Lead Scoring Case Study

By

Anusha V L

Shilpa Patil

Somasundar K U

PROBLEM STATEMENT

X Education is an organization which provides online courses for industry professional. The company marks its courses on several popular websites like google.

X Education wants to select most promising leads that can be converted to paying customers.

Although the company generates a lot of leads only a few are converted into paying customers, wherein the company wants a higher lead conversion. Leads come through numerous modes like email, advertisements on websites, google searches etc.

The company has had 30% conversion rate through the whole process of turning leads into customers by approaching those leads which are to be found having interest in taking the course. The implementation process of lead generating attributes are not efficient in helping conversions.

GOALS OF THE CASE STUDY

The company requires a model to be built for selecting most promising leads.

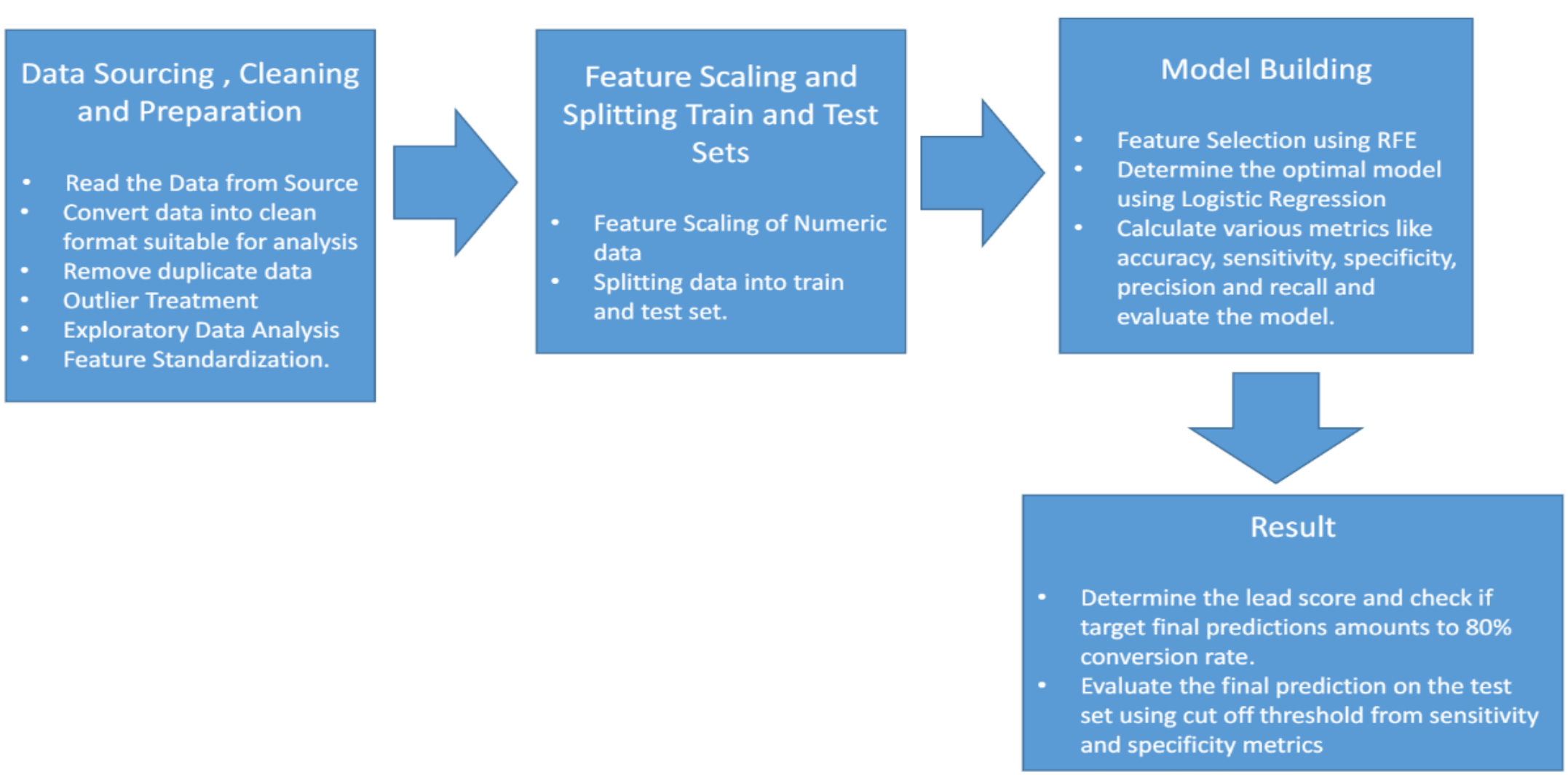
Lead score to be given to each leads such that it indicates how promising the lead could be. The higher the lead score the more promising the lead to get converted, the lower it is the lesser the chances of conversion.

The model to be built in lead conversion rate around 80% or more.

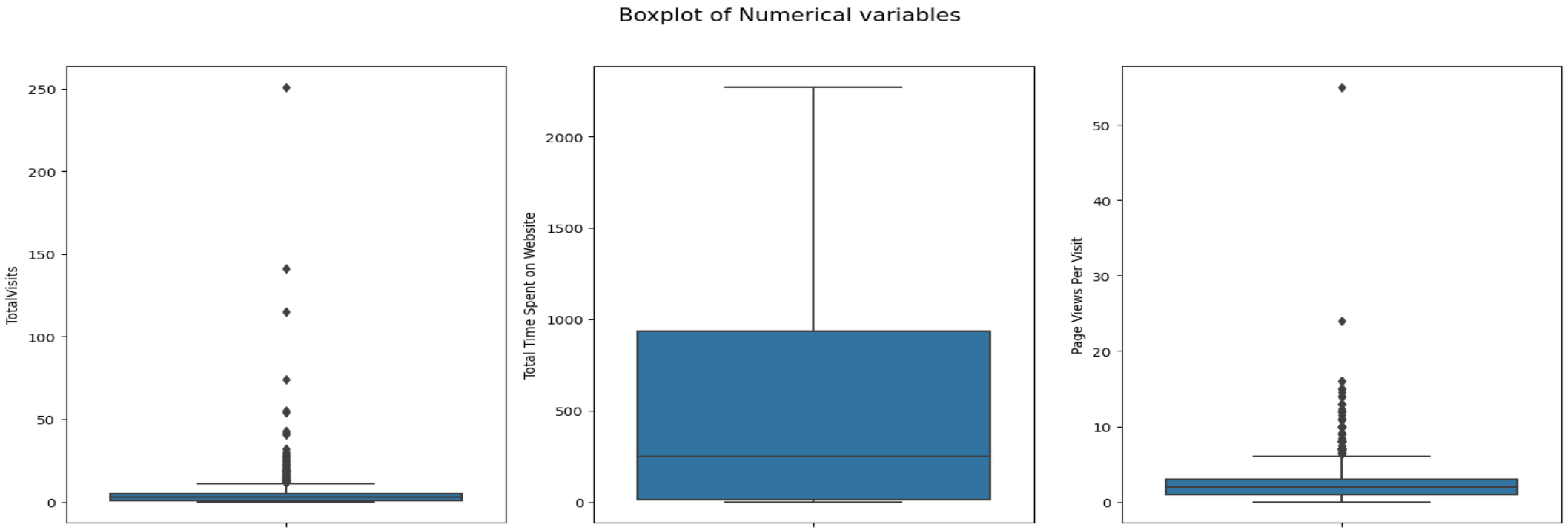
STRATEGY

- Import data.
- Clean and prepare the acquired data for further analysis
- Exploratory data analysis for figuring out most helpful attributes for conversion
- Scaling features
- Prepare the data for model building
- Build a logistic regression model
- Assign a lead score for each leads
- Test the model on train set
- Evaluate model by different measures and metrics
- Test the model on test set
- Measure the accuracy of the model and other metrics for evaluation

