

# Lead Scoring Case Study

## Group Members

- 1) Ish Bhargava
- 2) Peeyush Saxena

# Problem Statement

- ❑ X Education sells online courses to industry professionals.
- ❑ X Education gets a lot of leads, its lead conversion rate is very poor. For example, if, say, they acquire 100 leads in a day, only about 30 of them are converted.
- ❑ To make this process more efficient, the company wishes to identify the most potential leads, also known as 'Hot Leads'.
- ❑ If they successfully identify this set of leads, the lead conversion rate should go up as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone.

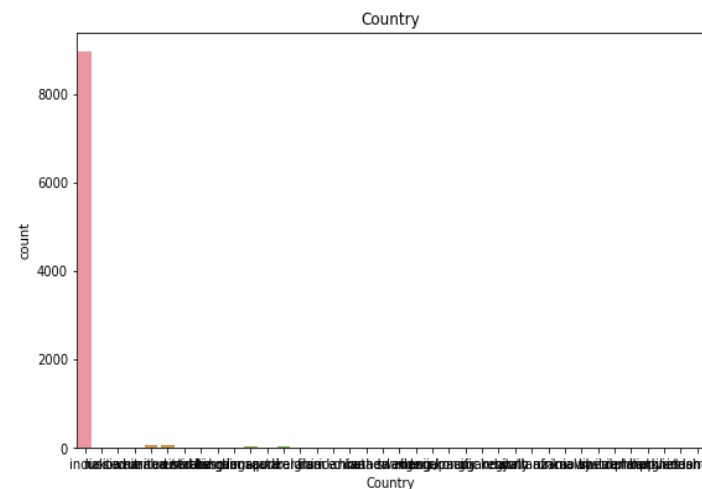
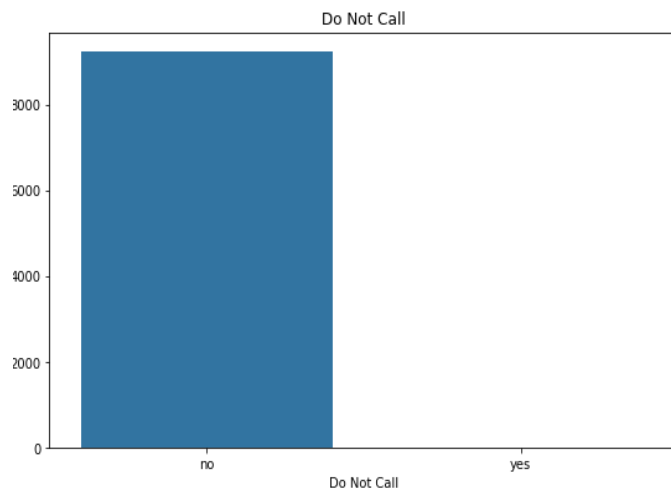
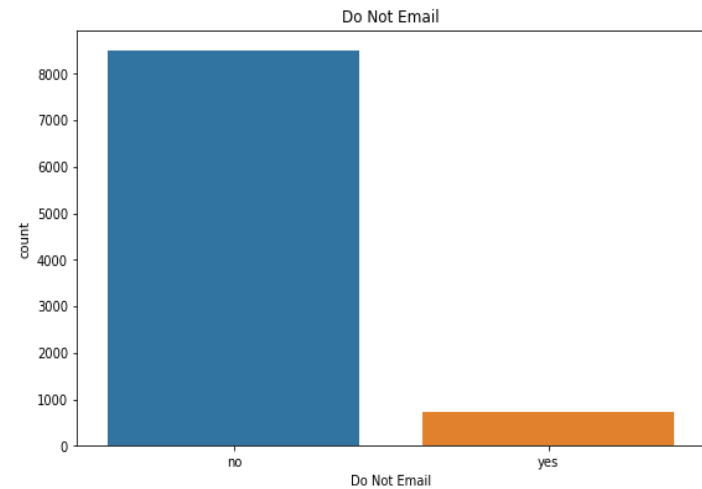
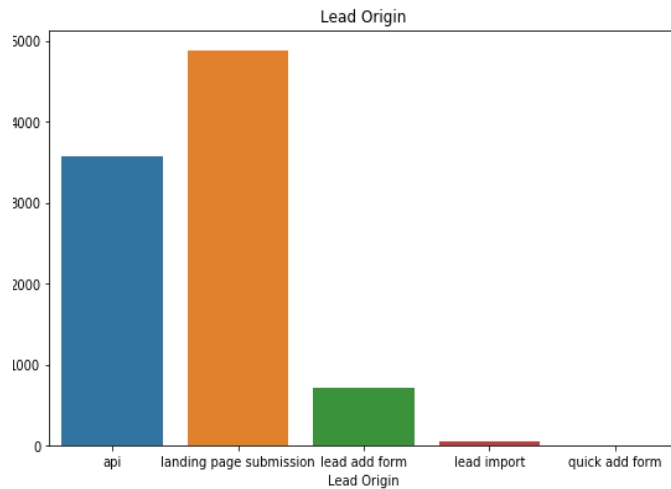
# Analysis Approach



