

Introduction:

Companies are storing vast amounts of data about the customers and their business, which is further used for analytical purposes and gaining valuable insights out of it. These insights can help the business or company to formula effective strategies for its growth.

This data analysis assignment is more focused to enhance marketing strategies and maximize the revenue generated.

The fictitious e-commerce company called ABC corporation is eagerly looking out for ways to optimize the strategies in order to develop its business. Below are the objectives and methodology used for analyzing ABC corporation.

Objectives:

The major objective of this analysis is to dig deeper on the data collected from the sources like web, social media, email campaigns, and customer database, which can help further to optimize marketing strategies for ABC corporation and can result in financial growth. The major analysis carried out in this process are,

- Prediction revenue from new email campaigns by using the past data.
- Segmentation of their customers by their transaction details
- Formulating A/B testing to enhance the new campaign performance.

Methodology:

As the ABC corporation is a fictional company, the methodology for the analysis involved,

1. **Data Creation** – Created four types of datasets using R
2. **Data preparation** – The created datasets were pre-processed in R to make it analytical ready.
3. **Exploratory analysis** – Several data insights were drawn by exploring data using R.
4. **Data Visualization** – Several insights drawn from the previous steps were visualized in R using ggplot2 package.
5. **Predictive modeling** - We used predictive modeling techniques to predict revenue generation for any new email campaigns by ABC Corporation.
6. **A/B testing** - We used A/B testing to compare the email campaigns to identify the similarity between.
7. **Visual reports** - Used Shiny package from R to create reports

Data collection:

The ABC Corporation is a fictitious company, and therefore I used R to create four datasets. We are creating the data sets for the duration of six months starting March 1st to August 31st 2024. Below are the details about the datasets,

Web_Souce_data:

- **Date:** The date on which the data about the website is collected (6 months).
- **Exit_rate:** It denotes in percentage terms that no. of people exiting the website after visiting just the home page.
- **Potential_views:** It denotes in percentage terms that no. of people bought something when they opened the website.

- **No_of_users:** It denotes the number of users used the website that day.
- **Device_type:** It denotes the type of device users are using to access the website.
- **Old_Vs_New:** It denotes whether the users accessed the website are new or old.

Social media data:

- **Websites:** It denotes the social media platforms of ABC Corporation.
- **No_of_followers:** It denotes the number of followers ABC corporation has on social media platforms.
- **No_of_posts:** It denotes the number of posts ABC corporation posted on these platforms.
- **No_of_interactions:** It denotes the number of interactions on the posts of ABC Corporation across platforms.
- **No_of_reach:** It denotes the number of users the post has reached.
- **No_of_shares:** It denotes the number of shares the users have made for ABC Corporation.

Email Campaign data:

- **Campaign no:** It denotes the Email campaign ids
- **No_of_sent_emails:** It denotes the number of emails sent as part of the email campaigns.
- **Generated_revenue:** It denotes the revenue generated by each campaign.
- **Response_rate:** It denotes the percentage of the users giving positive response to the emails.

- **Delivery_failure_rate:** It denotes the percentage of mails that failed to be delivered.
- **Ignorance_rate:** It denotes the percentage of the users who did not open the emails.

Customer transaction data:

- **Customer_id:** It denotes the id of the customer.
- **Sex:** It denotes the gender of the customer.
- **No_of_products:** It denotes the number of products bought by the user.
- **Amount_spent:** It denotes the total amount spent by the user.
- **Location:** It denotes the location of the user.
- **Payment method:** It denotes the user's mode of payment.

Data Preprocessing:

Below are the data preprocessing methods used to make the data analytical ready,

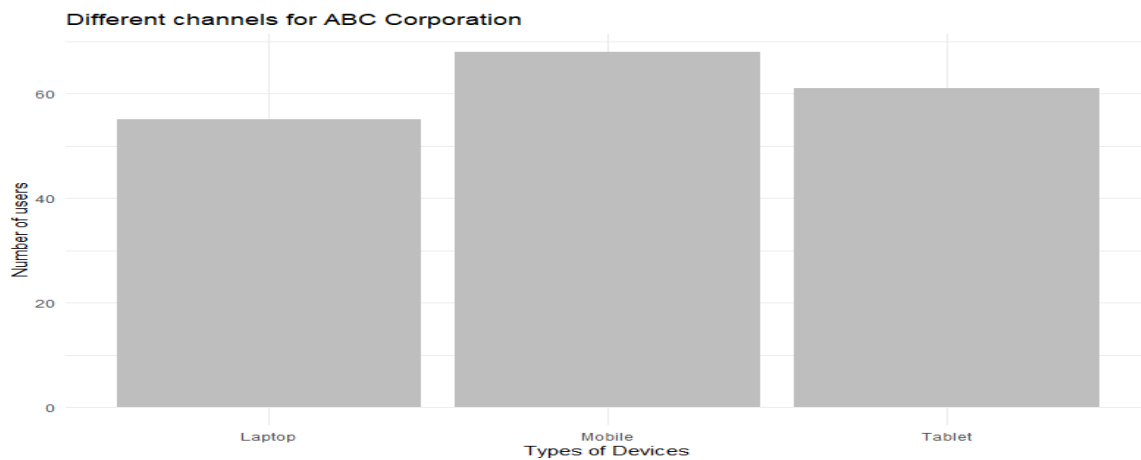
- **Data Cleansing:** Duplicate rows were removed in all the four data sets by using the function called Distinct. The column names were also standardized by using the function `tools::toTitleCase`
- **Datatype conversion:** Converted the datatypes of the columns which had categorical values to factors which will help models to work better.
- **Normalization and Scaling:** All the various rates in the datasets were converted to the percentage terms by multiplying the values by 100.
- **Missing values:** The missing values in the datasets were checked and filled in the by the appropriate values.

Exploratory Data Analysis (EDA):

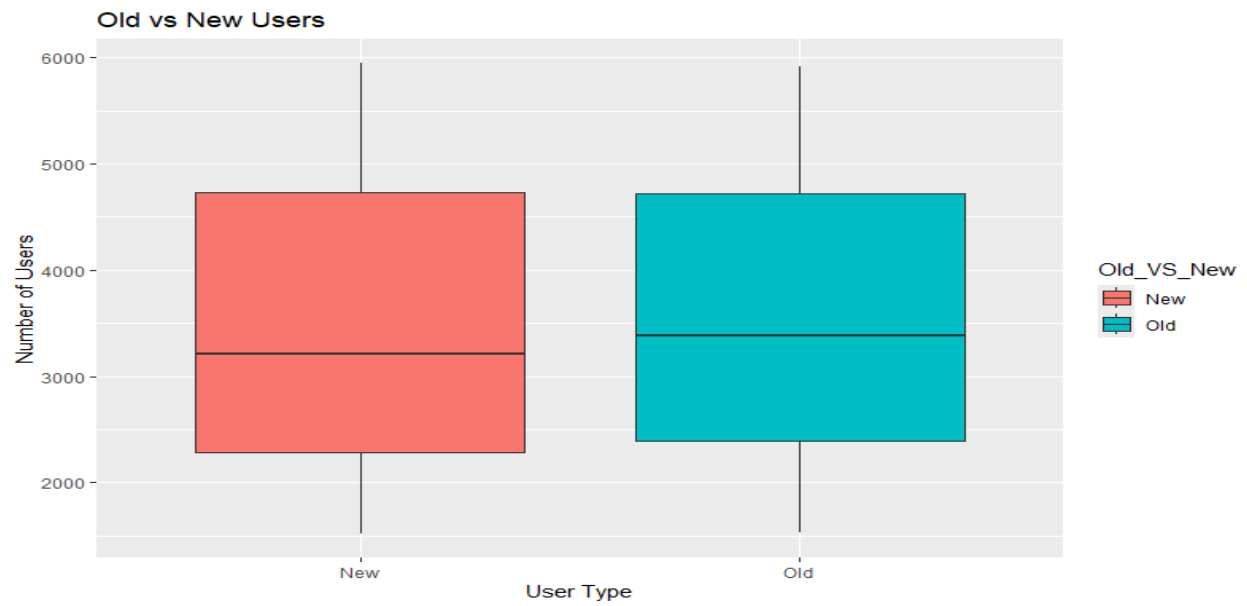
The extensive data analysis was carried out in all the datasets and various visualizations were created to gain insights about the data.

