

Marketing Campaign Report

Introduction

Marketing is crucial for the growth and sustainability of any business. Marketers can help build the company's brand, engage customers, grow revenue and increase sales. One of the key points for marketers is to know their customers and identify their needs by understanding the customer. Marketers can launch targeted ad campaigns that are tailored for their specific needs. In this case study, we have been tasked to create a targeted ad campaign by dividing the customers of the "X" company into distinctive groups.

Data Preprocessing and Exploratory Data Analysis

The company has extensive data on their customers for the duration of 2.5 years and it has these different features:

- **ORDERNUMBER:** Identification of order placed.
- **QUANTITYORDERED:** Number of items ordered.
- **PRICEEACH:** Price of each item.
- **ORDERDATE:** Date in which order is placed.
- **QTRID:** Quarter in which order is placed.
- **STATUS:** Status of the order.
- **MonthID:** Month in which order is placed.
- **YearID:** Year in which order is placed.
- **PRODUCTLINE:** Product Category.
- **CUSTOMMENAME:** Name of the customer.
- **PHONE:** Phone number.
- **ADDRESSLINE1:** Address to be shipped to.
- **ADDRESSLINE2:** Address to be shipped to.
- **CITY:** State in which customer resides.
- **POSTALCODE:** Postal code in which customer resides.
- **COUNTRY:** Country in which customer resides.
- **TERRITORY:** Territory in which customer resides.
- **DEALSIZE:** Contact person's first name.
- **CONTACTFRISTNAME:** Address to be shipped to.
- **CONTACTLASTNAME:** Contact person's last name.

Based on the client's needs, there's a lot of features that we can get rid of since they are not useful and they can for sure lower our model's performance. These features are:

- **ORDERNUMBER**
- **CUSTOMMERNAME**
- **PHONE**
- **ADDRESSLINE1**
- **ADDRESSLINE2**
- **CITY**
- **POSTALCODE**
- **TERRITORY**
- **CONTACTFRISTNAME**
- **CONTACTLASTNAME**

After dropping all these unnecessary columns, we ended up with this dataset.

	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	ORDERDATE	STATUS	QTR_ID	MONTH_ID	YEAR_ID	PRODUCTLINE	MSRP	PRODUCTCODE	COUNTRY	DEALSIZE
0	30	95.70	2	2871.00	2003-02-24	Shipped	1	2	2003	Motorcycles	95	S10_1678	USA	Small
1	34	81.35	5	2765.90	2003-05-07	Shipped	2	5	2003	Motorcycles	95	S10_1678	France	Small
2	41	94.74	2	3884.34	2003-07-01	Shipped	3	7	2003	Motorcycles	95	S10_1678	France	Medium
3	45	83.26	6	3746.70	2003-08-25	Shipped	3	8	2003	Motorcycles	95	S10_1678	USA	Medium
4	49	100.00	14	5205.27	2003-10-10	Shipped	4	10	2003	Motorcycles	95	S10_1678	USA	Medium

Figure 1: Dataset.

In order to solve our problem which is trying to divide the customers of the company, we first need to try and understand our data very well. For that analysing and visualizing data is a necessity. So i used Tableau to create different Dashboards that can help me figure out what we need to do later on.

