


# LEAD SCORE CASE STUDY



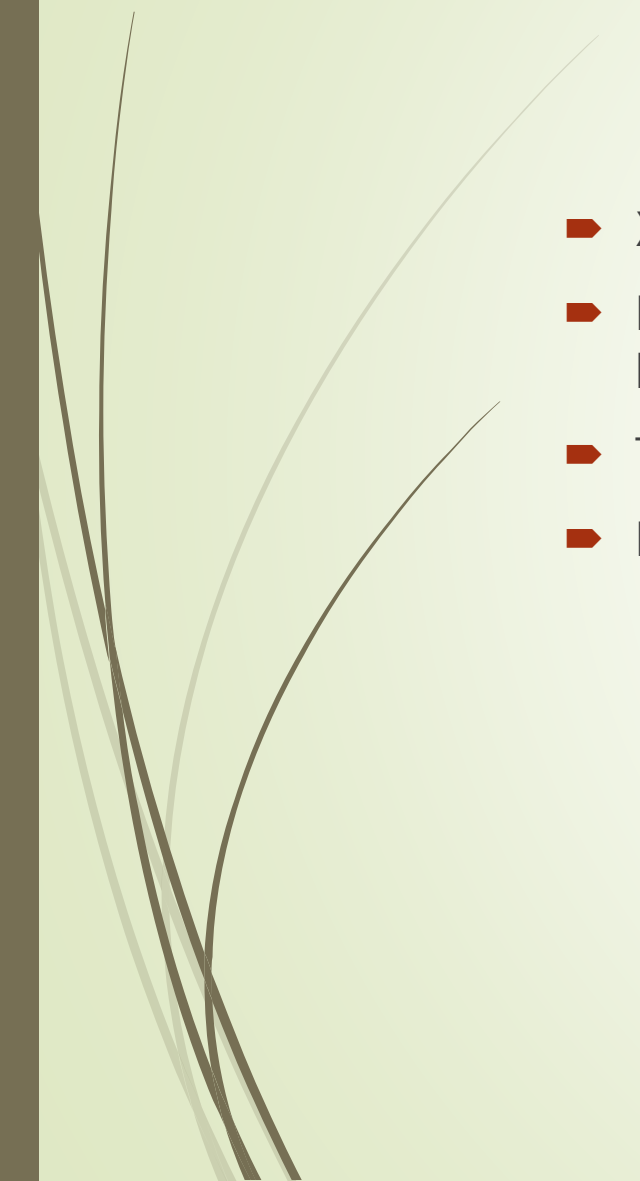


# PROBLEM

- X Education want to sell online courses to industry professionals.
  - X Education gets a lot of leads but conversion rate is very poor. They have only 30% Conversion Rate.
  - To make this process more efficient, the company wishes to identify the most potential leads, also known as 'Hot Leads'.
  - They want to successfully identify the set of leads so the conversion rate go up.
  - Sales team focusing more on communicating with the potential leads rather than making calls to everyone.
- 



# BUSINESS OBJECTIVE

- X education wants to know most promising leads.
  - Lead Score given to each Leads to indicate how promising the lead could be.
  - The Model built such as the conversion rate should around 80% Or more.
  - Deployment of the model for the future use.
- 