PROBLEM SOLVING METHODOLOGY

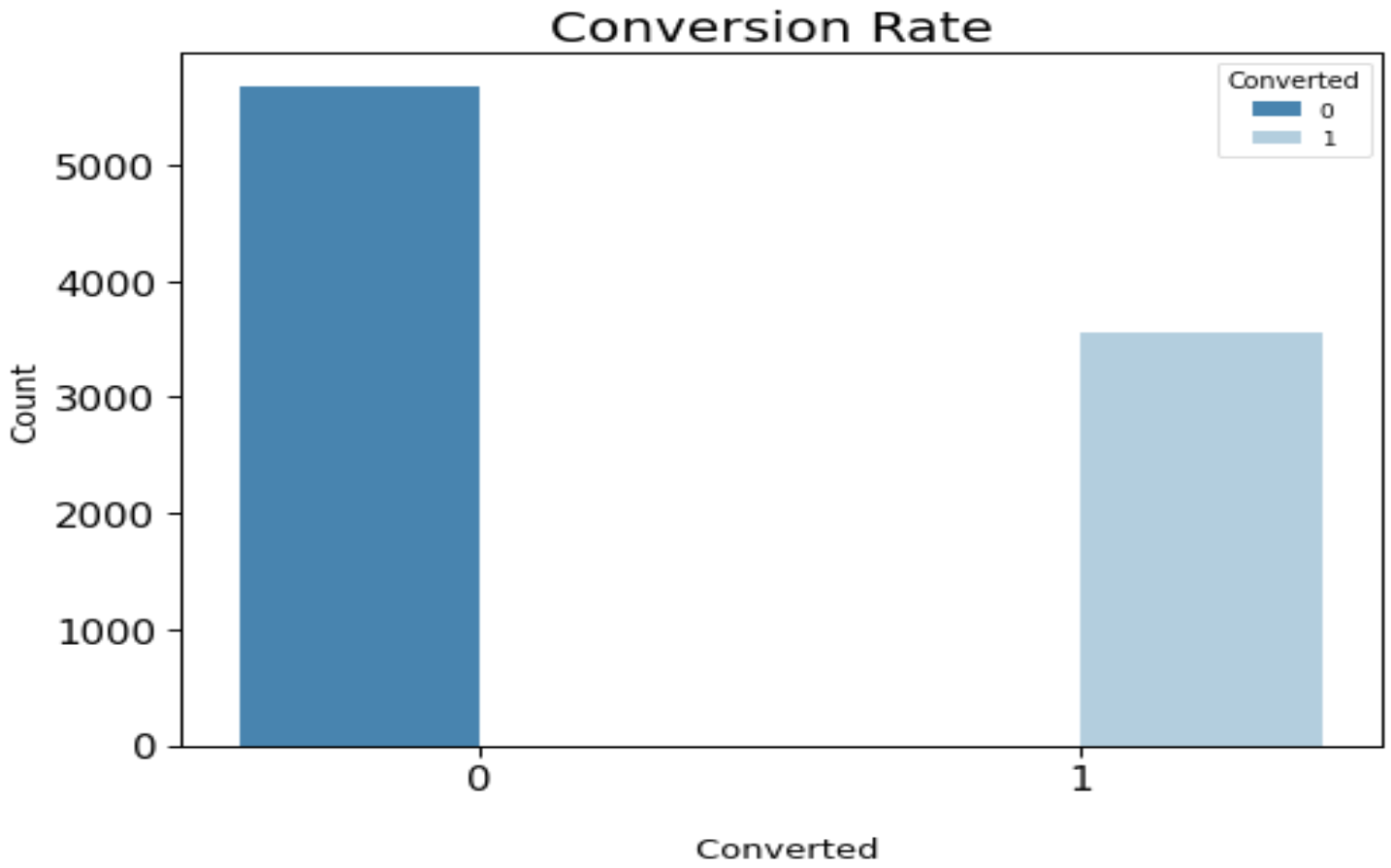


OUTLIERS



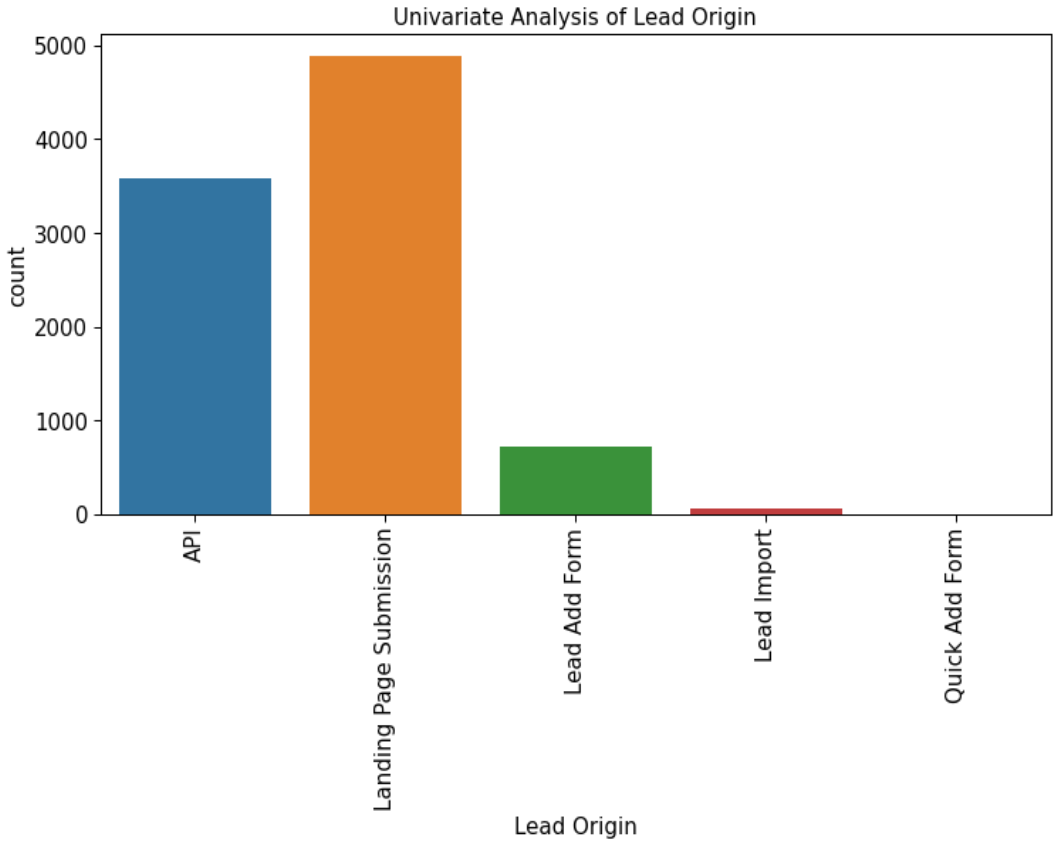
- TotalVisits and Page Views Per Visit has outliers
- Total Time Spent on Website has min value at 0 and most of values lies below 1000
- TotalVisits has most values lies between 1 and 5
- Page Views Per Visit has most values lies below 5 and above upper whisker at around 8 are outliers

