders fast which will put the customers in positive probability of purchasing products from the same company.
3) From Analysis 3, we can understand which products are sold together frequently and we can offer some discounts on the products in pair. We can also make a recommendation system which will recommend customers to products that they may be willing to purchase. This will lead to increased sales and revenue of the company.