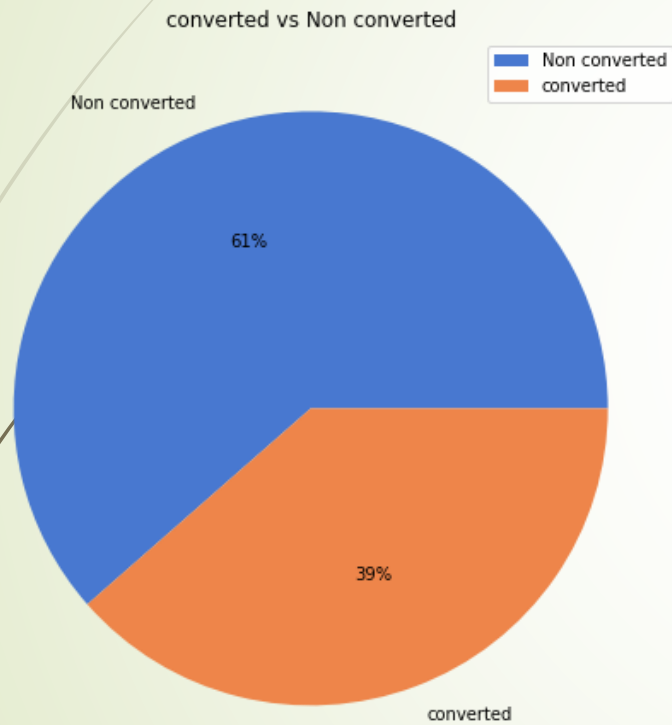


# PROCESS

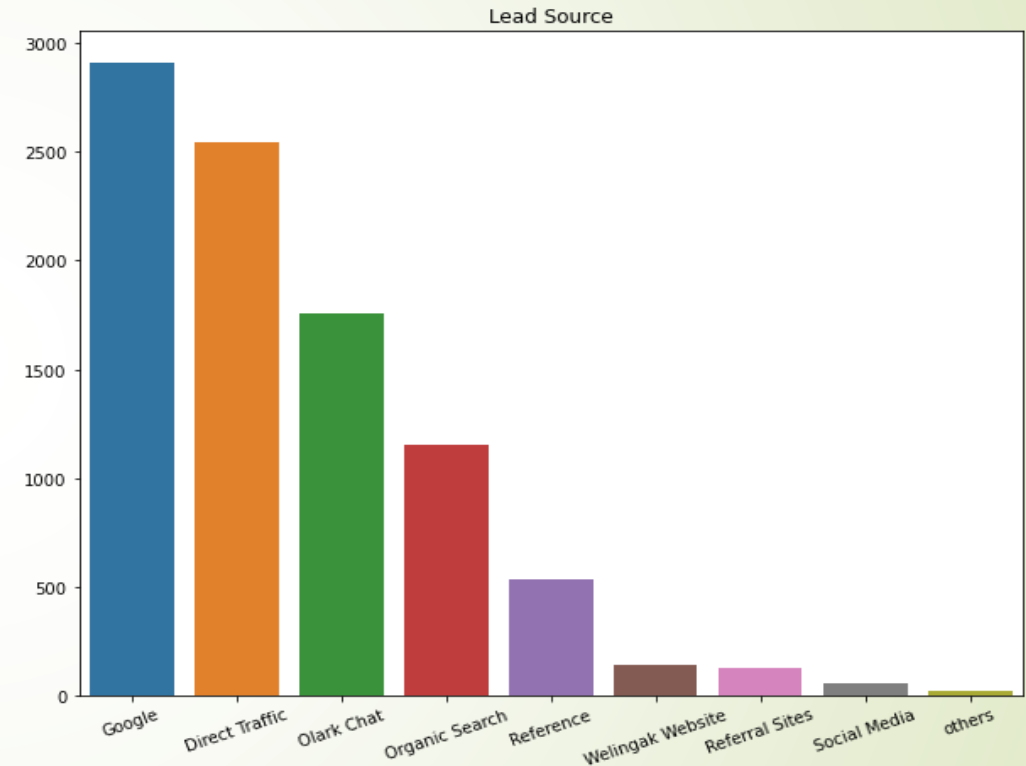


- Importing The Required dataset on which Analysis performed on.
- Data cleaning and data manipulation for further Analysis
- Check duplicate data, NAN values and missing values and get rid of this.
- Drop columns, if it contains large amount of missing values and not useful for the analysis.
- Imputation of the values.
- Check and handle outliers in data.
- EDA
  - 1. Univariate analysis
  - 2. Bivariate analysis
  - 3. Multivariate analysis
- Feature Scaling & Dummy Variables and encoding of the data.
- Classification technique: logistic regression used for the model making and prediction.
- Validation of the model.
- Model presentation.
- Conclusions and recommendations.