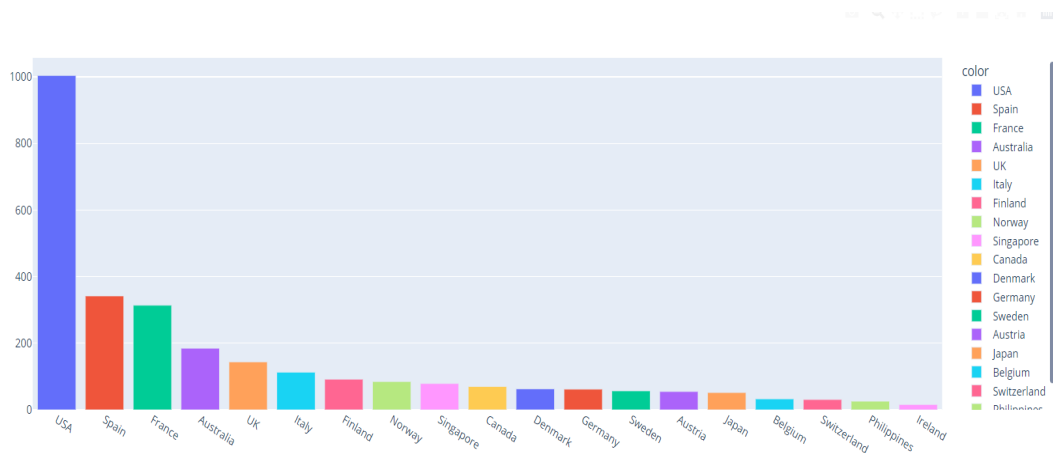


Figure 2: Sales for each country.

In the figure 2, we can see that USA has the most sales among all countries, we can also see that Europe plays a huge role when it comes to sales.

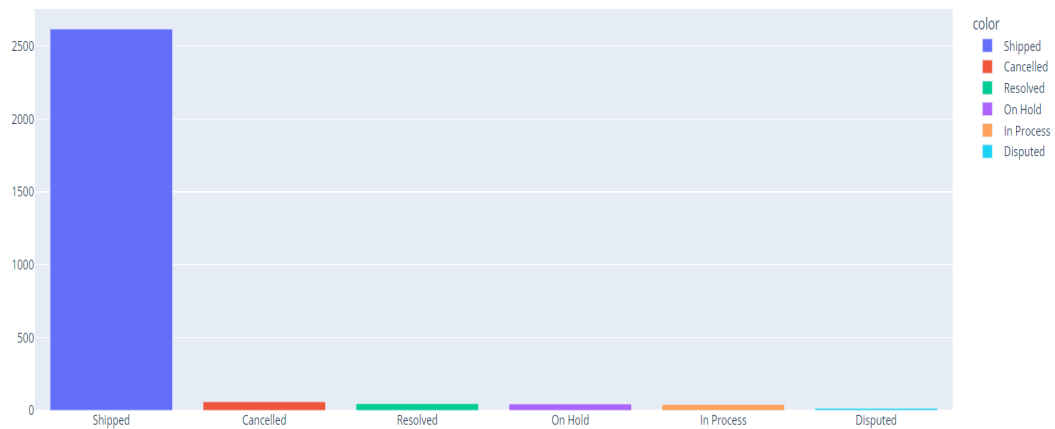


Figure 3: Status Distribution.

In the figure 3, we can see that our data is super unbalanced and it will have a huge impact later on that can lower the performance of our model. In this case we can't get more data and we also can't delete some rows of the data so the better option here is to drop this column.

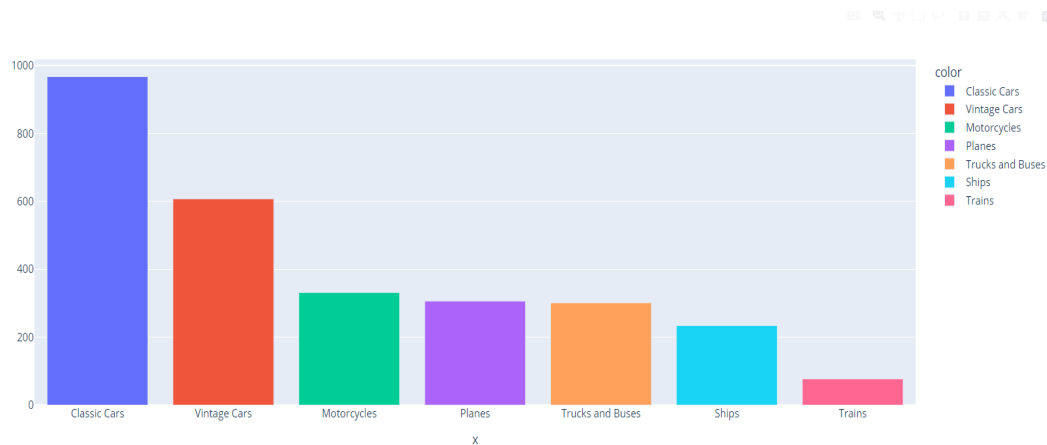


Figure 4: Products Distribution.