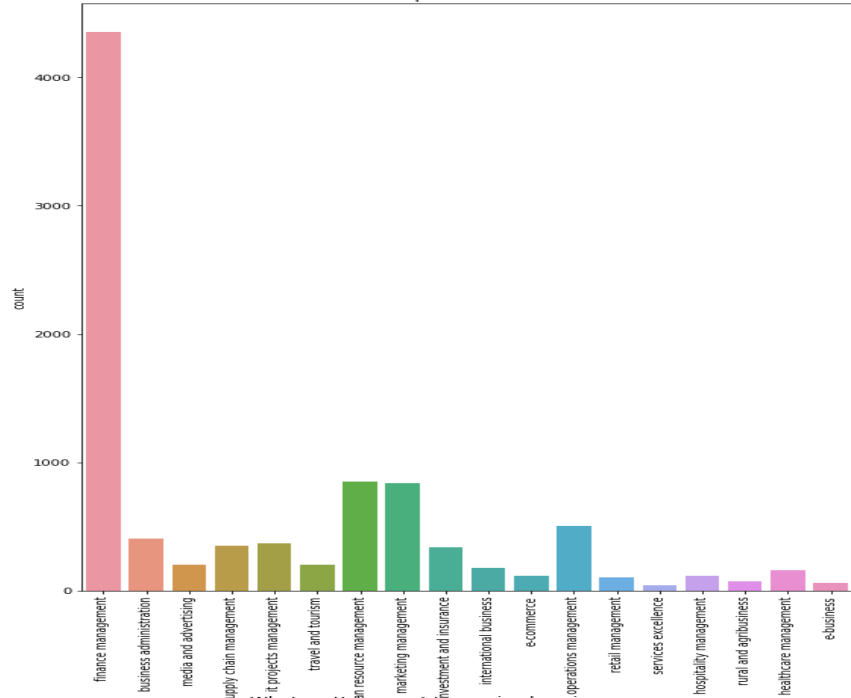
- Data Understanding
- Exploratory Data Analysis
- Feature Scaling
- Model Building
- Model Evaluation
- Conclusion

# EDA

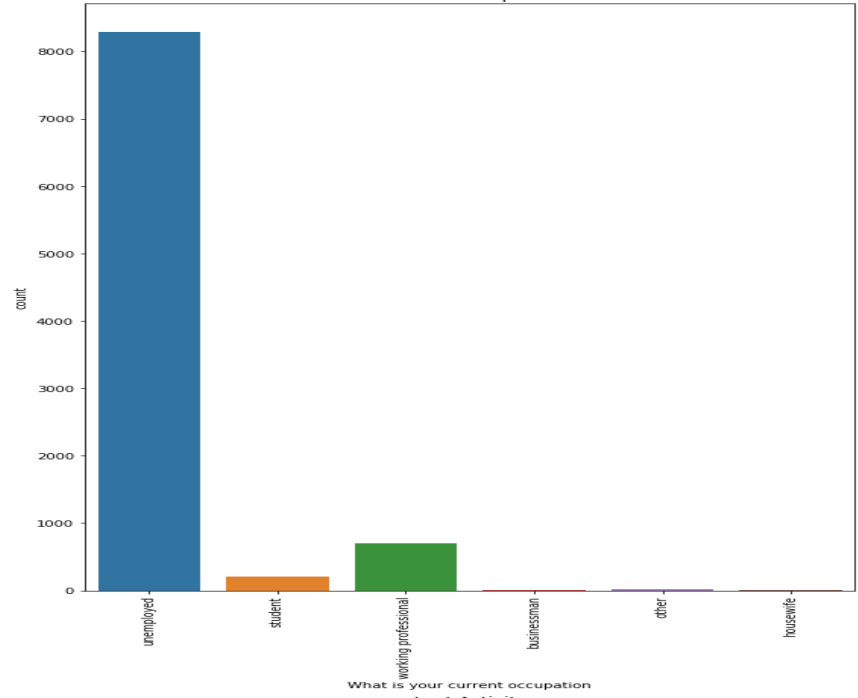
- First we checked for the missing values in data.
- Then we dropped the features having missing values greater than 30%.
- After that we impute the missing values.
- Univariate Analysis
- Bi-Variate Analysis