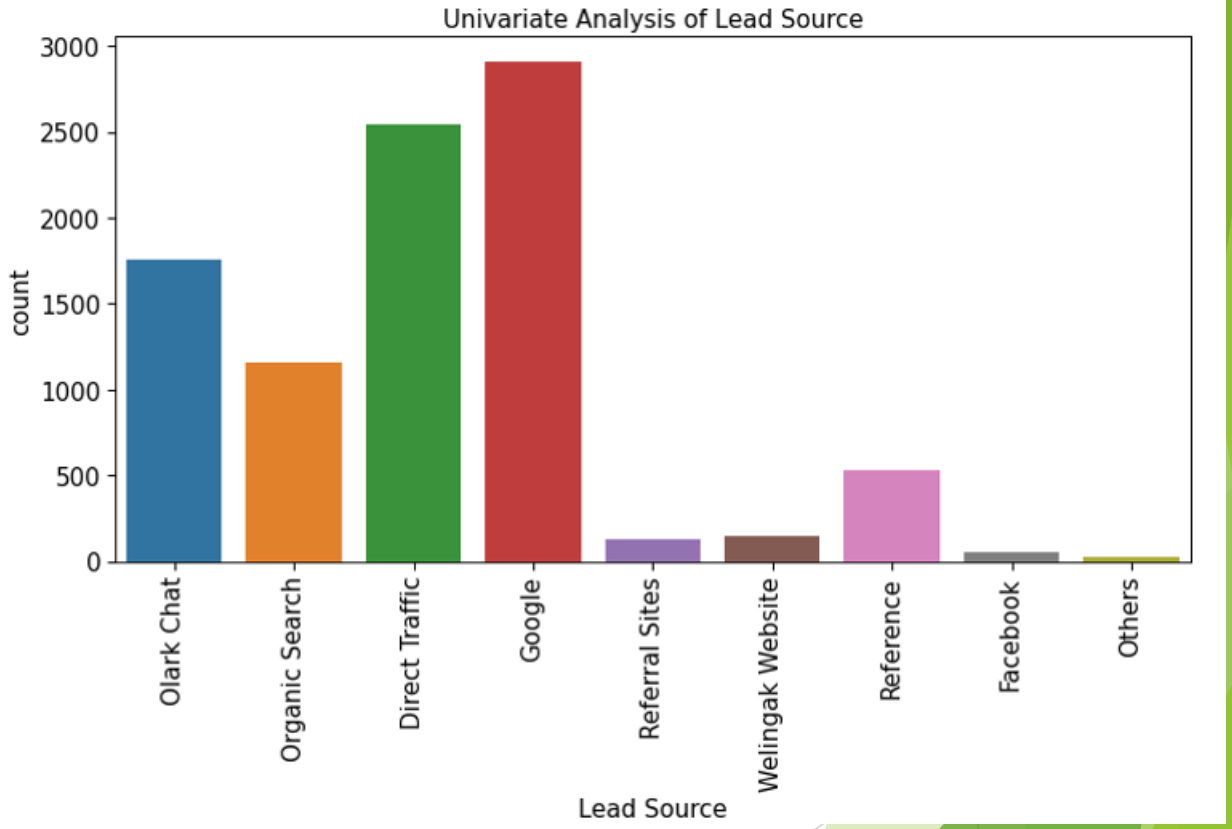
ANALYSIS OF TARGET VARIABLE



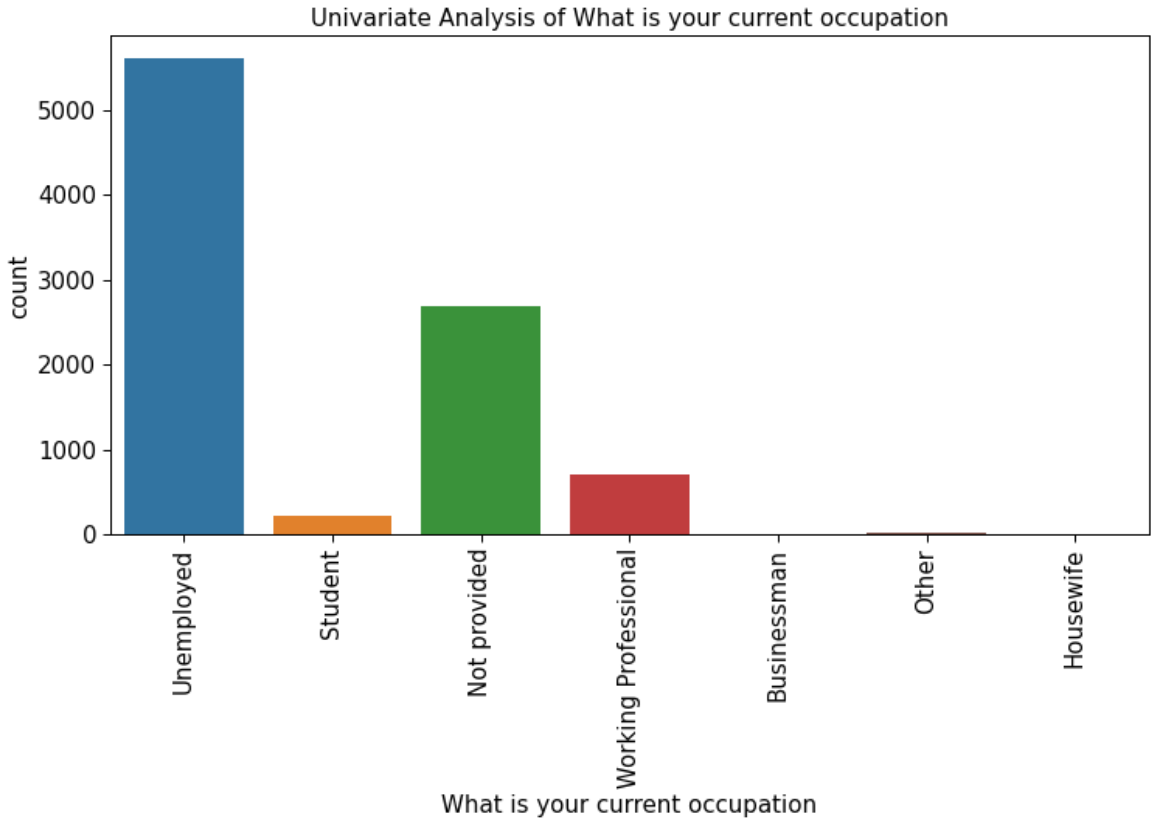
Conversion rate is around 38% only.Most of the leads are not converted