In the figure 4, we can tell that cars are the most sold products, rather than that, everything seems cool and fine when it comes to this column so no major changes needed.

Sales

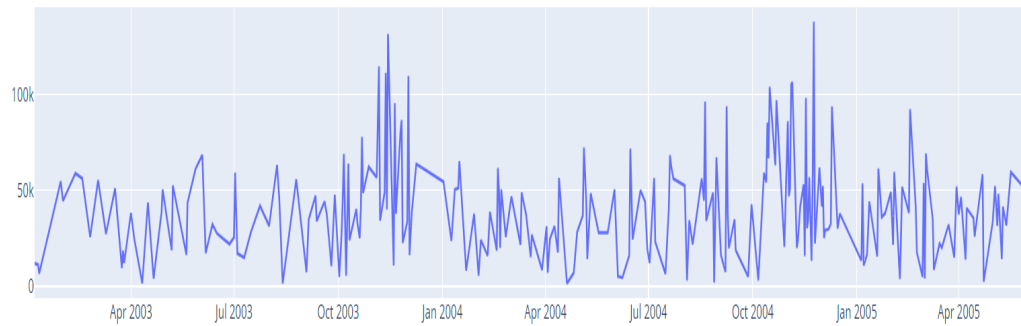


Figure 5: Sales overtime.

In the figure 5, we can tell that November has the highest sales each year, this might be a huge information since campaigns can be set around this month.



Figure 6: Correlation Matrix.

In the figure 6, We used the correlation matrix to see if there's any other information i can collect and i find out that the QTRID and MONTHID features are correlated so i can drop definitely drop the QTRID column.

K-Means Algorithm

K-means is an unsupervised learning algorithm that works by grouping some data points together in an unsupervised fashion. The algorithm groups observations with similar attribute values together by measuring the Euclidean distance between points.

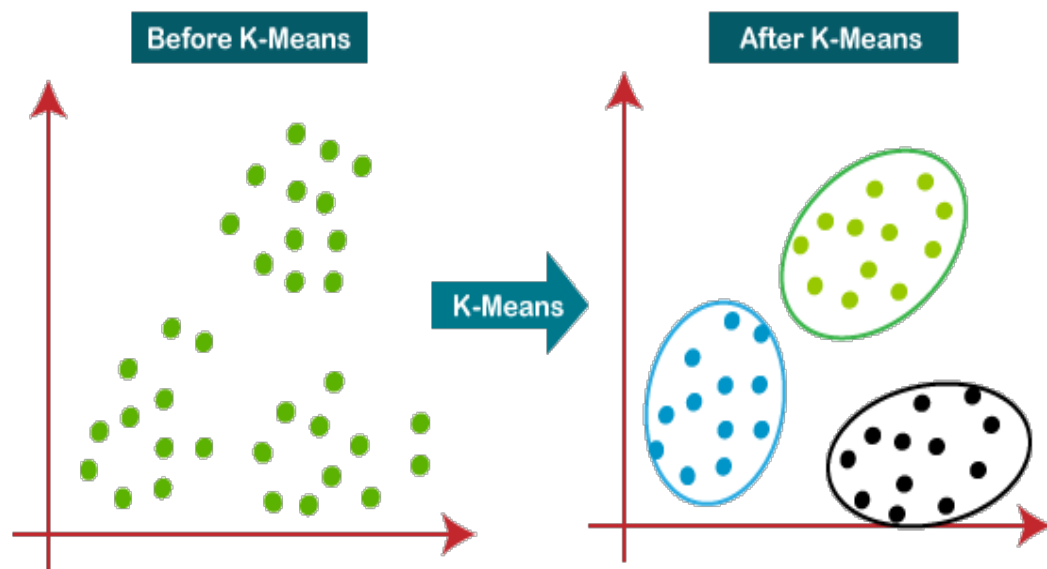


Figure 7: KMeans Algorithm.

