

Lead Scoring Case Study

Group Members -

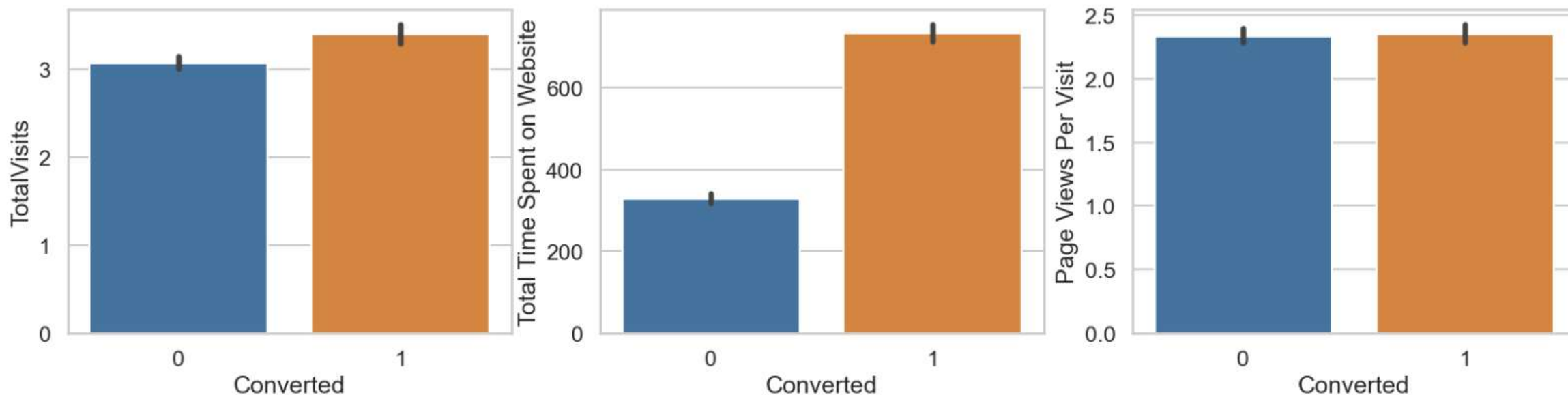
- 1) Aditi Shah
- 2) Bhimashankar Malbhage
- 3) Gurtek Singh

Solution Strategy and Methodology

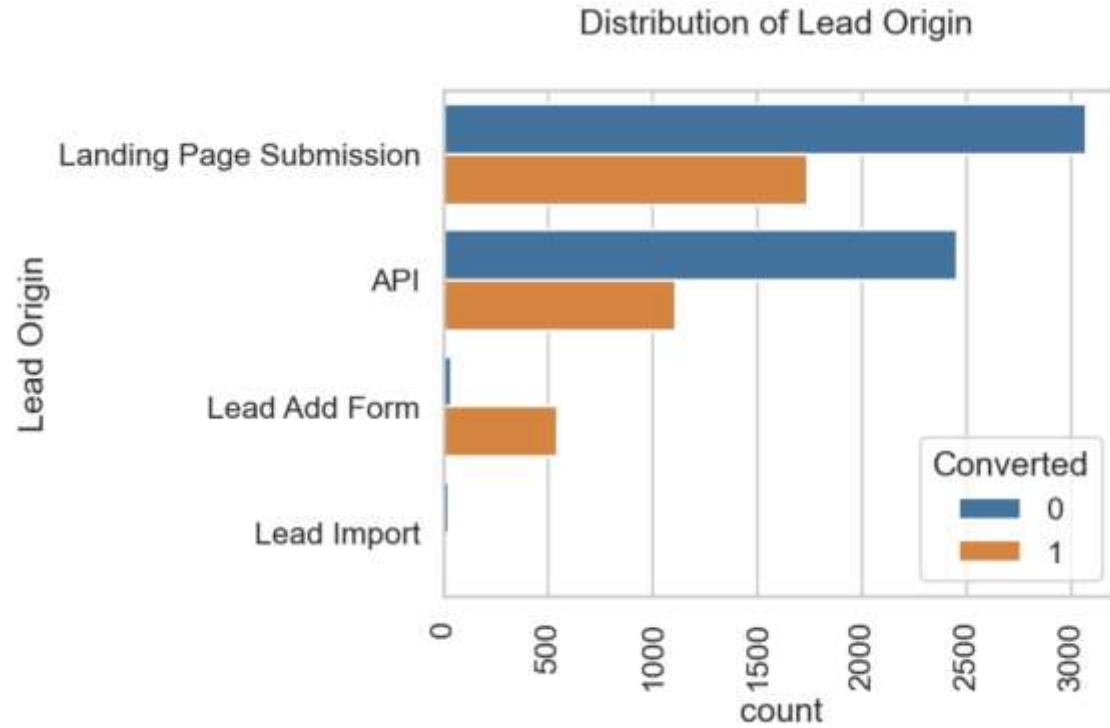
- Import the data for analysis
- Data Sanity Check
- Exploratory Data Analysis
- Data Preparation for Modelling
- Splitting the data into train and test dataset
- Building Logistic Regression Model
- Evaluating the model by using different metrics like Overall accuracy and Sensitivity and Specificity
- Finding the Optimal Cutoff
- Making Predictions on the Test Set