# EXPLORATORY DATA ANALYSIS

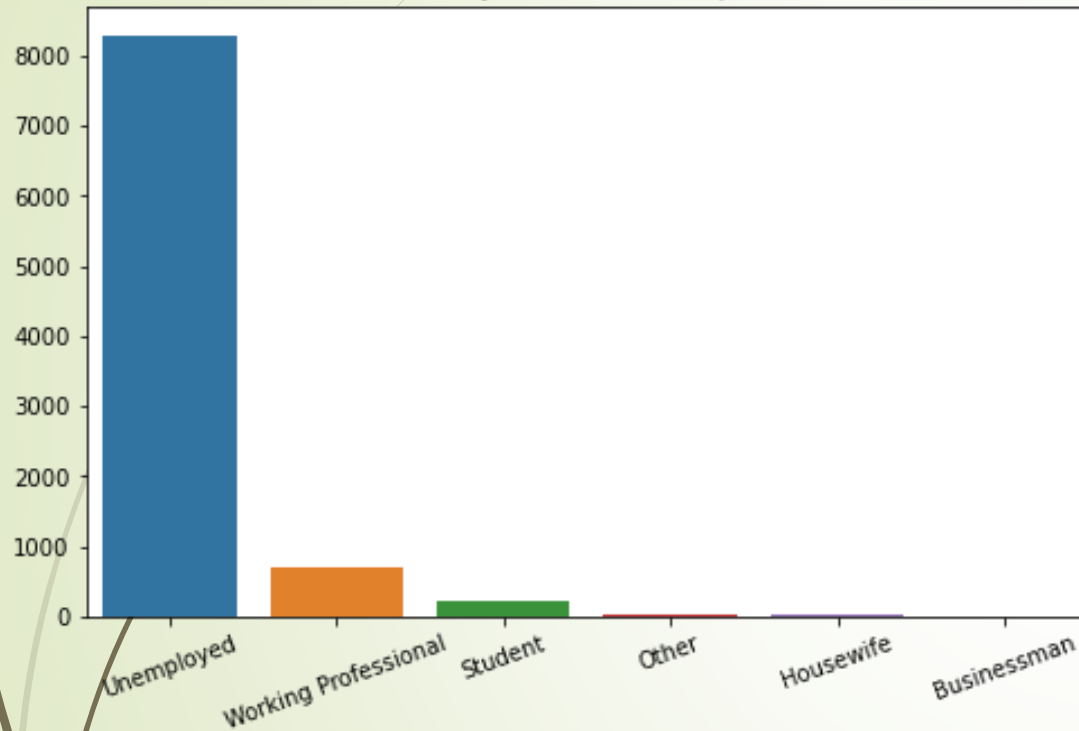


From This analysis we see that 39% Lead is converted.



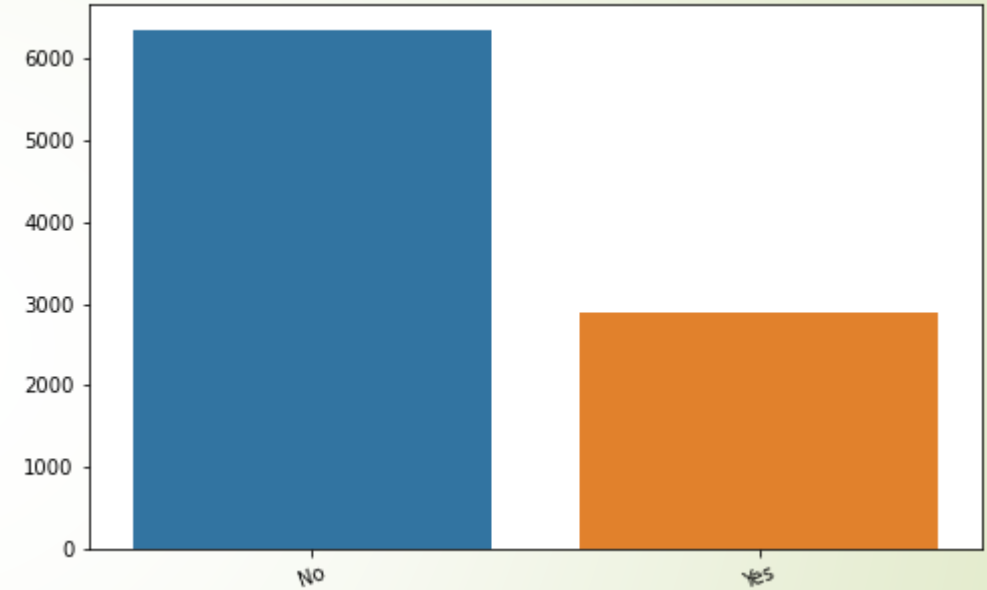
Google is most valuable lead source for generating leads