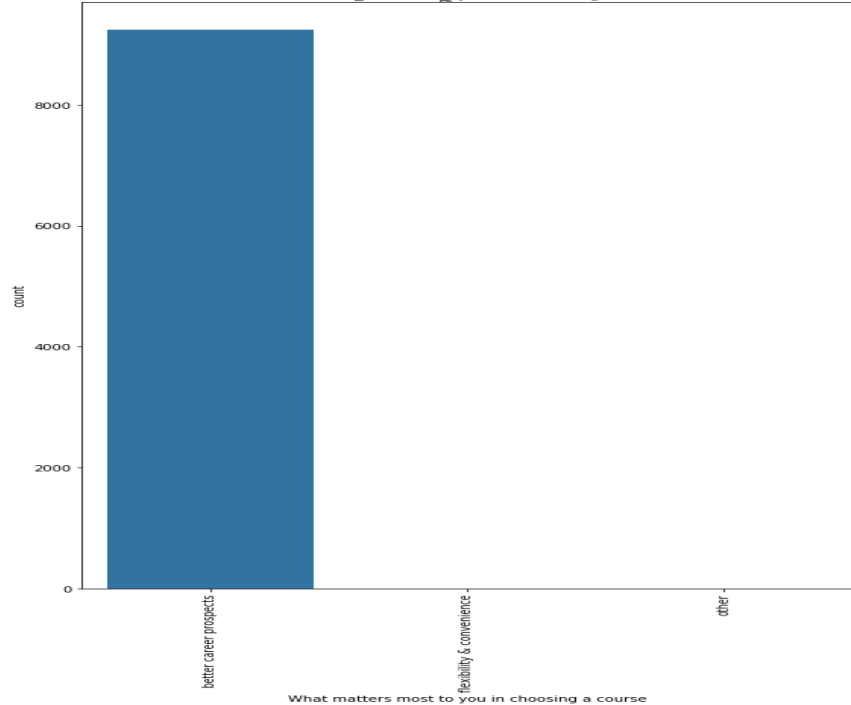
Specialization



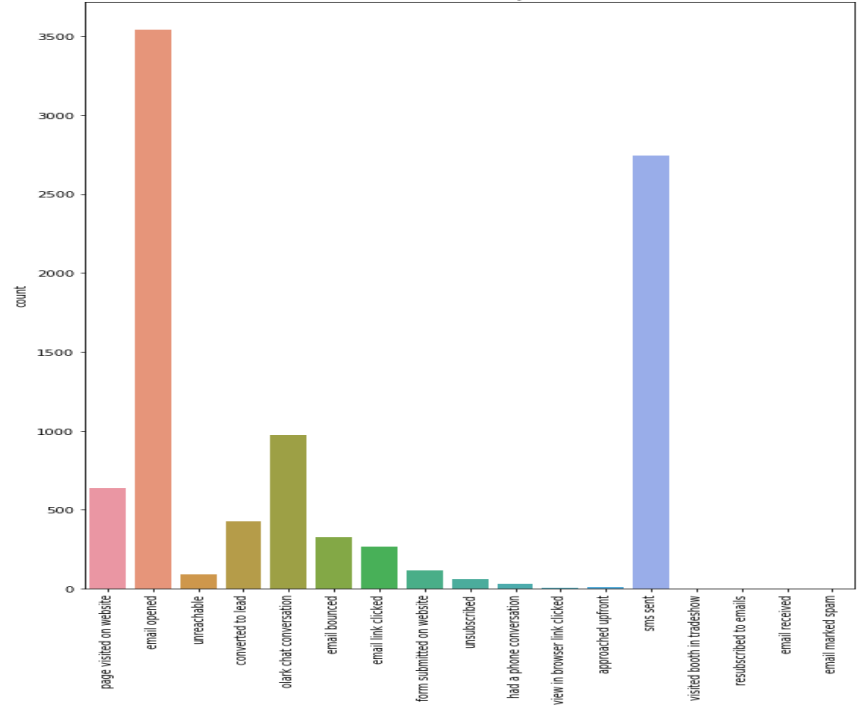