Exploratory Data Analysis

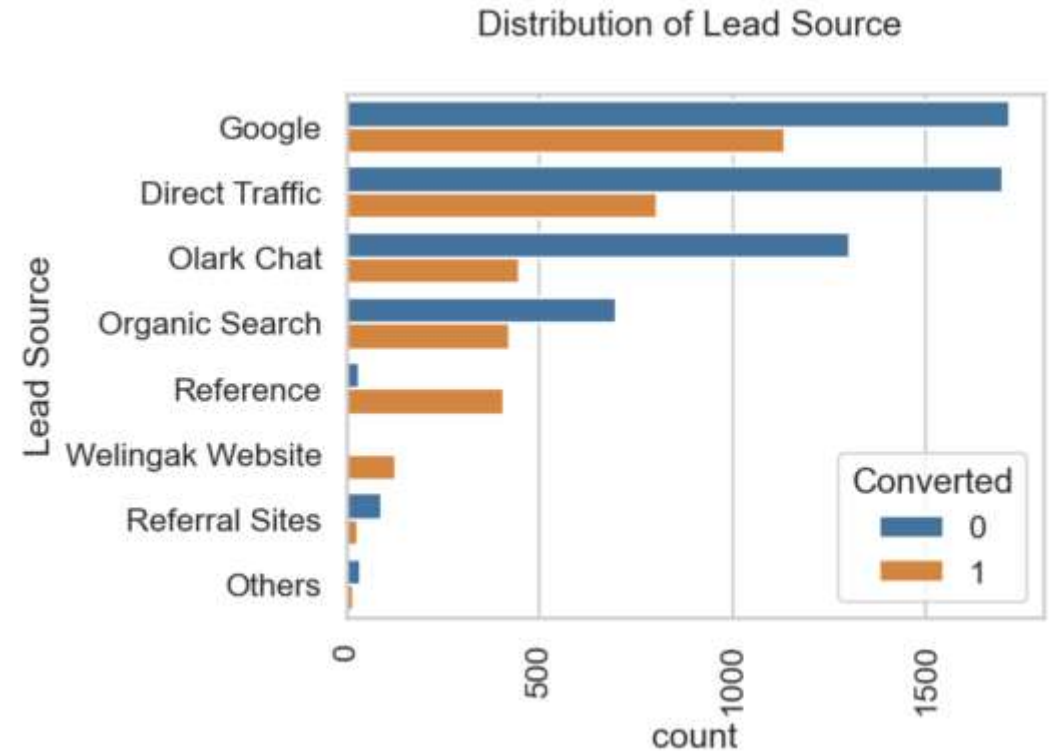
- The conversion rate were high for Total Visits more than 5
- The higher the Total Time Spent on Website, the more is the conversion
- Page Views Per Visit doesn't show a strong correlation with conversion