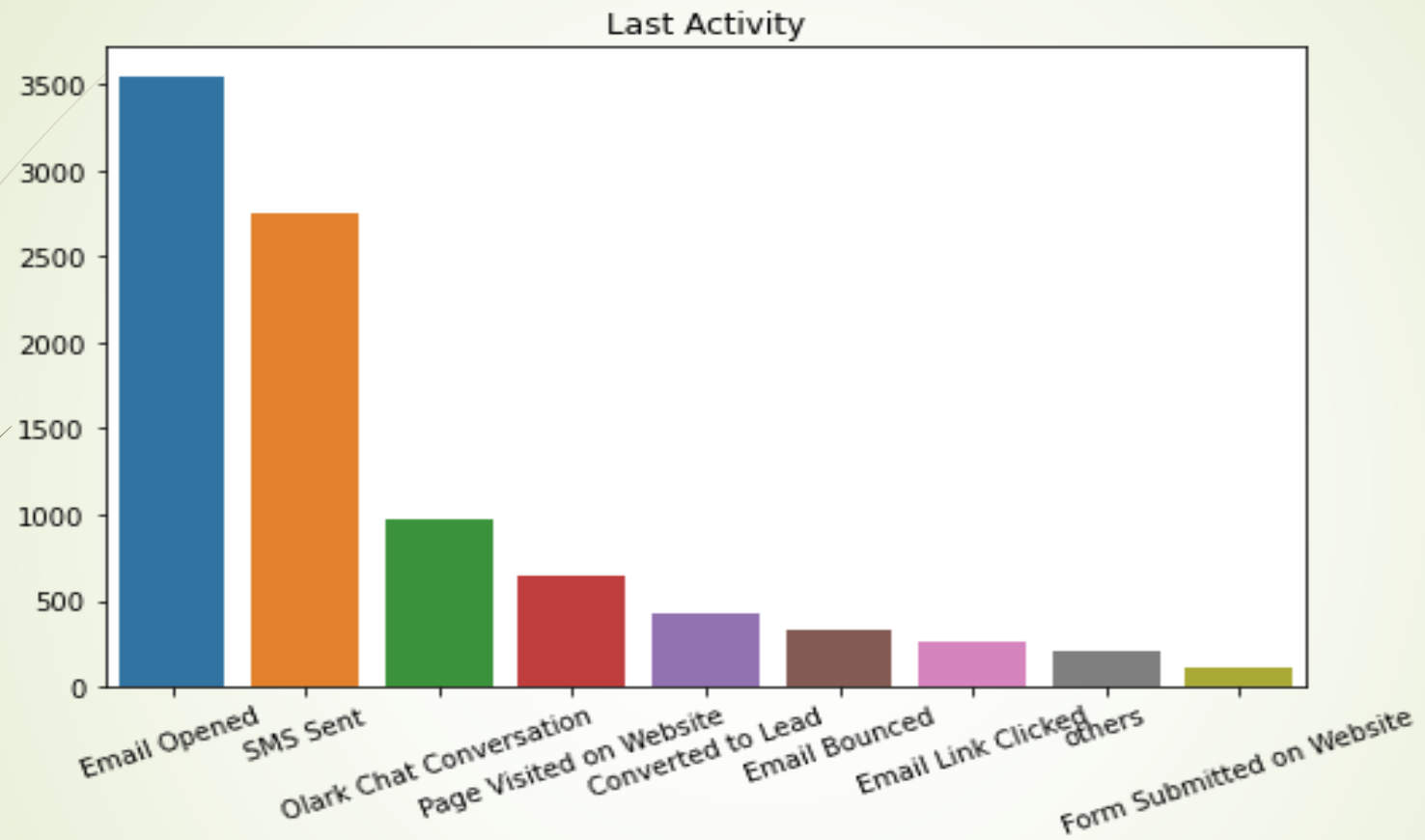
What is your current occupation



Unemployed are the most important lead then other occupation.