Web source data insights:

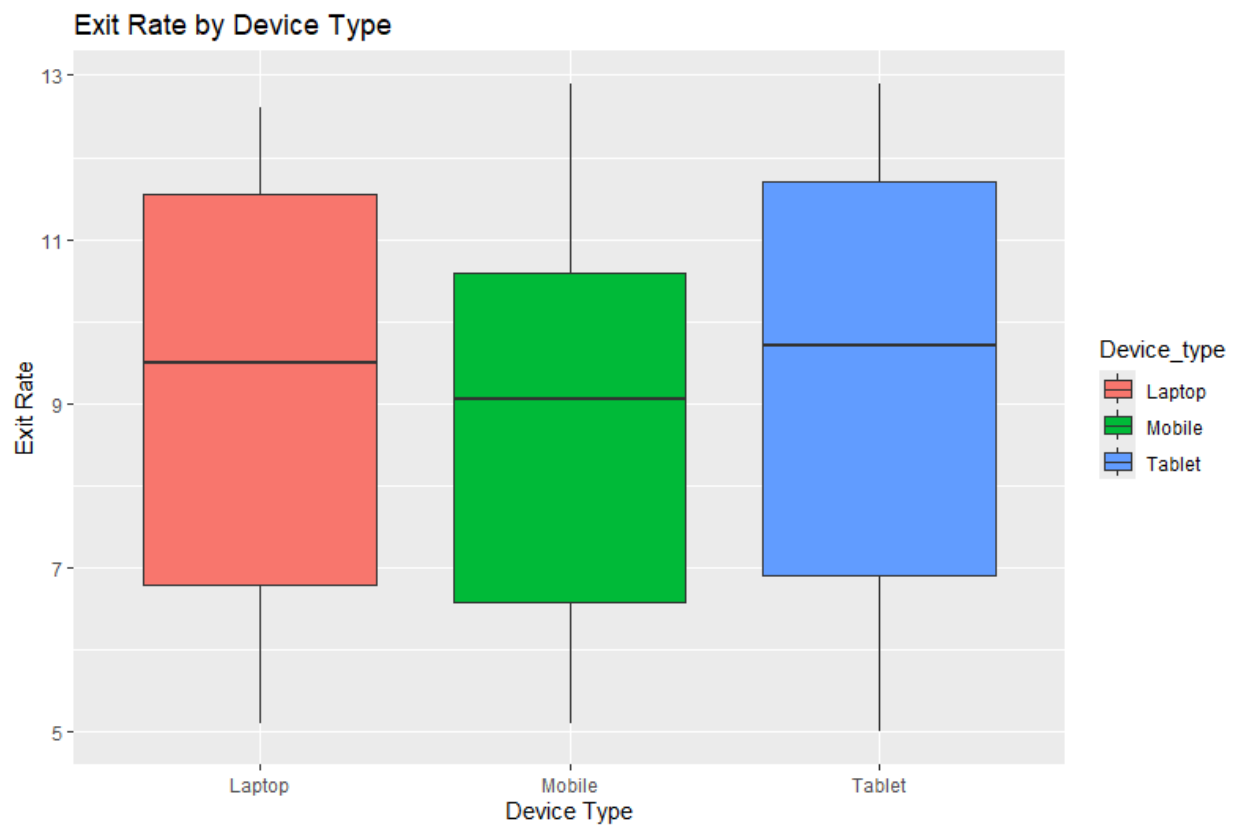
To understand the ABC Corporation better, we used ggplot2 to visualize and understand the nuances of the data,



Based on the above bar graph, it's understood that ABC Corporation has more mobile users and followed by Tablet and lastly Laptop users.

