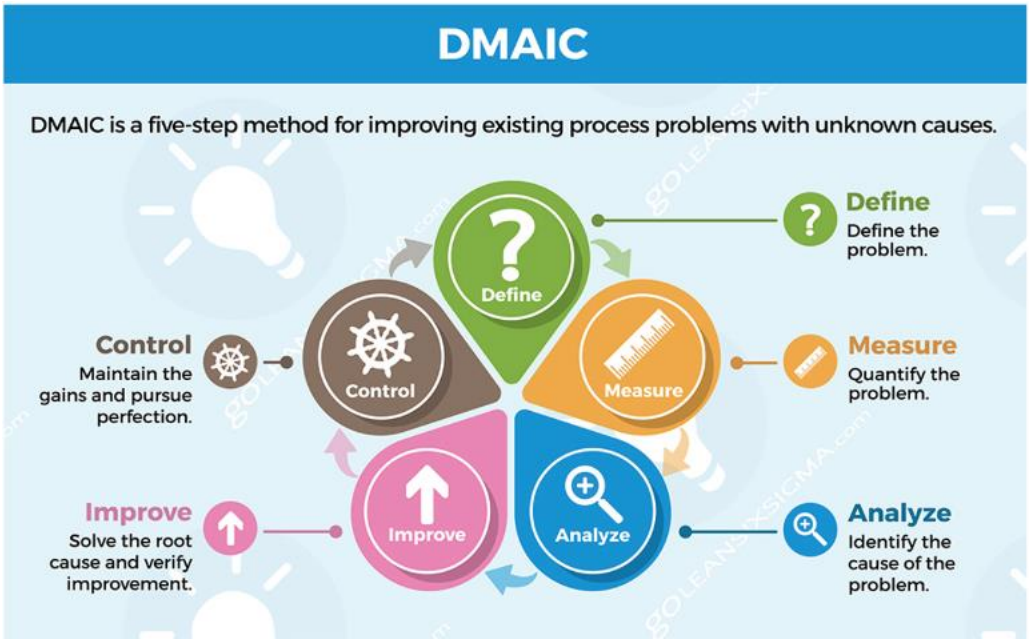
DMAIC Of Amazon Sale Prediction

What is DMAIC?

DMAIC is an acronym for Define, Measure, Analyze, Improve, and Control. It is a guideline for delivering sustainable and quantifiable results.



Define

As per DMAIC, the first and the most important step is the Define phase which is important for knowing areas of improvement. This project was created for resolving our three main problem statements:

1. Release date of our sale :- As an e-commerce website their first and the primary goal is to sell as many products as possible and increase their number of customers. To attain this goal rapidly e-commerce websites, launch sale. A sale is a campaign in which the customers are provided with huge discount and some great offers which benefits both that means the customer and seller. A sale can be launched either monthly or annually but in this case the website is launching a three-day annual sale which would benefit their business and the research aims to discover the three days that will be the most successful in the upcoming year so that customers can profit the most from the campaign and obtain the best discounts.

   KPI :- The Key Performance Indicators (KPI) are the critical success factors that are used to overcome the problem statement, so the KPI in this case are Price each, Sales, Quantity ordered, Month name, Day, Time.

CTQ (Critical to Quality): - The results of the study will be advantageous to the organization and its customers. If the promotion started on a day when business would be brisk, both customers and the company would benefit from the discount it offered.

2. Supply Chain Optimization :- Our second problem statement is to find out the sale of most and least demanded product state-wise. This will give us the idea of company warehousing that means the sale of products will tell us either we should increase or decrease the warehousing of product in that particular region. It will also help us by saving the cost of a warehouse of two different states as if the demand is concurrent for both the regions, there is no need to build another warehouse. One is enough to supply all the needs.

   KPI :- The KPI in this case are City States, Product, Quantity Ordered

   CTQ :- This study's results would be advantageous to the company because it would give them idea to reduce and manage its warehousing costs. This study would be beneficial to customers as well because it would result in faster and more accurate delivery.

3. Recommendation and Association :- Our third problem statements is to provide the customers suggestions of the product which are often purchased along the product which customer is actually purchasing.