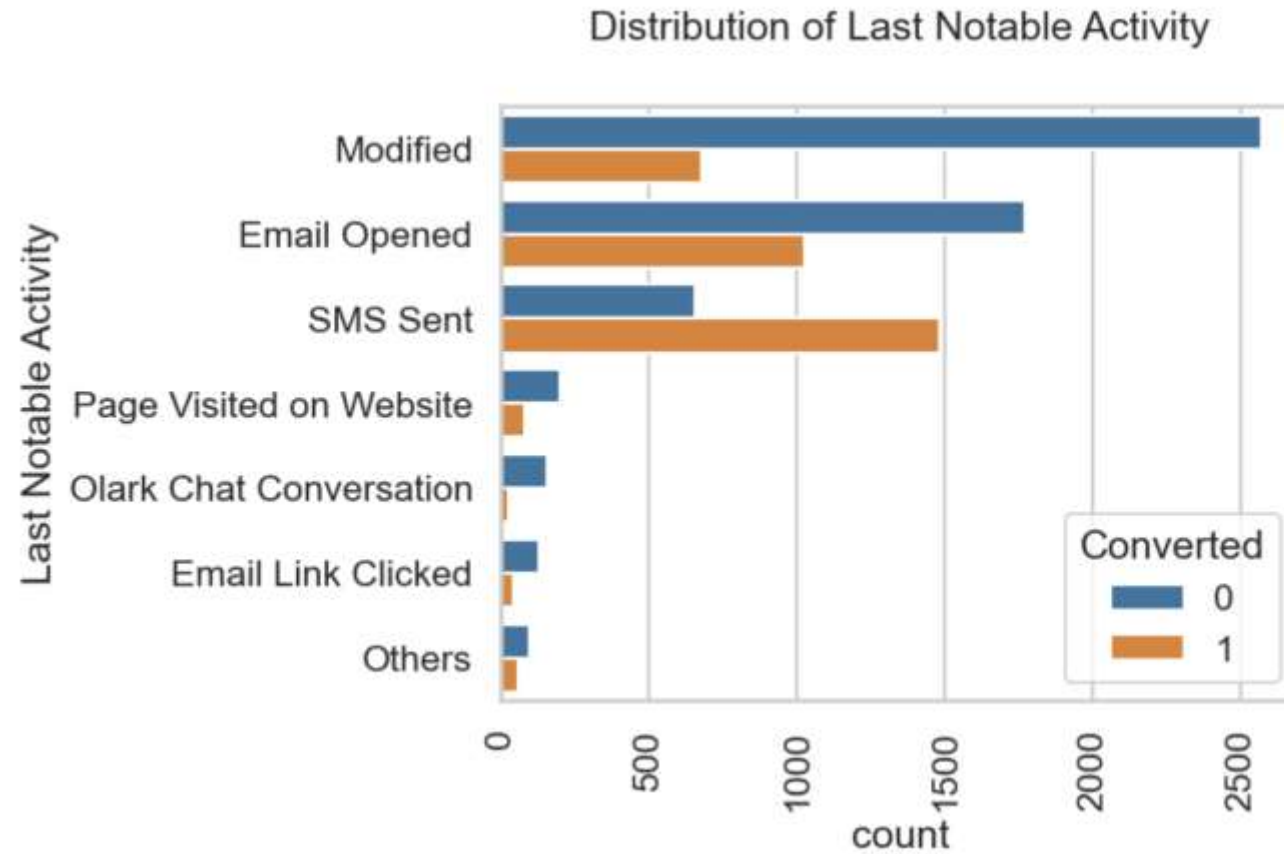
- Lead Add Form as Lead origin definitely leads to Conversion



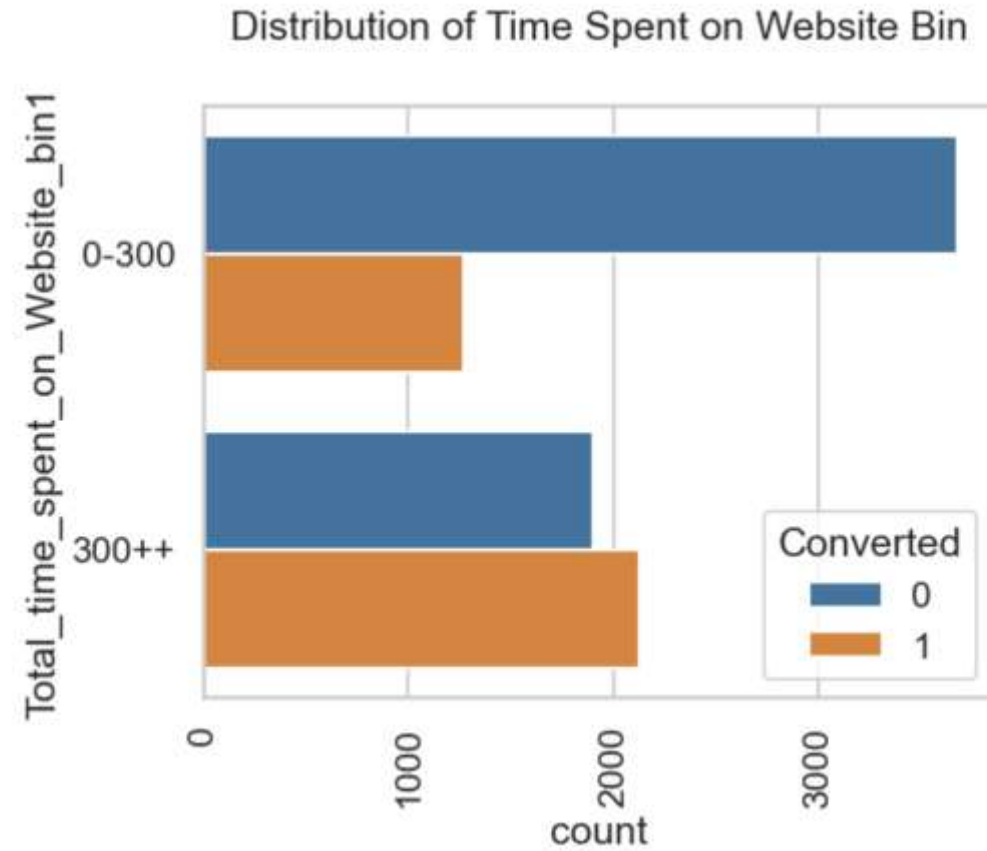
- Reference and Organic search have a better conversion rate