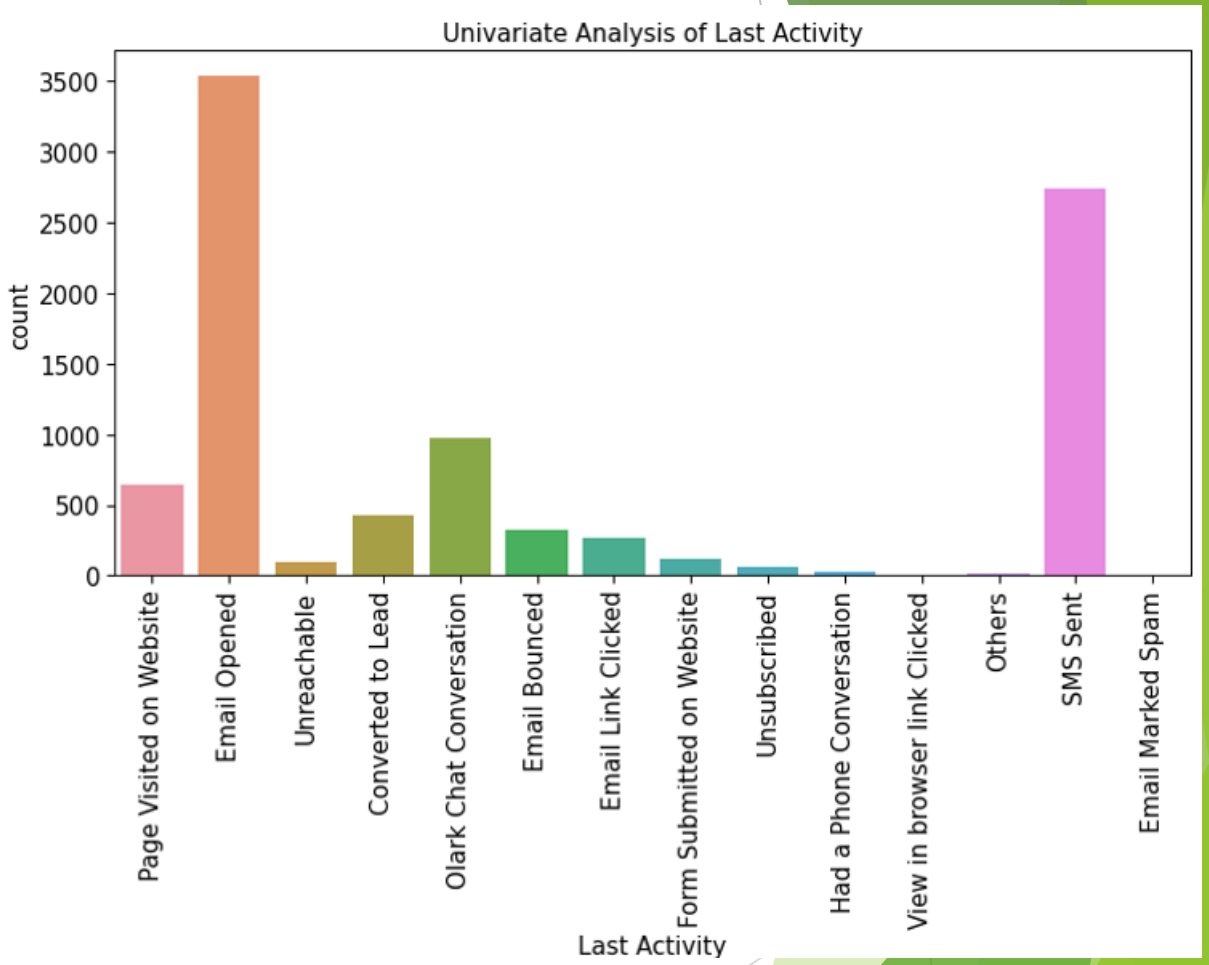
UNIVARIATE -CATEGORICAL





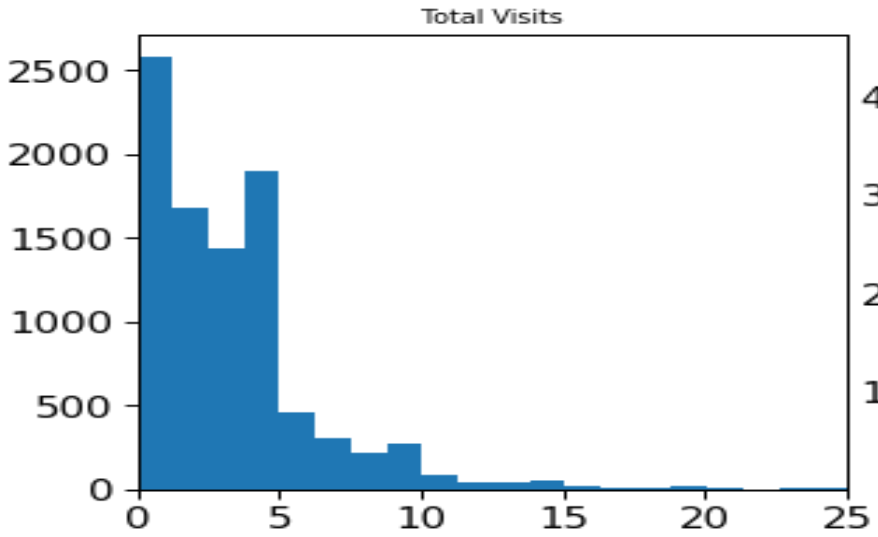
- Most of leads are generated from 'Landing Page Submission' followed by API
- Google and direct traffic produce the most prospects.

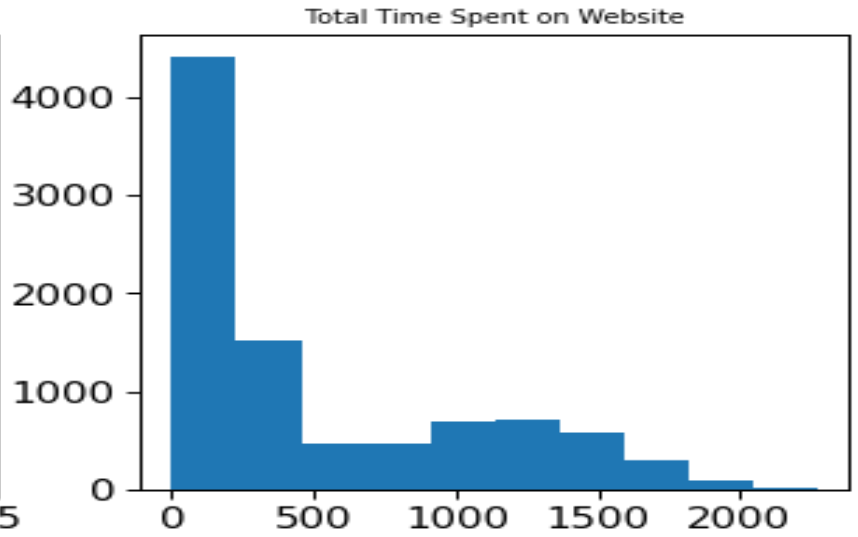


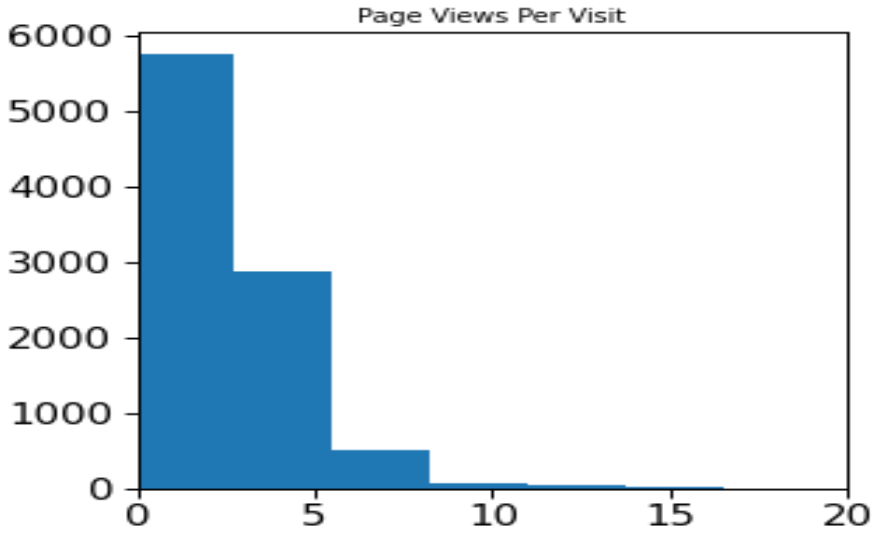


- Most leads are generated when the last action was an email opened followed by SMS sent
- Around 90% of the leads generated are unemployed followed by working professionals