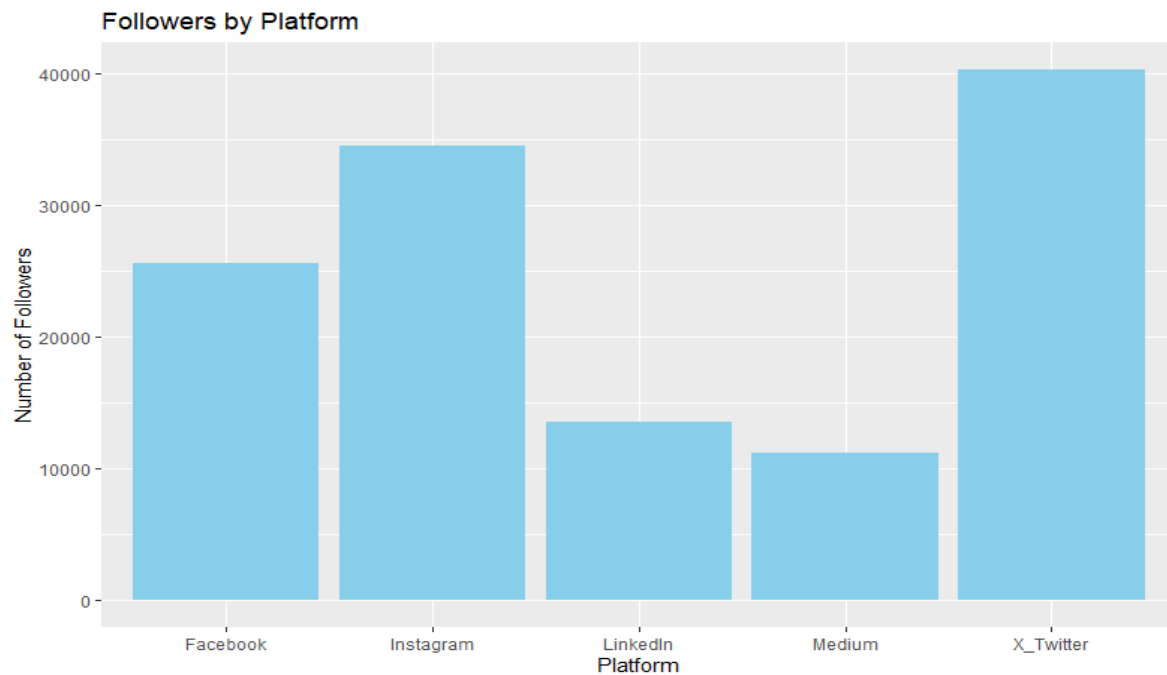


According to the above box plot, it's understood that we have almost equal number of old and new users in the six months data we had.



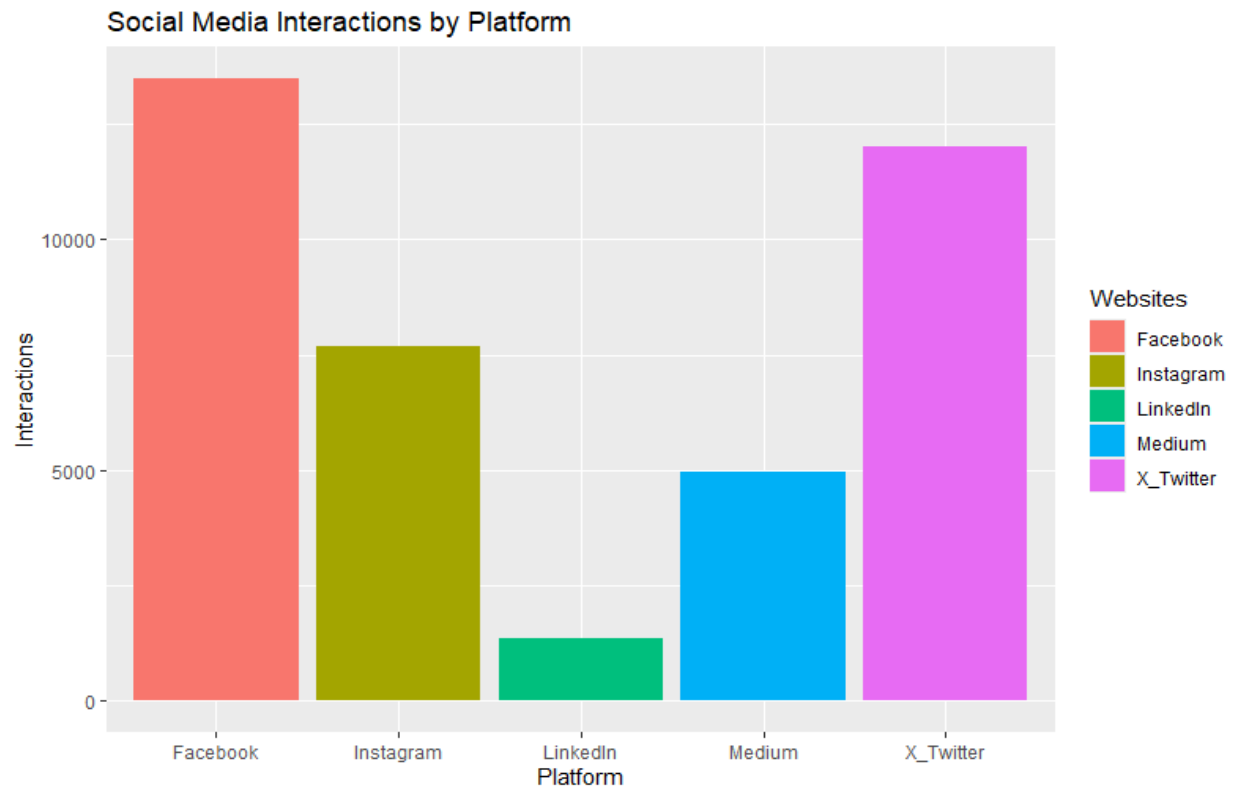
Based on the above visualization, it is evident that tablet users are exiting more frequently and followed by laptop and lastly the mobile users. It is interpreted that mobile users are spending more time on ABC Corporation websites than any other users.

