

## **Summary**

**The following technical steps are used:-**

**1. Data Cleaning:**

- As a first step in cleaning the dataset, we decide to eliminate any redundant features or variables.
- The option "Select" has to be replaced with a null value because it did not provide us with much information, and the data set was mostly clean aside from a few null values.
- Removed the high null value proportion of above 40%.
- Determined how many distinct categories there were in each category column.

**2. Exploratory Data Analysis:**

- Conducted univariate analyses for categorical and continuous variables.
- Conducted a Bivariate Analysis concerning the Target variable.

**3. Dummy Variables:**

- For each of the categorical columns, dummy variables are produced.

**4. Scaling:**

- Scaled the data for continuous variables using standard scalar.

**5. RFE-Based Feature Selection:**

- By using RFE with provided 20 variables. It gives top 20 relevant variables.

**6. Model Building:**

- Subsequently, the variables with a VIF less than 5 and a p-value of 0.05 were retained, while the unnecessary characteristics were manually eliminated based on the VIF values and p-value.

**7. Test Set Prediction:**

- A confusion matrix was made. Subsequently, the accuracy, sensitivity, and specificity were determined by utilizing the ROC curve to determine the ideal cut-off value, which was approximately 80%.

**8. Conclusion :**

As we've seen, the factors that matter most to prospective purchasers are:

- The total time spend on the Website.
- Total number of visits.
- When Olark Chat, Google, Direct Traffic, and Organic Search were the major sources
- When the last activity was: a. SMS b. Olark chat conversation.