Current Occupation

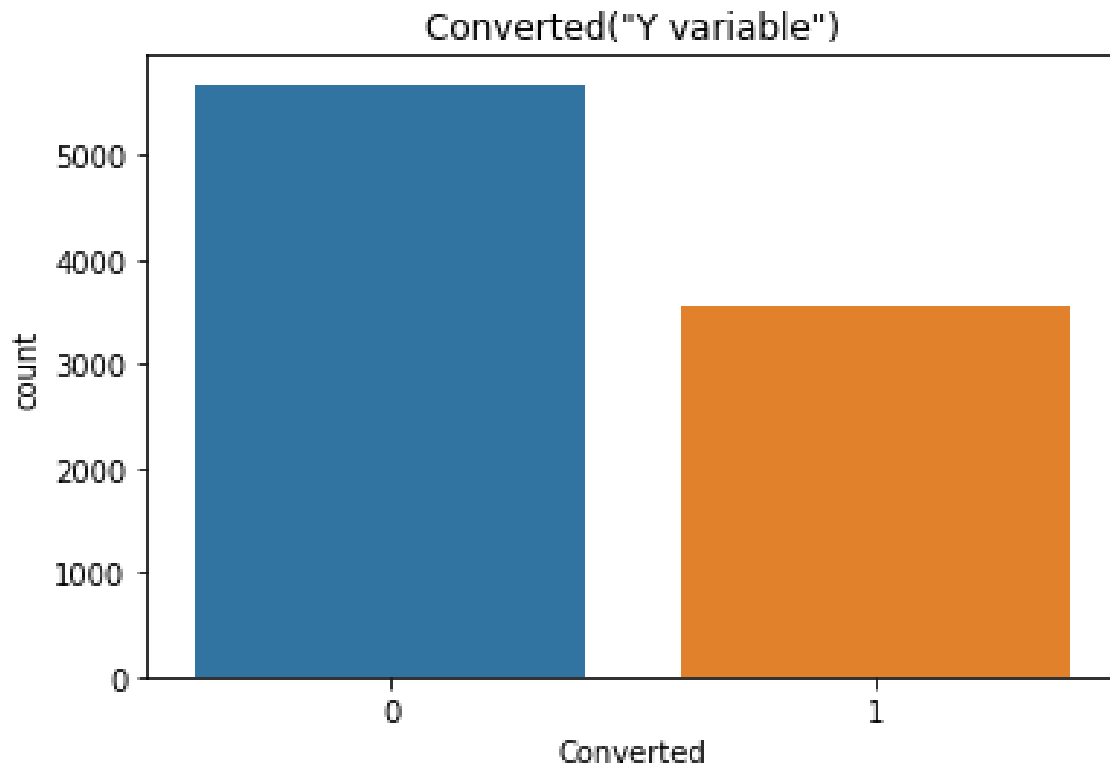


What matters most to you in choosing a course