Social media findings:



As per the above visualization, ABC Corporation has more followers on X, which is their official website. Secondly it is Instagram where they post about their new product offerings

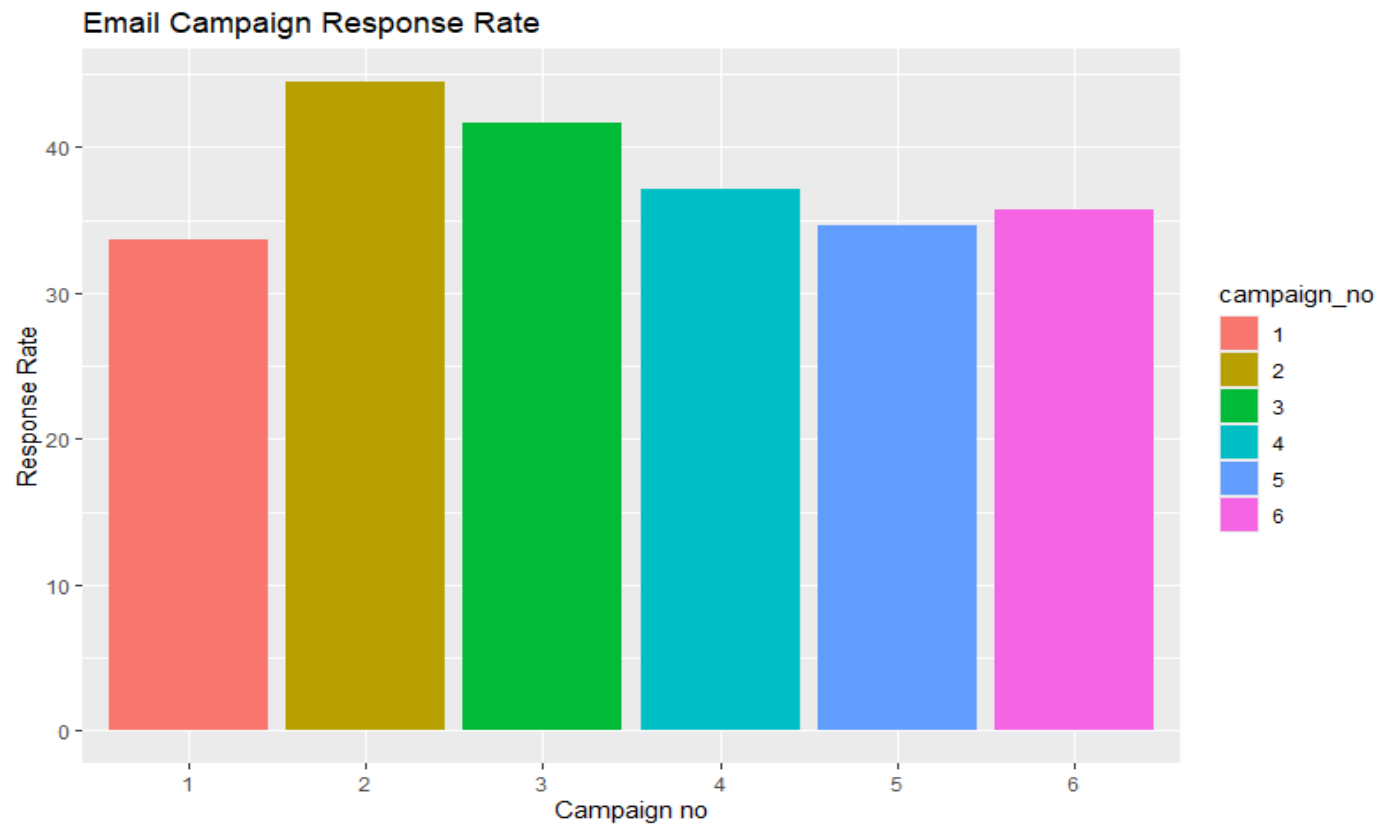
followed by Facebook, LinkedIn, and Medium.