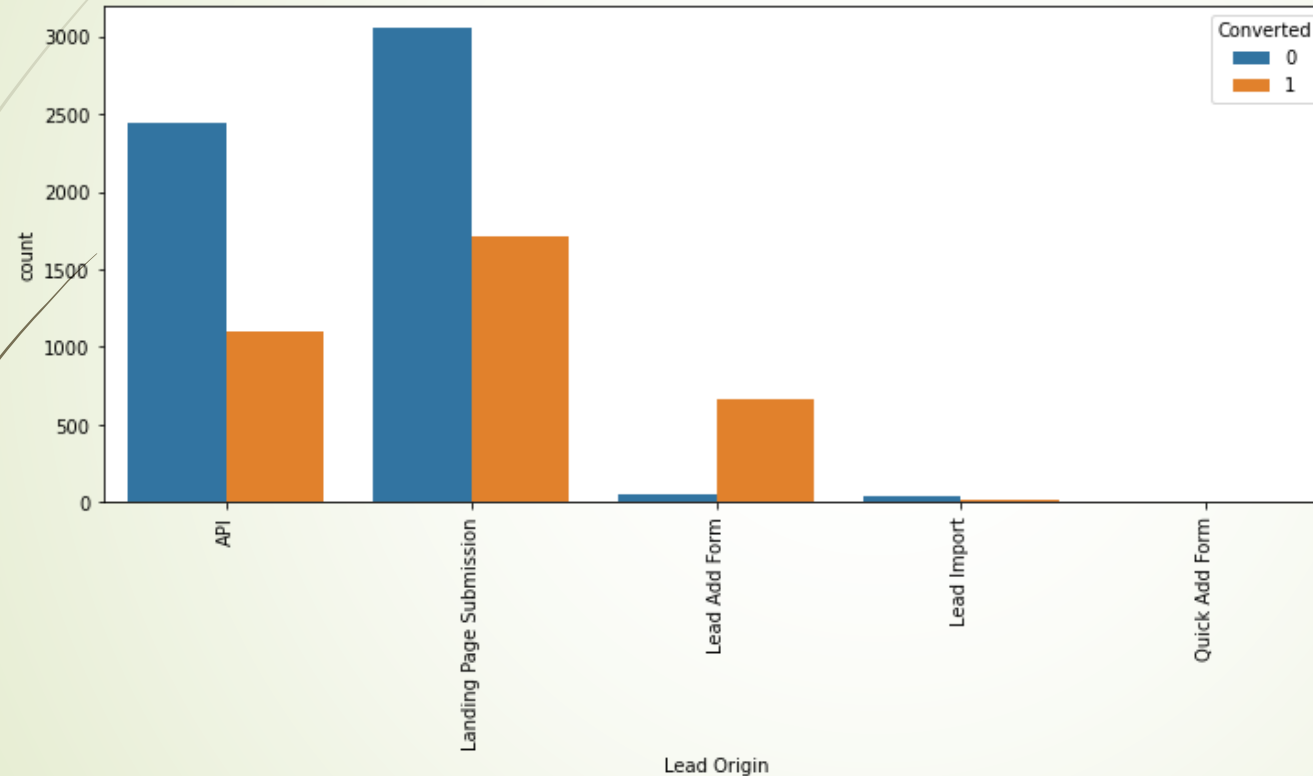
A free copy of Mastering The Interview





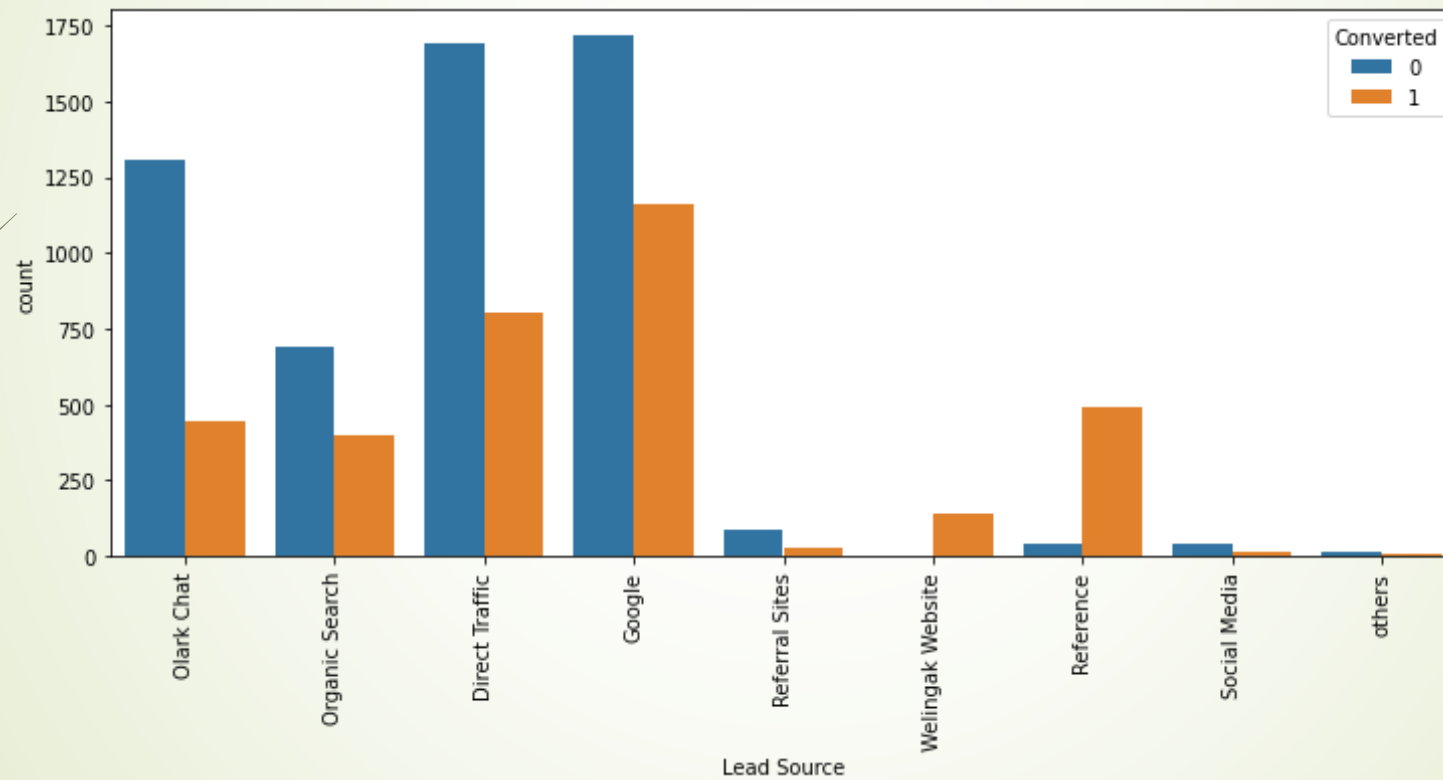
Email Opened and sms sent are the most frequent than other Activity leads