   KPI :- The KPI in this case are Order Id, Product, Quantity Ordered

   CTQ :- Customers stand to benefit most from this analysis since they will get better advice on what to buy and how to use the products they have selected. Sales would surely rise, which would be good for the business as well.

Measure

Amazon is one of the world's biggest E-commerce websites which deals with a humongous amount of the data. The amazon sales 2019 dataset on which we are doing our analysis is a mere part of that big data. The dataset was provided to us by our institution Asian Academy of Film and television. The dataset contains these following columns:-

- Order Id : The unique Id provided to the purchase a customer made from the website. The order id is assigned on the basis of purchase, if a customer purchases multiple items at a single time their order id would be same.
- Product : The name of the product which had been purchased by the customers.
- Quantity ordered : This column tell us the number of units of the particular product which had been purchased
- Price each : Price of each product

- Order date : date on which product was ordered
- Purchase address : address of the customer who purchased a product from the website

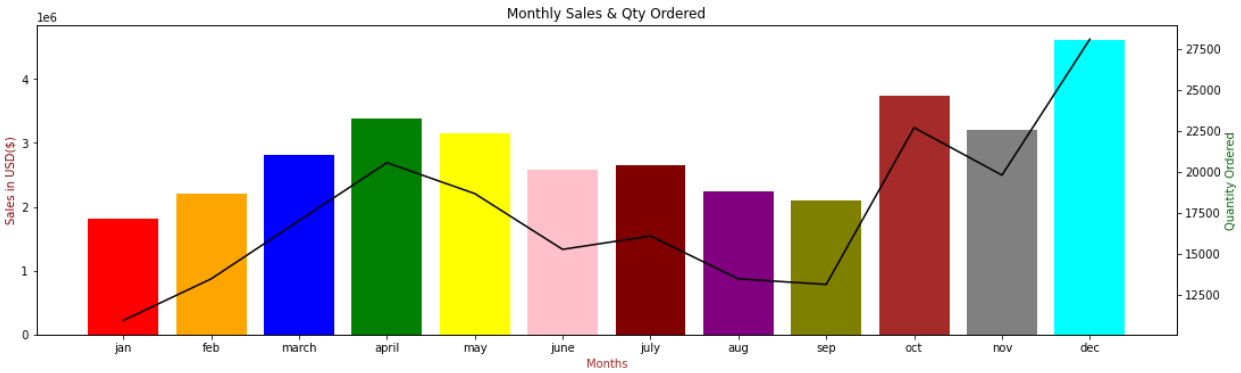
These were the default columns available in our dataset. During our analysis we created and added few more columns to the dataset which were essential for our project. The columns that we created were :-

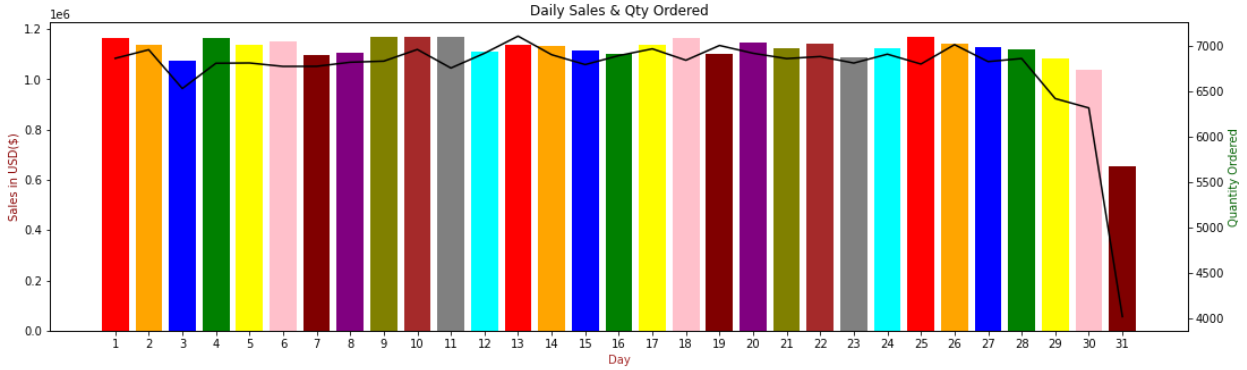
- Sales : This column tells us the amount of revenue a particular order has given. It was calculated by multiplying the values of quantity ordered and price each.
- Month: The name of the month on which the order was placed. It was extracted from order date column
- Day : The day of the month on which order was placed. It was also extracted from order date column
- Hour : The time of the day at which the orders are placed. It was also extracted from order date column.