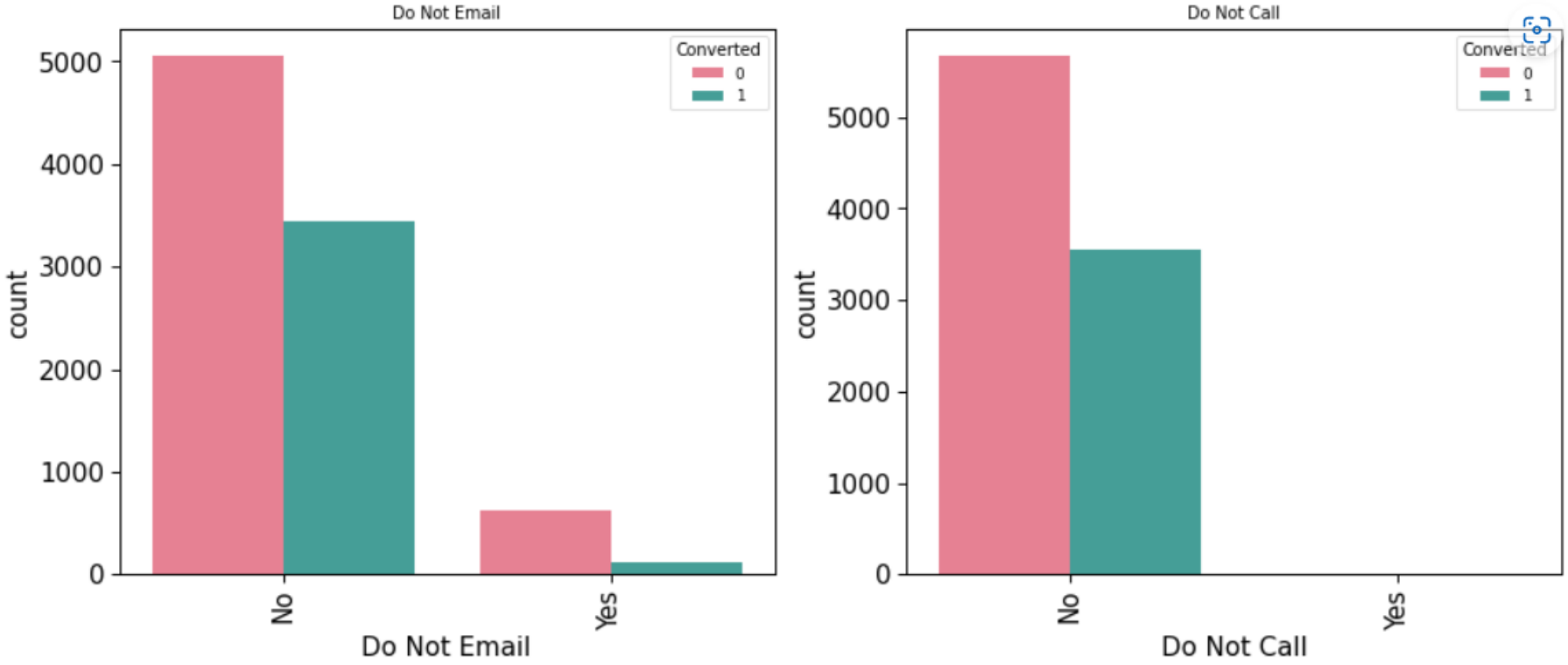
UNIVARIATE-NUMERICAL



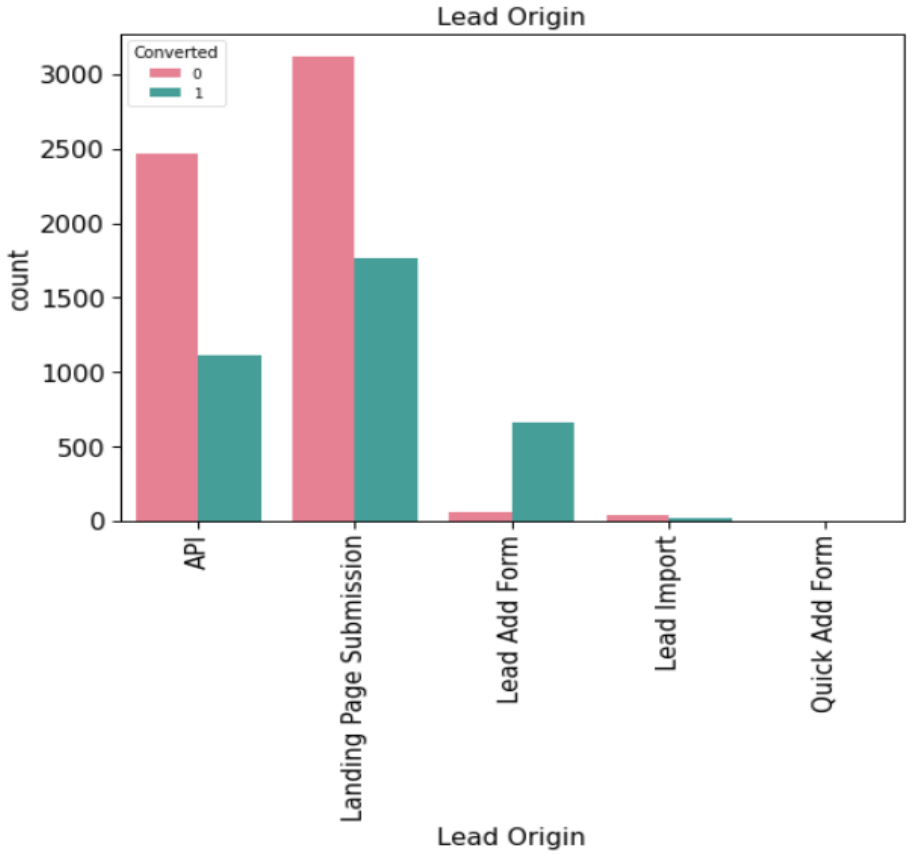


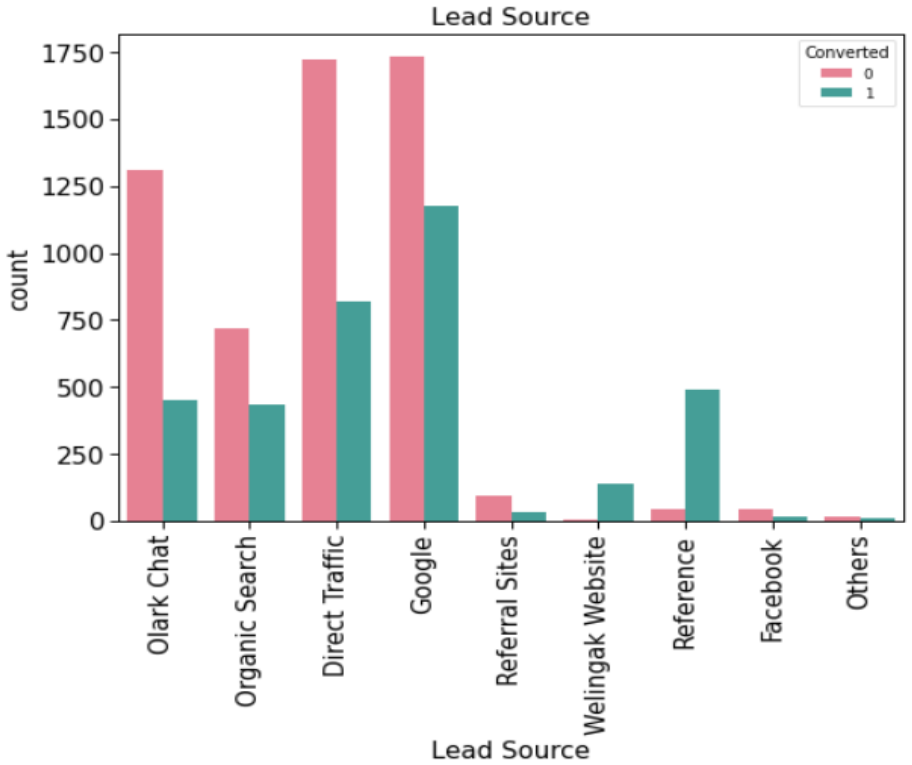


CATEGORICAL VARIABLES VS CONVERTED

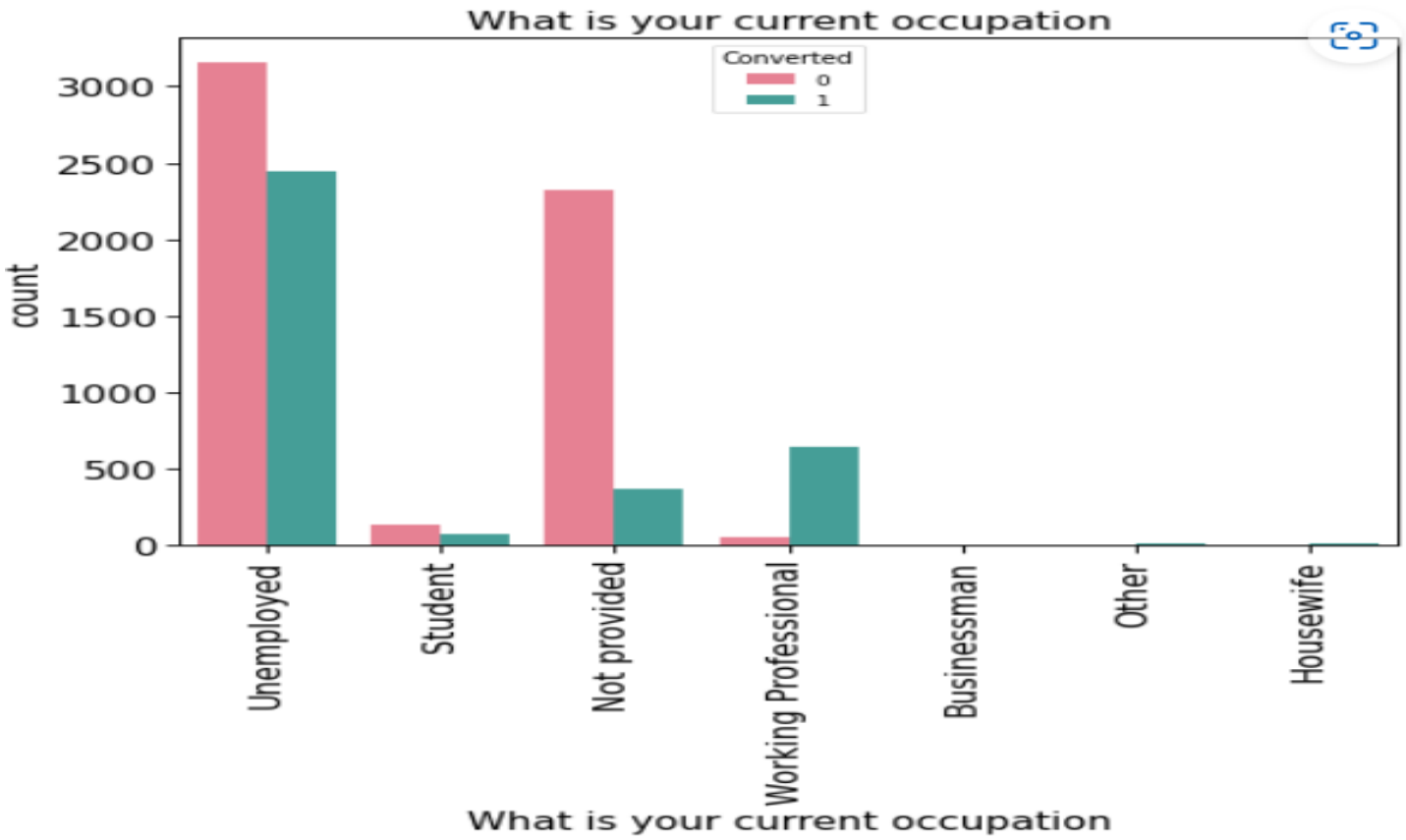


- Do not Call is imbalanced as almost all customers are in category 'No'. So this column cannot be used for further Analysis
- Do not Email has more conversion rate for customers who responded 'No'