# BIVARIATE ANALYSIS

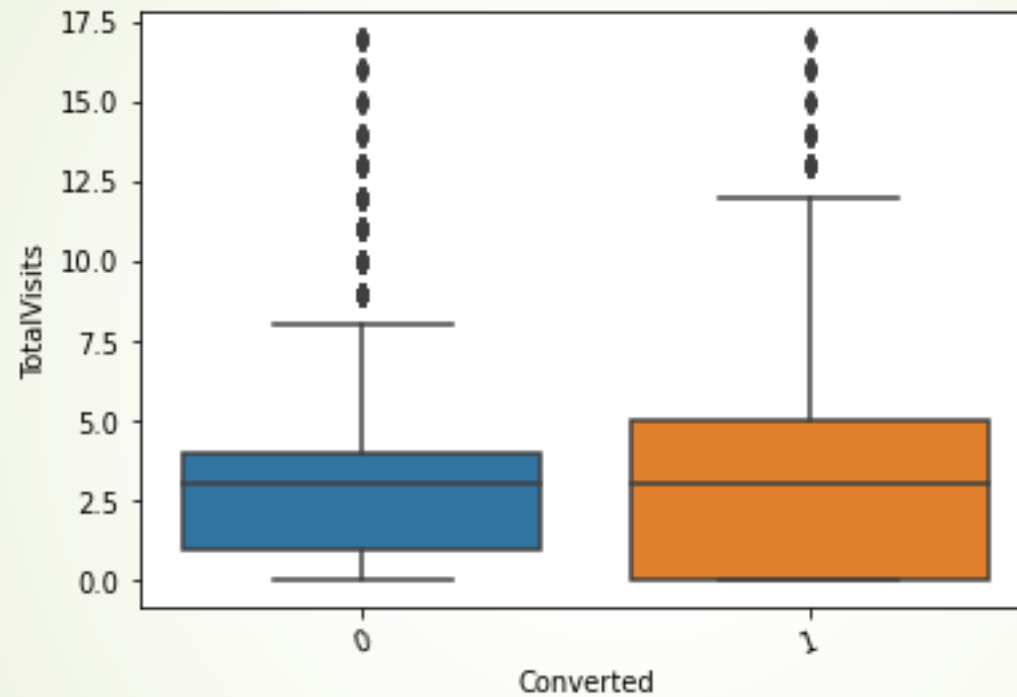


The Landing page Submission are more converting in numbers and the Lead Add form have high conversion rates.