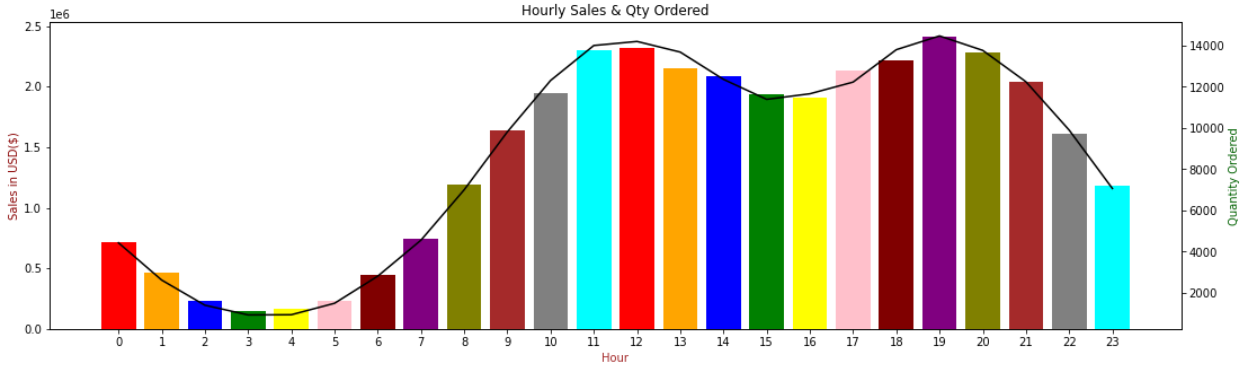
On basis of these following columns we are going to do our analysis

Analyze

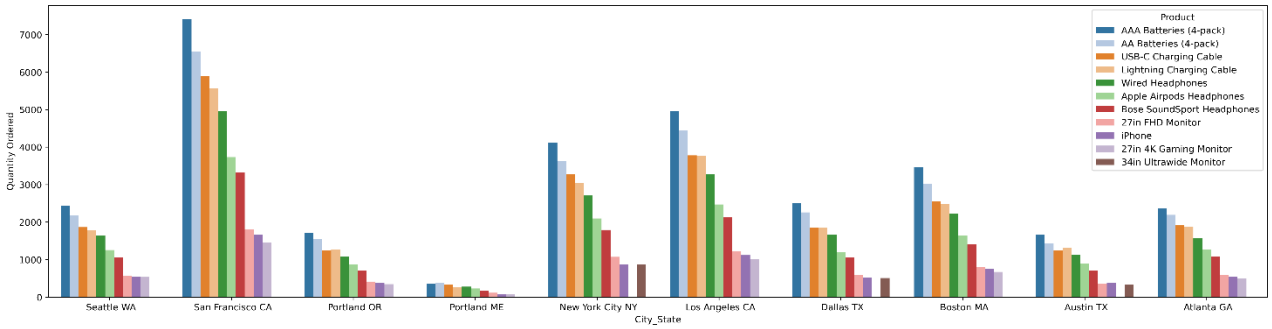
1. Finding the most appropriate month, day, and time for a three-day annual Sale was our first problem statement. We have created three graphs to show the month, day, and hour with the highest sales and quantity ordered. The graphs include the relevant feature (Month, day, or hour) on the x-axis along with the sales (bars) and quantity ordered (line) information on the y-axis. The order of these three graphs is as follows:

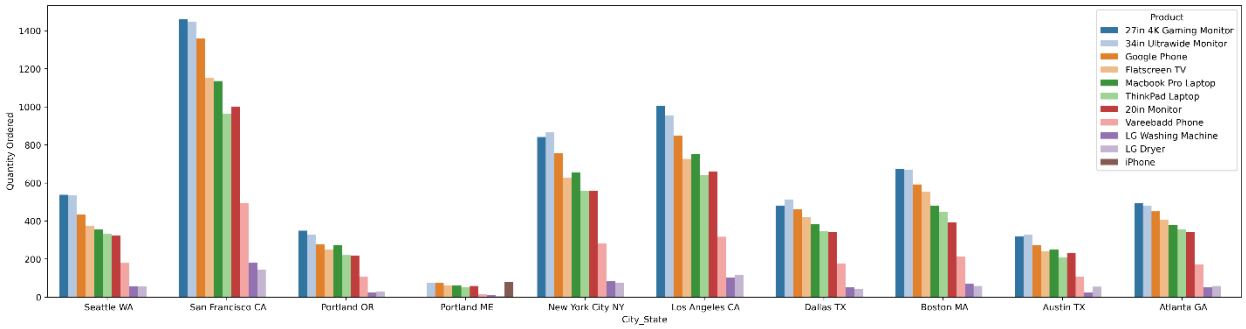




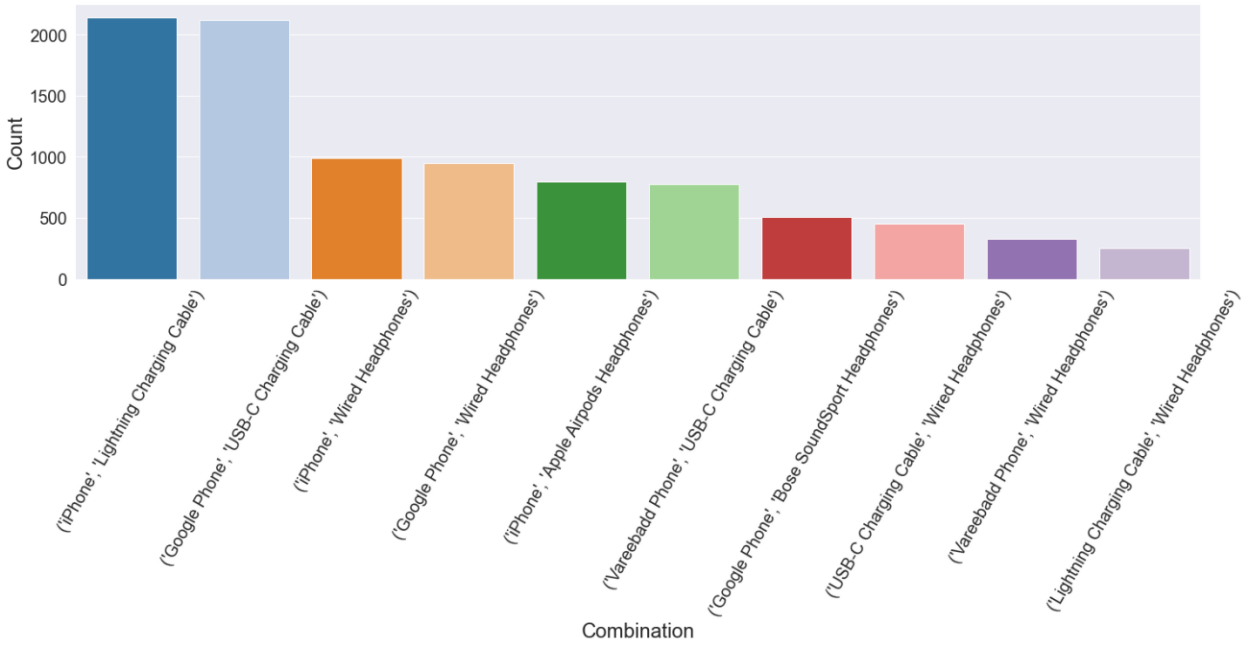


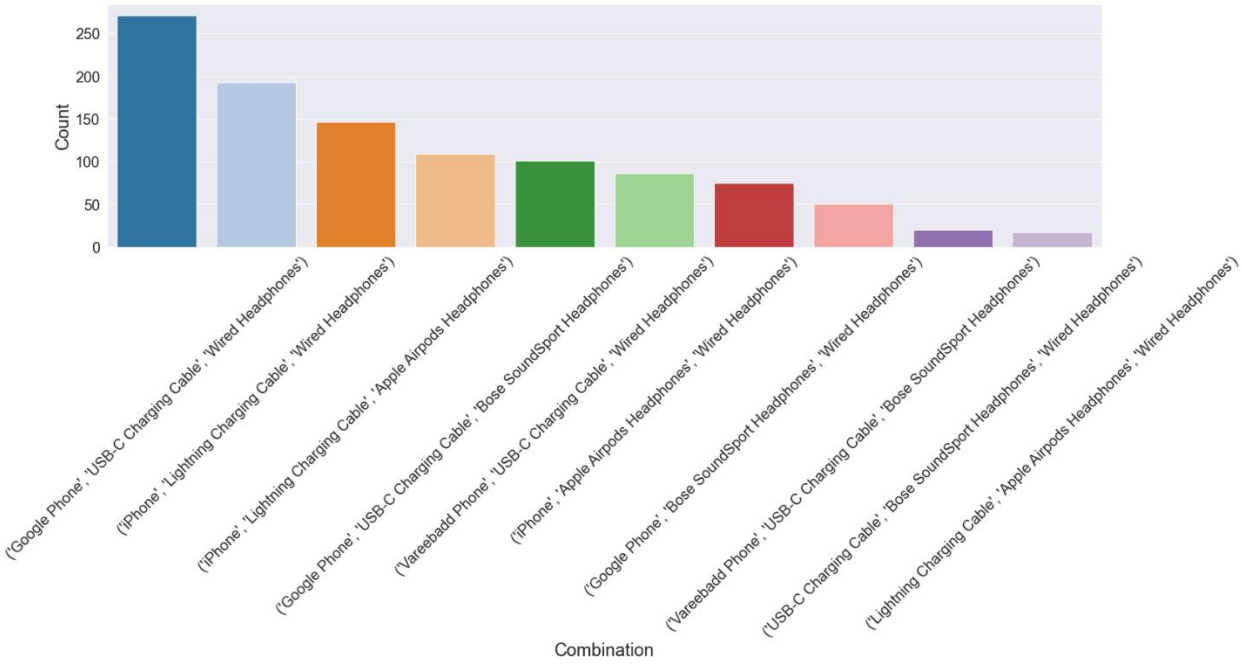
2. Addressing the management difficulty with warehousing was the goal of the second problem statement. This can be done by examining the quantity of each product that was separately ordered in each state. For this, we have created two graphs, one of which shows the number of products ordered in each state that is highest, and the other of which shows the amount that is lowest. Each stage is depicted on the x-axis, and the quantity requested is shown on the y-axis. The following two graphs display:





3. Our most recent problem statement concentrated on enhancing the consumer recommendation system. We were able to assemble a list of the most well-liked products by comparing the same order id for numerous orders. Two graphs have been made; the first displays the purchases of two things together, while the second displays the purchases of three products together. The following two graphs display:





Improve

- The first problem statement was to predict the best time, day, month for releasing our sale. According to our analysis report 7 PM as time, 25th as date and December as month are instances of highest sales and order. But we can't release our sale on 25th of December as the sale would be preceding Christmas so there would be less sale after 25th so we will release our sale before Christmas as at this time there is increase in orders and sales so we will release our sale on 23rd of December and the sale will remain three days till 25th. Since 7 PM is slight bit late in night, it is not recommended to release our sale at that time because more than half of the day has already passed and customers might be exhausted so we will release our sale at 10 AM as it would provide a great start to our sale

- The second analysis' graphs show that eleven products are ordered most and least frequently in each state. Given the strong demand in San Francisco, the business can build a sizable warehouse there with a lot of the goods with greater demand (as seen in the first graph) and a lot of the product with lower demand (as shown in the second graph). Furthermore, the business can set up a consolidated warehouse for multiple states based on their locations because Portland has very low demand. The demand graph would help us to manage the quantity of products bases upon the demand. As greater the demand we must increase our supply and store more units of that product and as lower is the demand we should reduce the quantity of our products which would save us money.

- Customers that purchase a "iPhone" or "Google Phone" also purchase a "USB-C Charging Cable," "Lightning Charging Cable," "Wired Headphones," or "Apple AirPods Headphones," according to the third study graph, which is an intriguing result. We may therefore suggest all of these items to a customer if they purchase an iPhone or a Google phone.