The steps of this algorithm are:

- K-Clusters(Centers) points are created at random locations.
- Assign each data point to the nearest center.
- The center points are moved to their means of their respective clusters.
- step 2 and 3 are repeated until no observation changes.

Dividing Customers into 2 groups

After studying the data, it turns out that it's optimal to divide the customers into 2 groups.

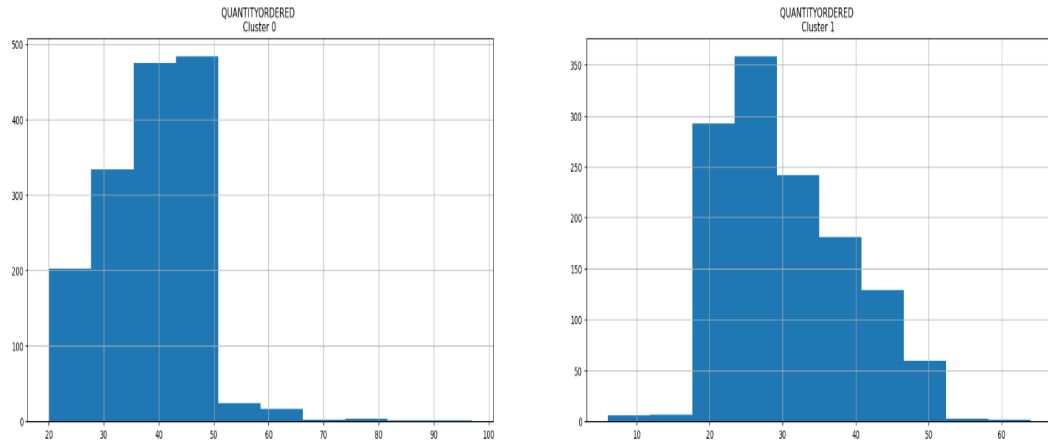


Figure 8: Quantity Ordered Distribution.

Based on the figure 8, we can see that the first group buys low quantities while the other group buys mid to high quantities.

Visualizing the Clusters

In order to visualize the cluster, We used PCA which is an unsupervised machine learning algorithm, it performs dimensional reductions while attempting at keeping the original information unchanged. It works by trying to find a new set of features called components which are composites of the uncorrelated given input features.

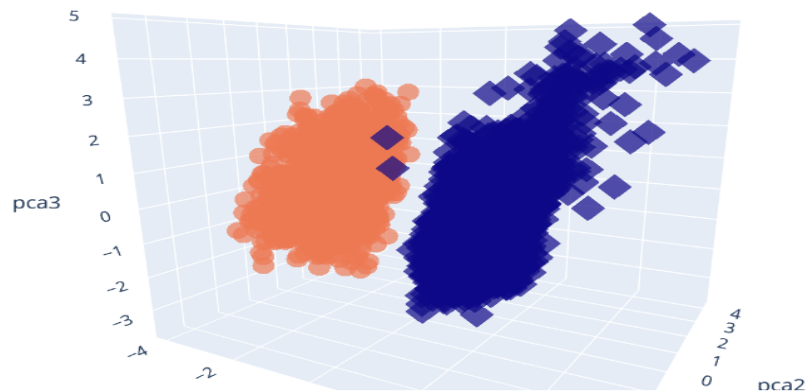


Figure 9: Visualizing the Clusters.