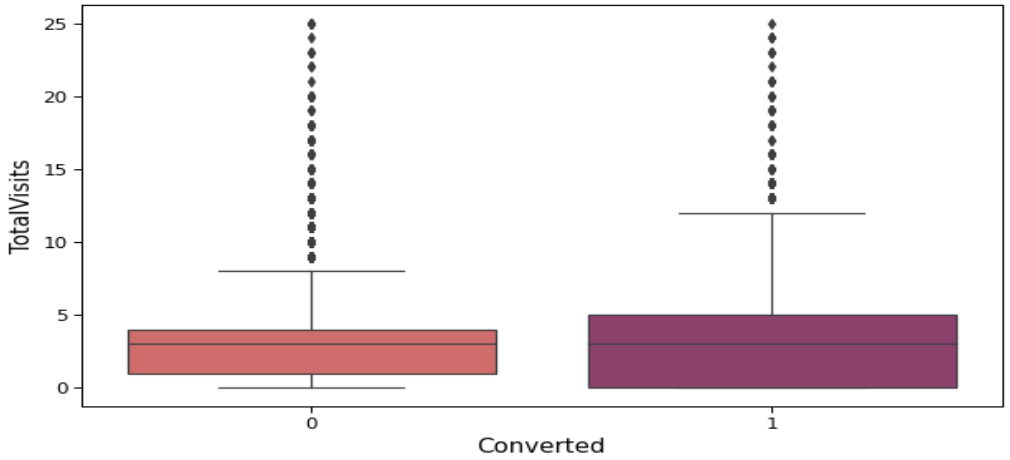


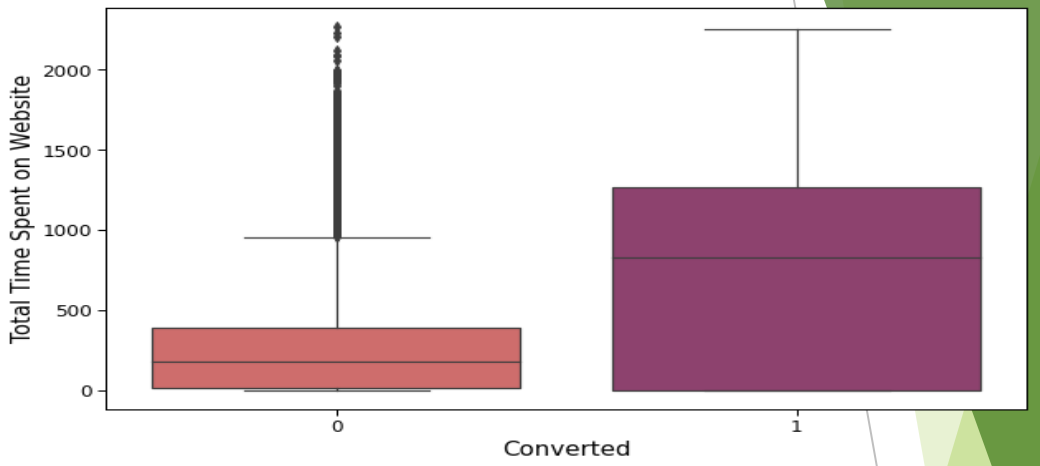
- Conversion is mostly from Landing Page Submission and API. Lead Ad form has more converted leads than nonconverted
- Google and Direct Traffic generated more leads and high in conversion also
- Reference and welingak Website has high conversions

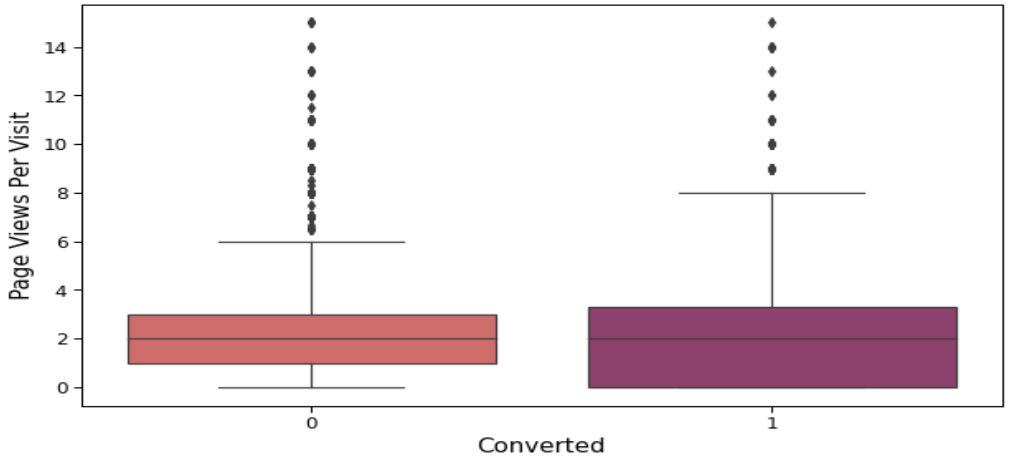


- Unemployed generated high leads and has high conversion rate also
- Working Professional has high conversion rate