According to the above graph plot, it is interesting to know that users are more engaged with ABC Corporation in Facebook through their likes, comments and shares. On the other hand, ABC Corporation's Linked posts have the least interaction amongst the users.

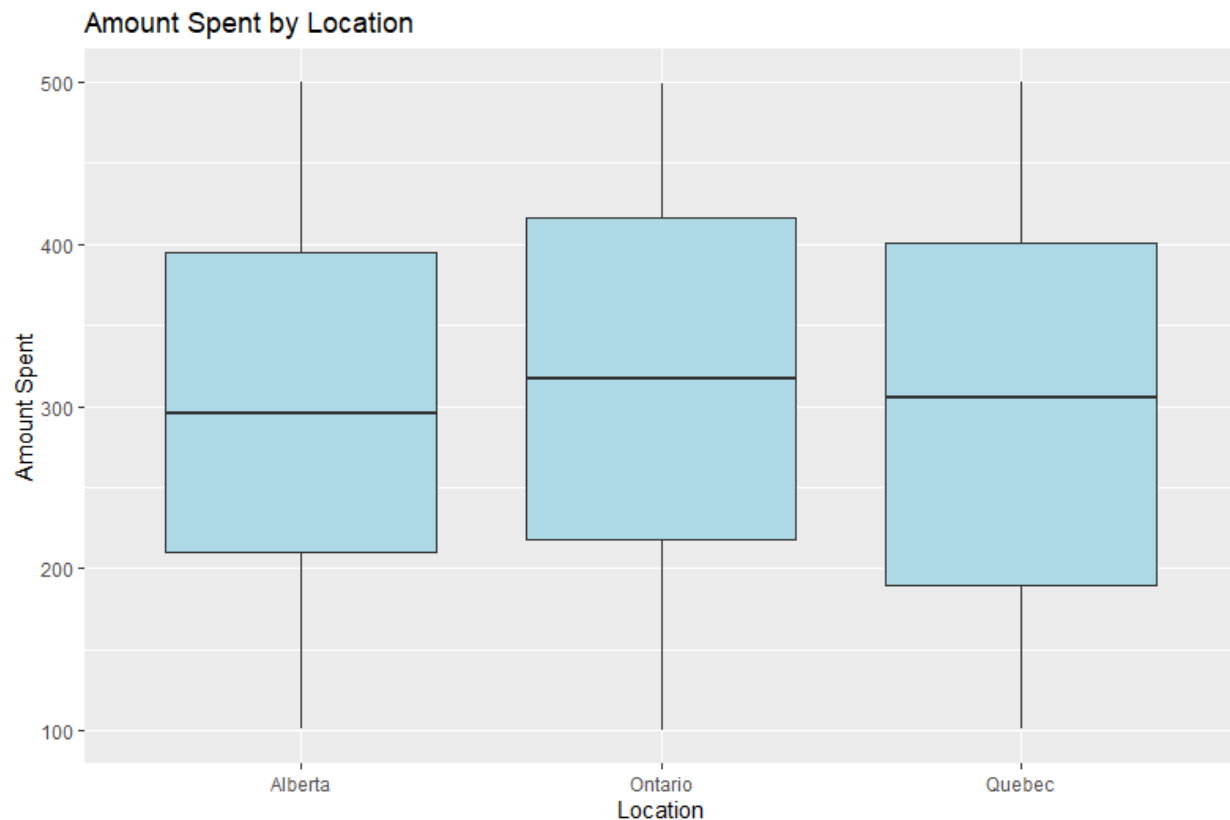
Email Campaign data insights:

ABC Corporation has the practice of sending discounts and new offering mails to their users. They have come across six email campaigns to offer new products and sale events.



The above visualization shows that Email campaign 2 has a higher response rate, (i.e) users have responded to the mails largely in the second email campaign. The first Email campaign has the least response rate than any other campaign.

Customer data insights:



Based on the above visualization, it is understood that ABC Corporation's Ontario users are more likely to spend higher when they make a purchase. Alberta users spend little less compared to other regions.

Customer Segmentation and Profiling:

ABC corporation's customer data is utilized for segmenting the customers based on features like No_of_products and Amount_spent. Furthermore, the three different clusters for the customers are identified as,

- Cluster 1: The users who spend less and buy fewer products in a purchase.

- Cluster 2: The users who spend on a medium scale with multiple products in a purchase.
- Cluster 3: The users who spend more and buy multiple products in a purchase.



Based on the above Kmeans clustering visualization, it is found out that,

Segment 1, who tends to spend less, mostly below \$300, and buy fewer products between 1 to 4.

Segment 2, who tends to buy more products typically between 5 to 8, but spends moderately between \$100 to \$400

Segment 3, who always spends more than \$300 to \$500 and buys a wide range of products.

ABC Corporation can utilize this segmentation and can formulate the marketing strategies for each group based on their purchasing practices.

Predictive Modeling and Forecasting:

I have used a linear regression model to predict the revenue generated from any new email campaigns based on the number of emails sent, response rate, and delivery failure rates.

ABC Corporation can input their desired number of mails that will be sent along with response rate and failure rates which can help them understand their revenue generation from the email campaign.

Model: Generated revenue ~ Emails sent + Response Rate + Delivery failure rate

I have used a random number for the inputs to the model, and it's predicted that the new email campaign has the potential to generate revenue of about \$13,832.

A/B testing and optimization:

I divided the six email campaigns as two sets to identify which campaign set is a winning set. To proceed with the analysis, I have used a statistical test (t-test) to compare the mean revenue between two sets.

The six campaigns are divided into two sets,

- Campaign set1 – Campaign 1, Campaign 2, and Campaign 3.
- Campaign set2 – Campaign 4, Campaign 5, and Campaign 6.

The p-value from the t-test will help us to know if there is a statistically significant difference in the revenue generated by the two campaigns. If the p-value is less than the significance level of 0.05 then we can conclude that there is significant difference between the two campaign sets.

T test result:

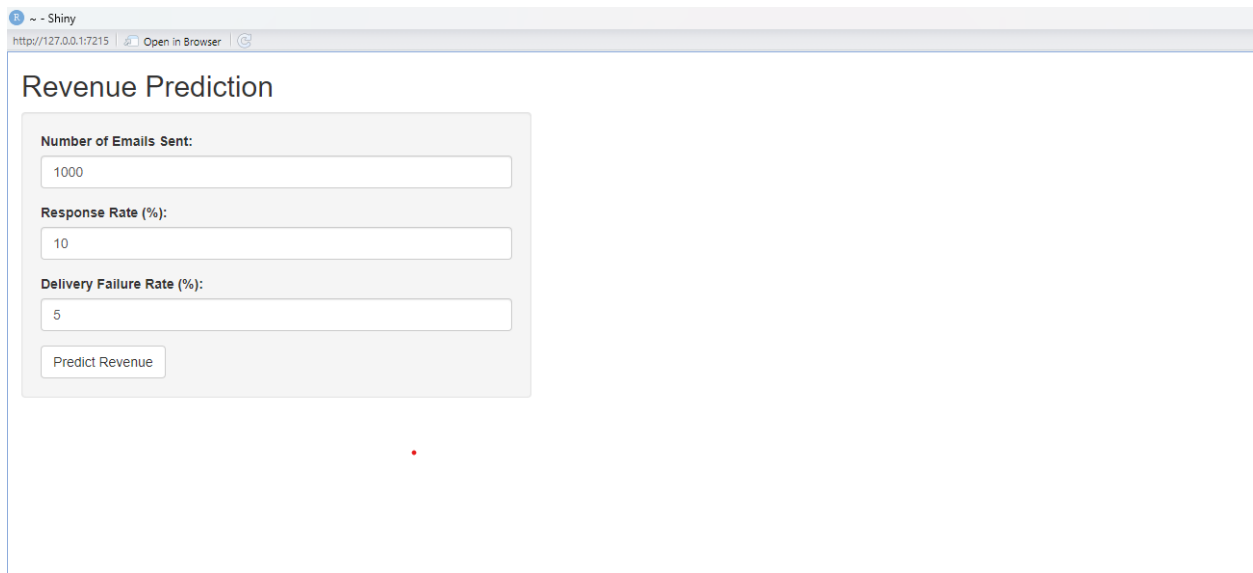
```
welch Two Sample t-test  
  
data: campaign_set1 and campaign_set2  
t = 0.16623, df = 3.8356, p-value = 0.8764  
alternative hypothesis: true difference in means is not equal to 0  
95 percent confidence interval:  
-6928.346  7795.012  
sample estimates:  
mean of x mean of y  
12266.67 11833.33
```

Based on the above results it is concluded the p-value is 0.8764 and therefore both the campaign sets are similar to each other and not significantly different. Hence, ABC Corporation has to come up with new marketing strategies for their next email campaign.

Reporting and Visualization:

I have used shiny package in R to create an UI interface for ABC Corporation to input requirements of a new email campaigns which can help us to know the revenue generated out of it.

Below is a snapshot of how the interface looks. This can be shared to any of the ABC corporation's stakeholders to understand the email campaign's revenue generation.



The screenshot shows a web browser window with a Shiny application. The browser's address bar displays 'http://127.0.0.1:7215' and includes an 'Open in Browser' button. The application's title bar reads '~ - Shiny'. The main content area is titled 'Revenue Prediction' and contains a form with three input fields: 'Number of Emails Sent:' with the value '1000', 'Response Rate (%):' with the value '10', and 'Delivery Failure Rate (%):' with the value '5'. Below these fields is a 'Predict Revenue' button. The background of the application is a light gray color.

The user of this interface should enter the details of the new email campaign such as Number of Emails sent, response rate, and Delivery failure rate in percentage terms. Once the information is entered, the user should press the Predict Revenue button to know the revenue generation prediction from the campaign. The user can use the interface to trial-and-error different inputs and can settle down on the desired revenue they want from the new email campaign.

Conclusion:

Based on the data analysis of ABC corporation, below are the key findings measured.

1. **Revenue Prediction model:** The linear regression model effectively calculates the revenue generation based on the factors inputted. This can help ABC Corporation to

forecast better and can help them to come up with enhanced and effective Email campaign strategies in the future.

2. **A/B testing:** A/B testing helped in understanding which email campaign sets were better in terms of revenue generation. The results show that both the email campaign sets are not significantly different and therefore ABC Corporation has to come up with effective strategies and not to follow the old method to achieve greater growth.
3. **Customer Segmentation:** With the use of clustering analysis, customers were divided based on purchasing behavior. This segmentation can help ABC corporation to formulate effective marketing strategies based on their user's spending habits.

Below are the few recommendations based on the above findings,

1. **Incline towards effective campaigns:** ABC Corporation should more focus on marketing campaigns that have lower delivery failure rates and higher response rates. This can help them to drive up the revenue.
2. **Marketing strategies based on customer segments:** Based on the customer segments provided to ABC Corporation, they should create marketing strategies pertaining to their buying pattern. Low spenders should be focused more on promoting cheaper products and high spenders should be marketed with high priced products frequently to improve the revenue generation.
3. **Social media strategies:** ABC Corporation should focus more on promoting their products in Facebook rather than any other social websites as they have more interactions compared to others.