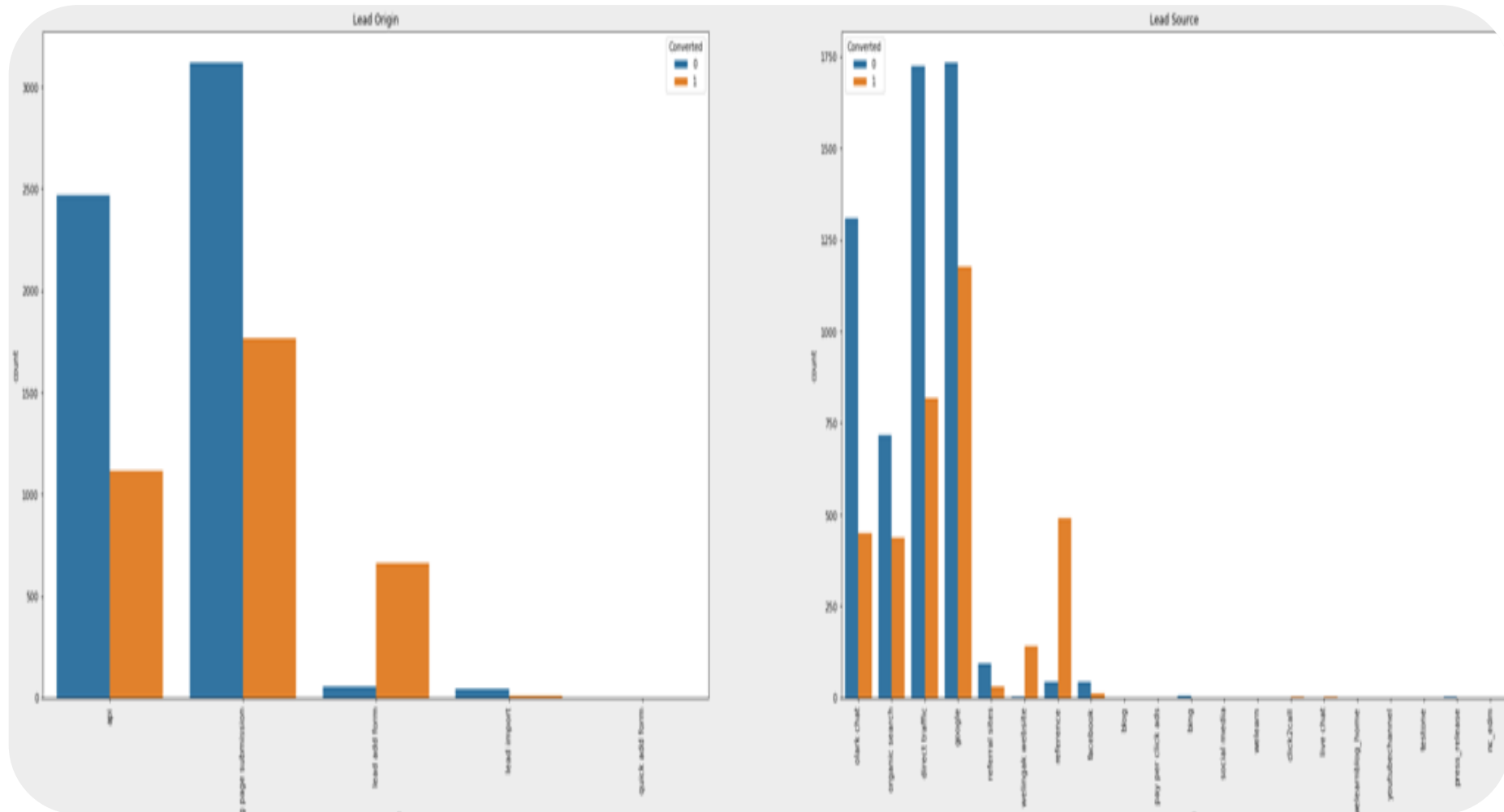
Last Activity



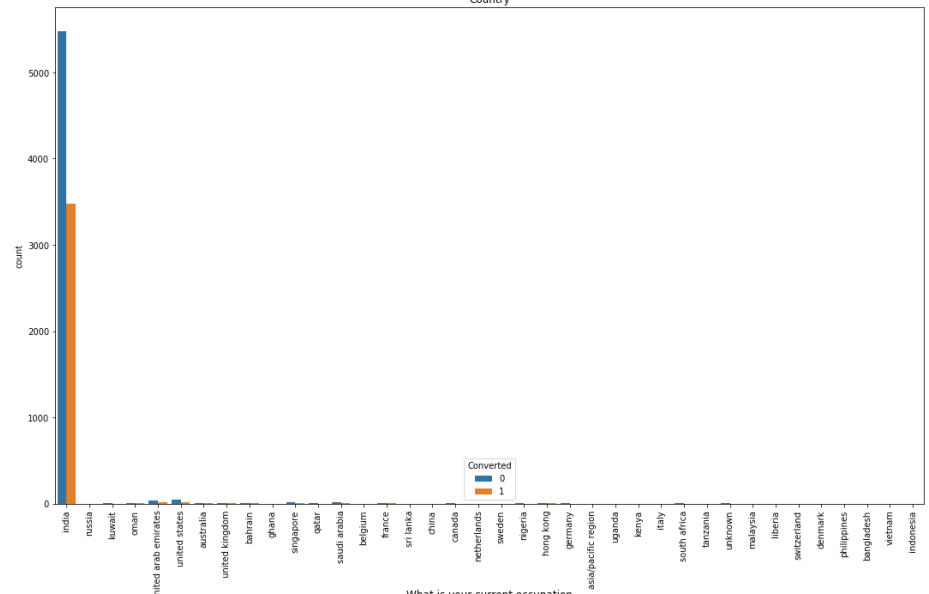