NUMERICAL VARIABLES VS CONVERTED

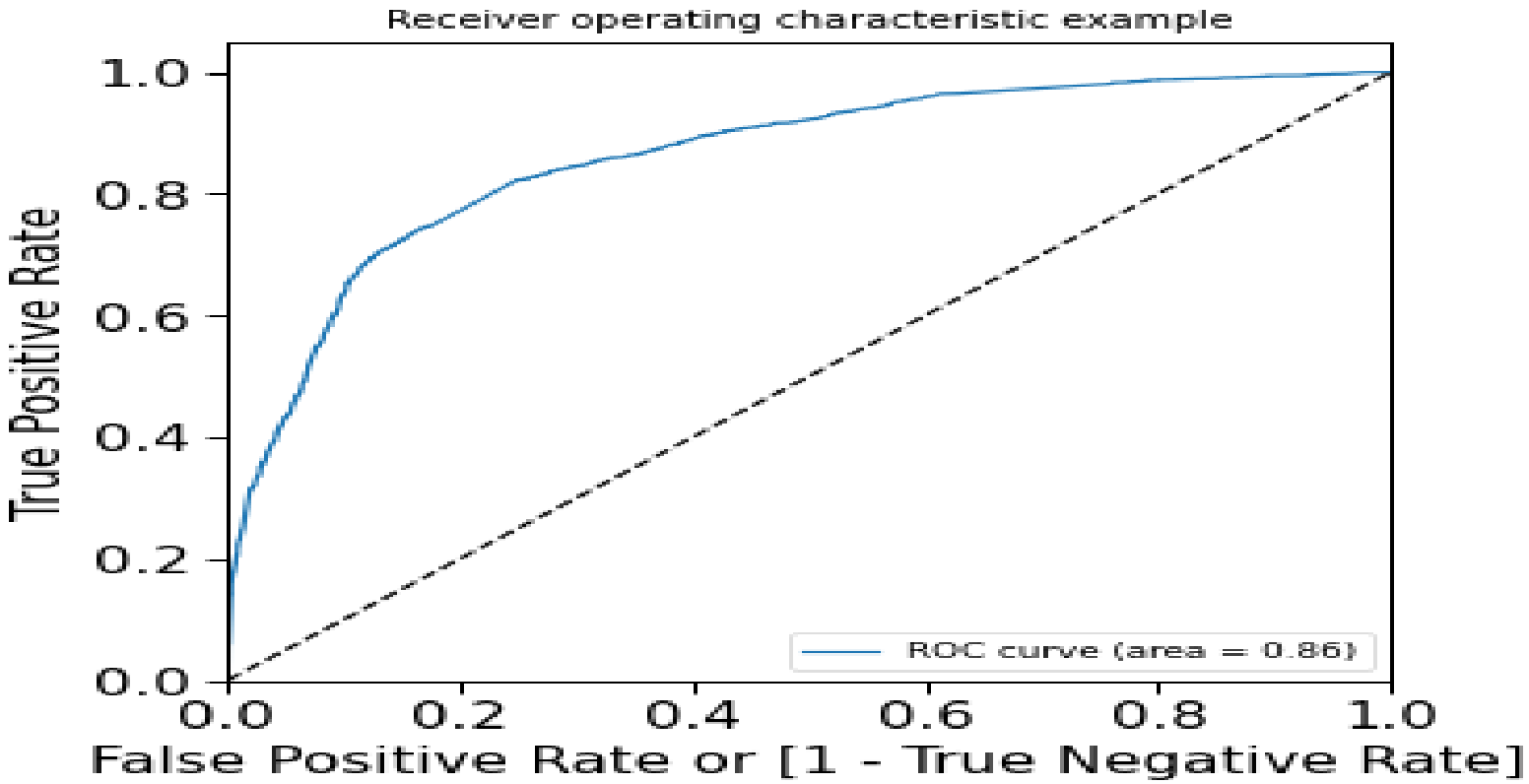






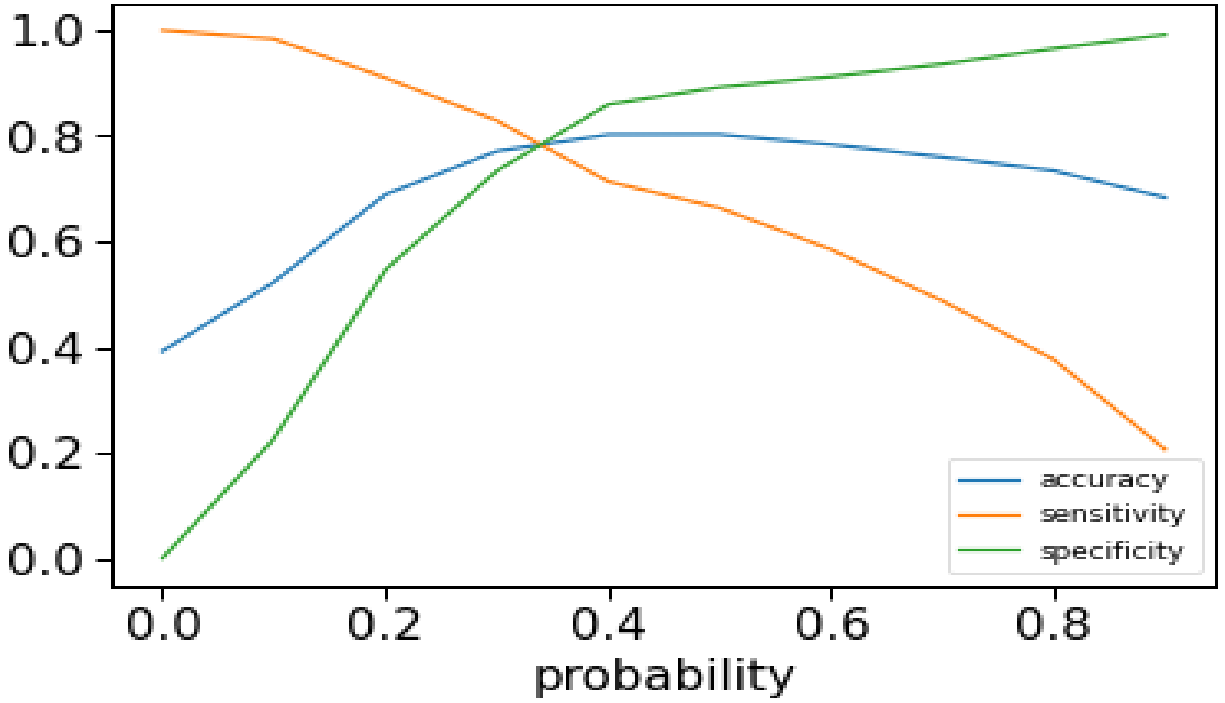
- For TotalVisits,median of both converted and not converted are same
- we can see an increase in conversion rate for those who spent more time on website.so to boost conversion company should make website more reliable and engagable
- Median for page views per visit for both category are same

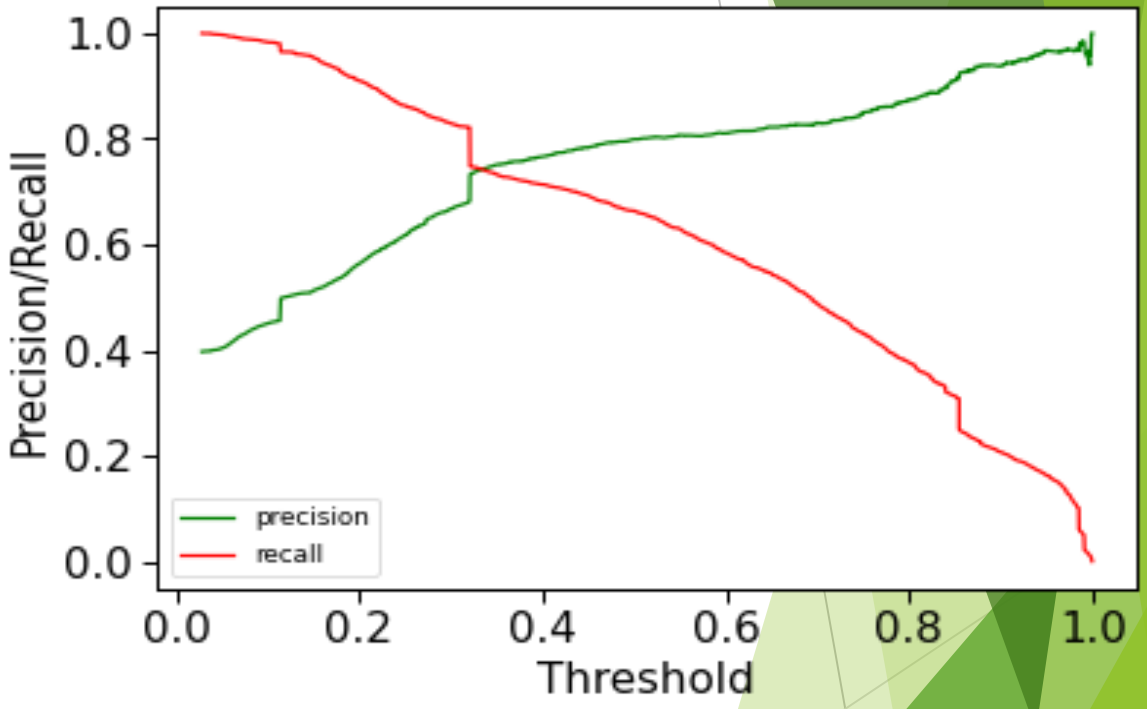
ROC CURVE



- we got a ROC curve with area of 0.86 which indicates a good model

OPTIMAL CUTOFF POINT





- From the curve above,we got 0.3 is the optimum point from Sensitivity-Specificity graph,As we are focusing on high Sensitivity we can take it as a cutoff probability.

MODEL EVALUATION

The model appears to be working well with an ROC curve of area 0.86

We have following Metrics for Train dataset

- Accuracy:**77%**
- Sensitivity:**83%**
- Specificity:**73%**

We have following Metrics for Test dataset

- Accuracy:**77.4%**
- Sensitivity:**84.8%**
- Specificity:**73.11%**

CONCLUSION

- The test set's values for accuracy, sensitivity, and specificity are around 77%, 84%, and 73.11% which are closer to the corresponding figures derived using the training set.
- As conversion rate is above 80%,We should be able to give the CEO confidence in making wise decisions based on the model since it appears to predict the Conversion Rate quite accurately.
- Important features that contribute most to the likelihood of a lead being converted or those that have a high conversion rate are:
  - What is your current occupation_Working Professional
  - Lead Origin_Lead Add Form
  - Lead Source_Welingak Website

RECOMMENDATION

- Develop Strategies for attracting customers from potential lead sources
- Invest more on website development to attract more customers
- Focus more on working professionals as they are hot leads.Give them attractive offers for courses
- More money will be allocated for advertising on welingak website.
- For marketing tactics, prioritise features with a positive coefficient.