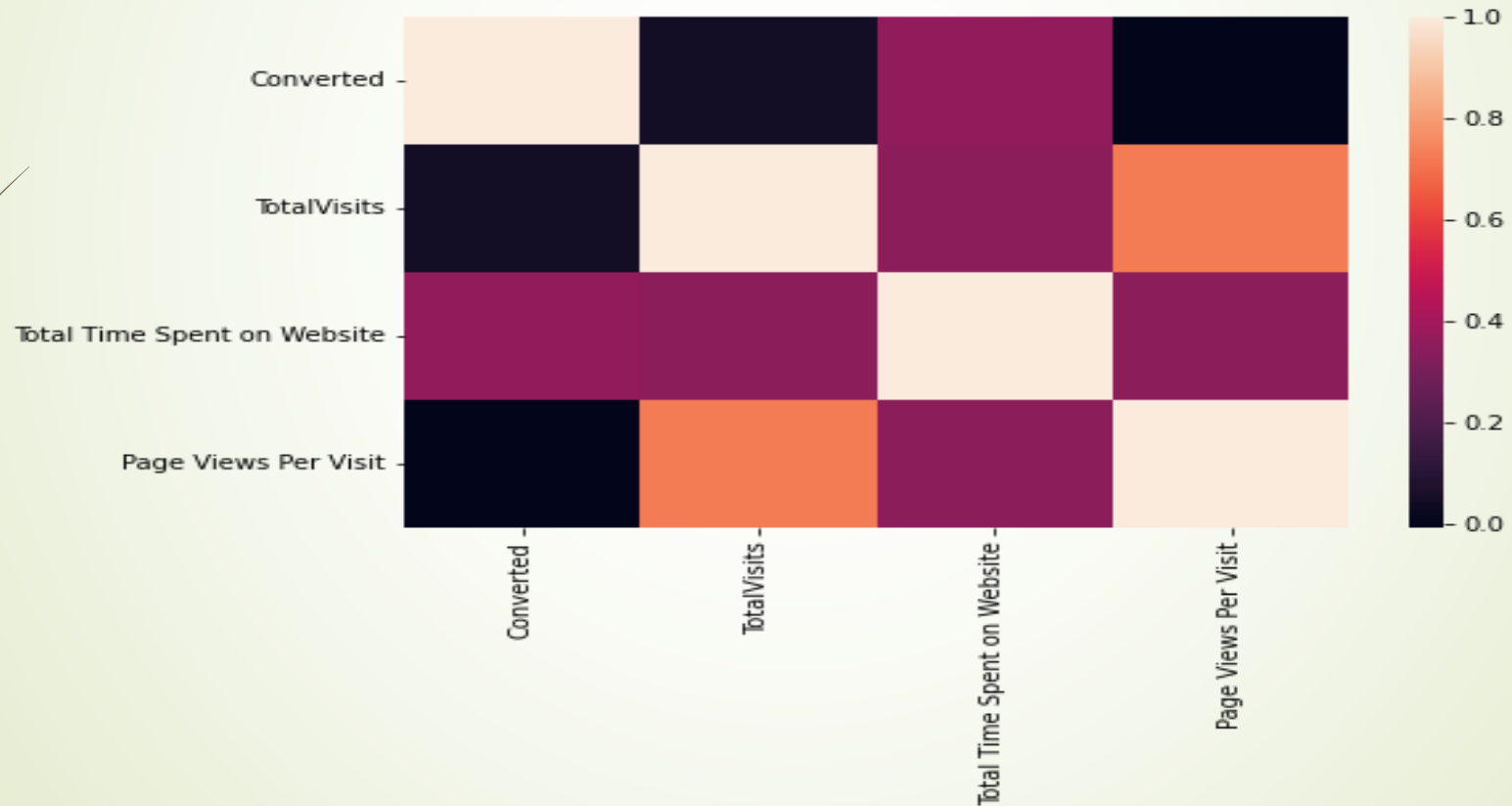


Lead Source from Google is higher Value but reference have high in conversion.



The similar trend is seen in converted and non converted but the upper and the lower range differs but they have similar median range.

