Summary

1. This case study required the outcome to be a 0 or a 1, ie. Will a lead be converted ot not, thus the same was carried out using Logistic Regression model as it was the best fit for the business problem.

2. It is seen that most of the leads are from India and belong to city Mumbai.

3. Several columns were found in which there is a level called 'Select' which basically means that the applicant had not selected the option for that particular column which is why it shows 'Select'. To get some useful information regarding the applicant it is very important that we make a compulsory selection for e.g. in columns like Customer occupation, Specialization, etc.

4. The high number of total visits & Total time spent on platform may increase the chances of lead to be converted.

5. The leads who converted previous, joined a course for Better Career Prospects, most of them having a Specialization in Finance Management. Leads from HR, Finance & marketing management specializations are highly probable that they will convert.

6. Getting in touch with the last notable Activity, making improvement in customer engagement through email & calls will help to convert leads. As the leads going through/reading their email have a high probability to convert, the case is the same with Sending a SMS which is also be a benefit.

7. It was also found the maximum converted leads are generated by Google, followed by direct traffic. Conversion Rate of reference and welingak website is high.

8. It was found that leads spending more time on the website are more likely to be converted.

9. We observed that there were a lot of variables present in the dataset. Thus, the best way to approach this was to select a small set of features from this pool of variables using RFE.

10. RFE implementation helped us pick 15 useful variables to build a perfect model for prediction.

11. Finally observing the p-values and VIF of all features and dropping the ones whose values were above the threshold, we conclude the following features were most important :

const
TotalVisits
Total Time Spent on Website
Lead Origin_Lead Add Form
Lead Source_Olark Chat
Lead Source_Welingak Website
Last Activity_Email Bounced
Last Activity_Email Opened
Last Activity_Other_Activity
Last Activity_SMS Sent
What is your current occupation_Working Professional
Last Notable Activity_Unreachable

12. Once we found the ROC is 0.87 and checked the accuracy, sensitivity, and specificity tradeoff, it helped us find the best cutoff value to train the model to predict conversions.

13. Finally after training the model, we tested the same on the test set, and it was found that the model predicted the conversion rate with a precision of 78.4% and a recall of 78.3 %, which seemed to be a good overall solution.