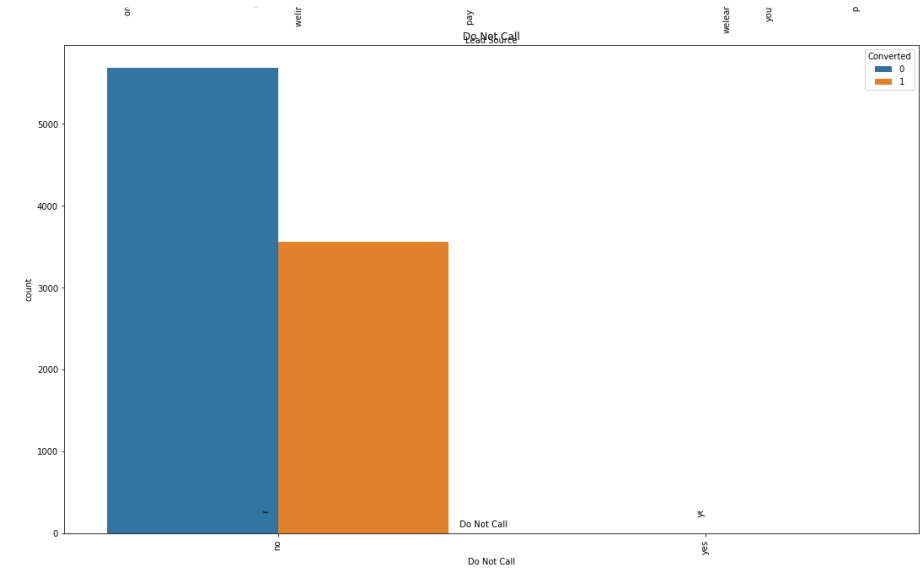
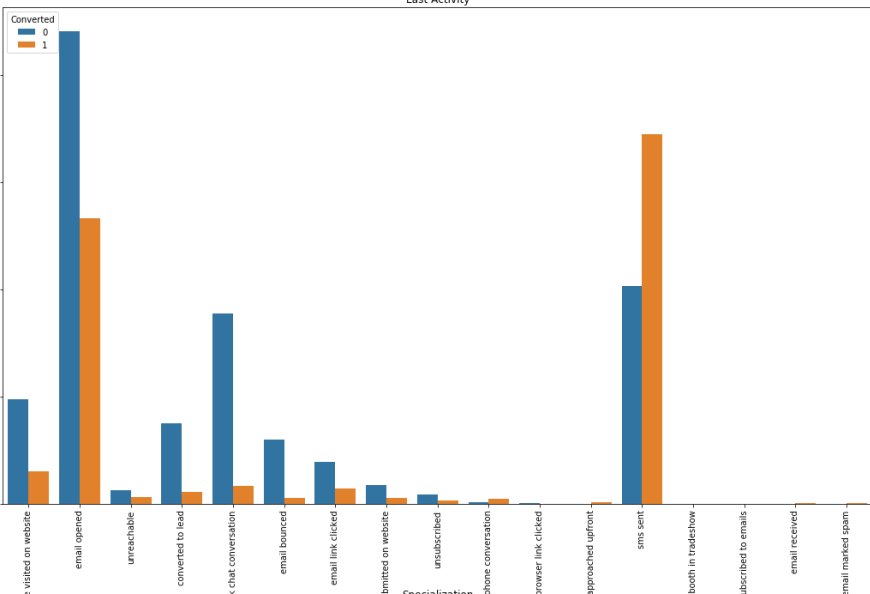
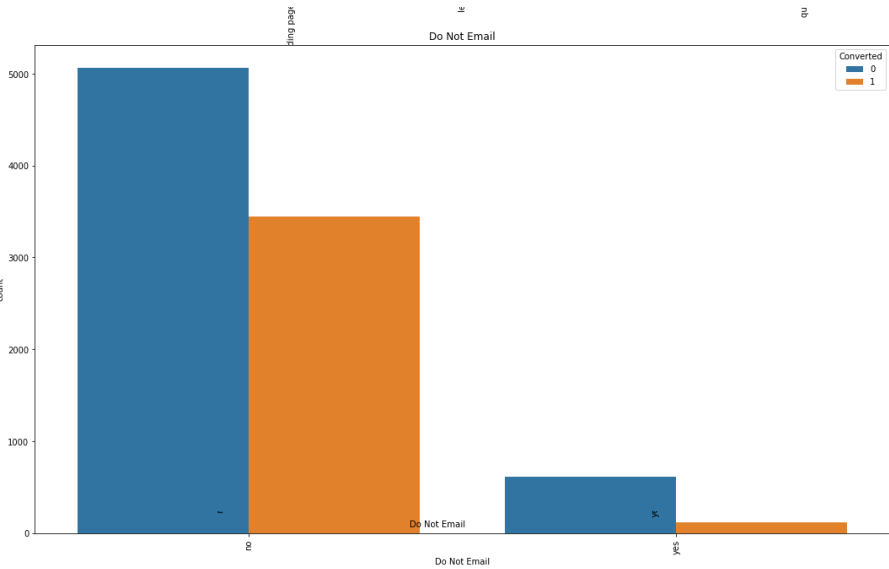