# MULTIVARIATE ANALYSIS

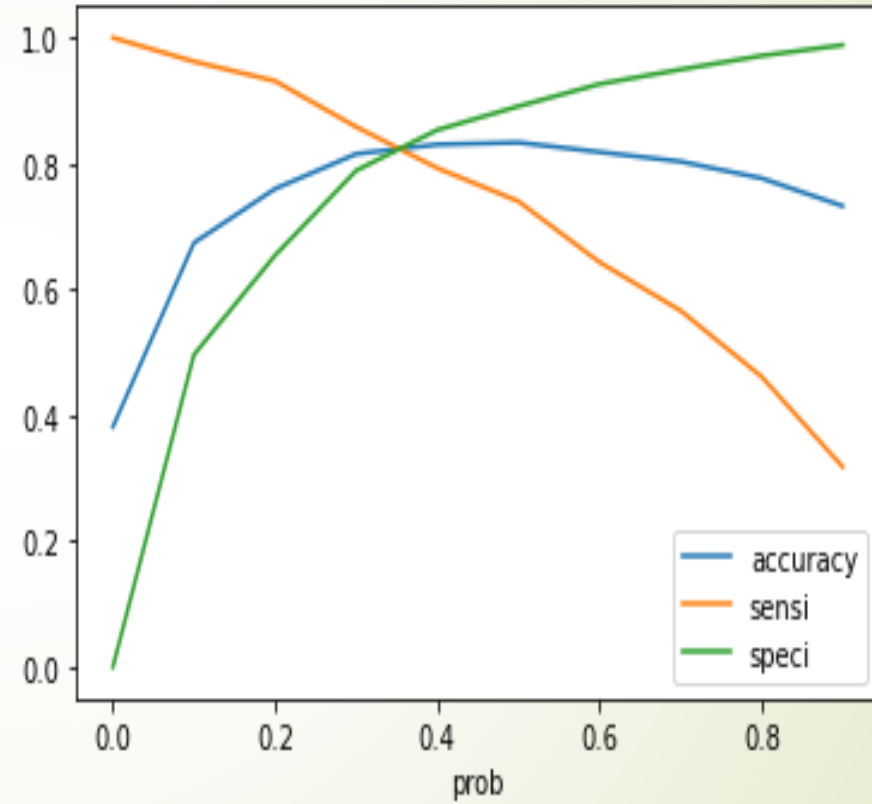
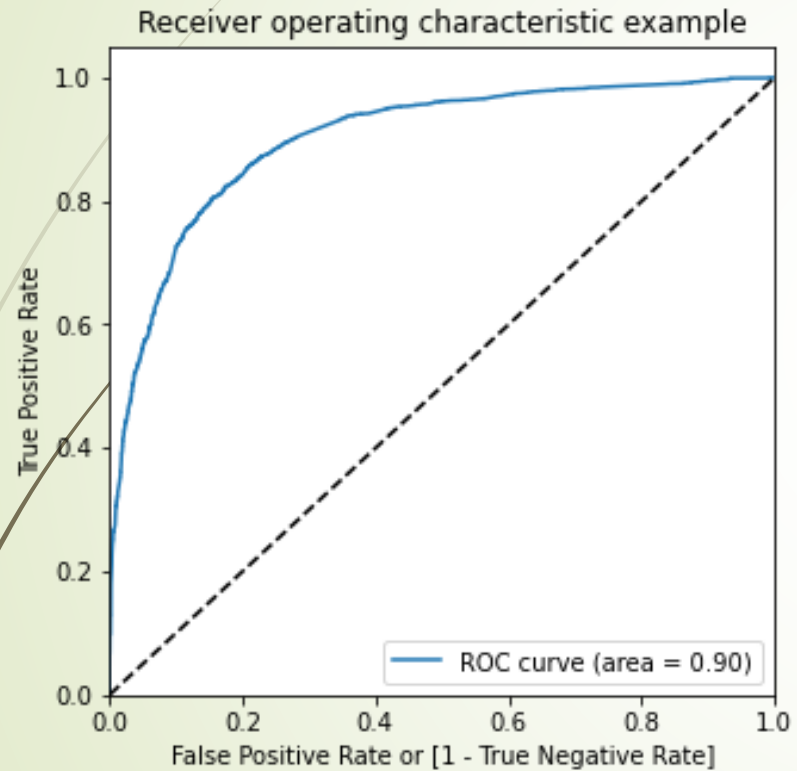




# MODEL BUILDING

- Splitting the Data into Training and Testing Sets
- The first basic step for regression is performing a train-test split.
- Use RFE for Feature Selection
- Running RFE with 18 variables as output
- Building Model by removing the variable whose p- value is greater.
- Predictions on test data set
- Overall accuracy 81.4%

# ROC CURVE





# CONCLUSION

- The total time spend on the Website have higher chance of conversion.
- Total number of visits.
- When the lead source was Google, Direct traffic, website and Sms have higher chance of conversion.
- When current occupation is working professional, Keeping this in mind the X Education can flourish as they have a very high chance to get almost all the potential buyers to change their mind and buy their courses.
- The Accuracy is more than 81%.



**THANK YOU**