- Like Last Activity, Last notable activity shows higher conversion with SMS Sent followed by Email Opened

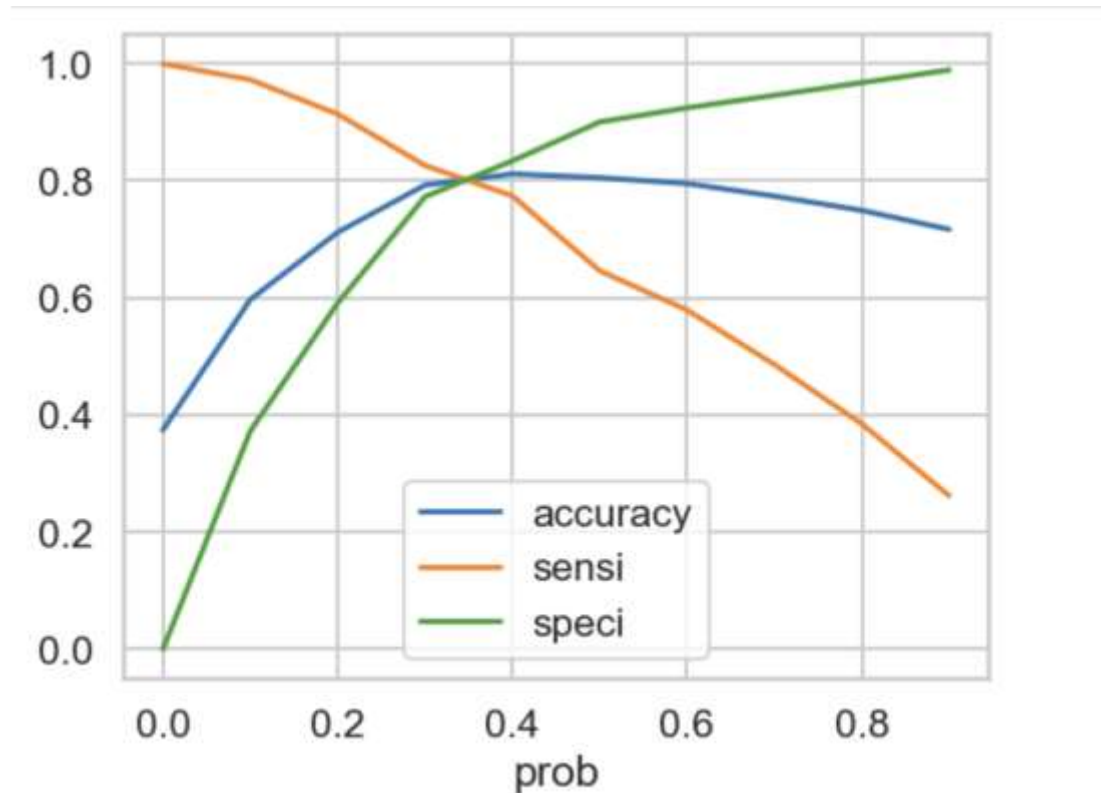


- Customers with more than 300 Seconds spent on website have a significantly high conversion rate.



Model Evaluation on Train Data Set

- The graph depicts an optimal cut off of 0.35 based on Accuracy, Sensitivity and Specificity



Accuracy - 80.59%

Sensitivity - 79.88%

Specificity - 81.01%

Model Evaluation on Test Data Set

- Accuracy – 81.47%
- Precision – 76.04%
- Recall – 76.26%

Confusion Matrix :

```
array([[1397, 251],  
       [ 248, 797]])
```


Key Variables

const	-1.1136
Do Not Email	-0.7603
Total Time Spent on Website	4.7311
Lead Origin_Landing Page Submission	-0.3260
Lead Origin_Lead Add Form	4.8161
Lead Source_Olark Chat	0.9797
Lead Source_Reference	-1.4199
Last Activity_Email Bounced	-1.8227
Last Activity_Olark Chat Conversation	-1.3557
Last Activity_Page Visited on Website	-0.8051
What is your current occupation_Unknown	-1.0867
What is your current occupation_Working Professional	2.5207
Last Notable Activity_Modified	-0.8259
Page_view_bin1_1-2	-0.3970
dtype: float64	

Conclusion

- The top variables that contribute for lead getting converted in the model are:
 - Total Visits
 - Total Time Spent on Website
 - Page Views Per Visit
- We have considered the optimal cut off based on Accuracy, Sensitivity and Specificity for calculating the final prediction.