

Started to import all the require libraries .

1. Data reading
2. Data cleaning
 - Checking for the null values.
 - Checking for the object types are mentioned correctly.
 - Replacing the select with NAN
 - In this case study changing the select to NAN
 - Dropping the columns of no use with the analysis and the unique values=1 as well as the unique values count = no or rows of the dataset
 - Removing the columns with more than 35% of the null's
 - For the rest removing the null rows as there is sufficient number of columns for the analysis
 - No outliers are seen
3. EDA
 - Exploring on the categorical and the numeric variables separately
 - Drawing the count plots, histplots , heat maps
 - Checked how the conversion rate is with all the categorical variables
 - Got the correlation using the heat map this is details which parameters is much closely associated with conversion
4. Dummy Variables
 - Extracting the Dummies from the categorical variables for more clear analysis
 - Used minmax scaler to scale numeric variables
5. Train and test split
 - Split the data into 70% training and 30% test
6. Model building
 - Logistic Regression to build the model from sklearn
 - RFE for feature selection and selected top 15 features for the further analysis
 - Using the manual method to eliminate features until the below conditions satisfy for all the features
 - $VIF < 5$ and $P < 0.05$
7. Model Evaluation
 - Evaluating the build model by using confusion matrix, TP, FN, sensitivity, specificity, accuracy, ROC curve
 - 79% accuracy, sensitivity and specificity being 74% and 83%
 - Area under ROC curve is 0.86
 - Cut off as 0.41 using the sensitivity, specificity and accuracy
8. Prediction
 - Predicting using the test set
 - 75% accuracy, 83% sensitivity, 69% specificity
9. Precision and Recall
 - Got the cut off as 0.41
 - Precision as 73% and recall as 85% on test data.

Parameters found to be useful are as follows:

1. Total time spent on the website
2. Lead source is mostly rely on the direct traffic, google and organic searches
3. Last activity SMS and then the Emails
4. What is your current occupation - unemployed
5. Lead origin as API , Lead add form, landing page submission
6. Country - India
7. Specialization with marketing and banking and health care sectors