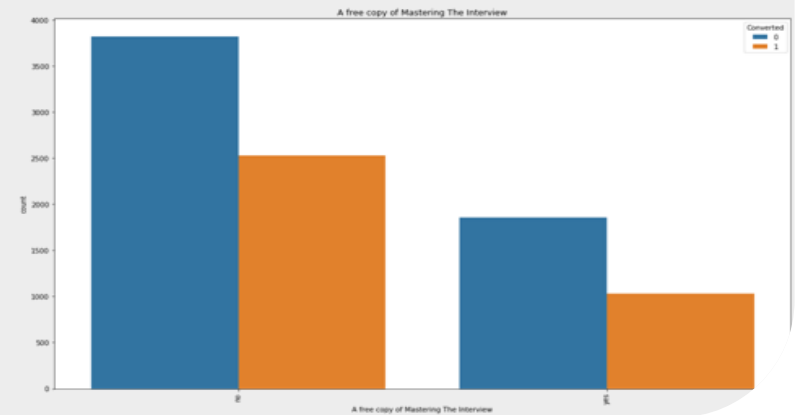
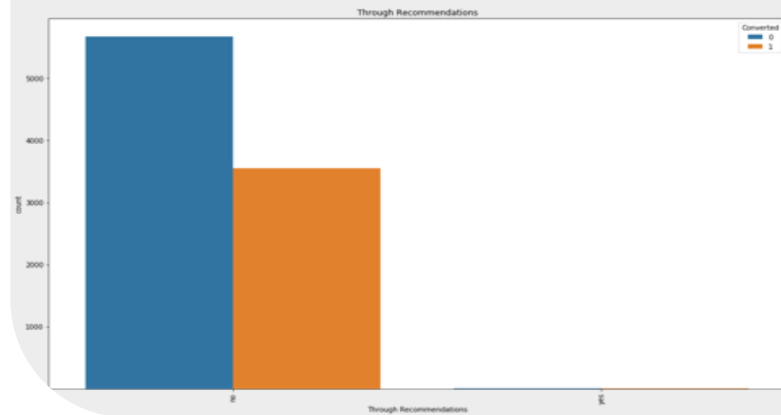
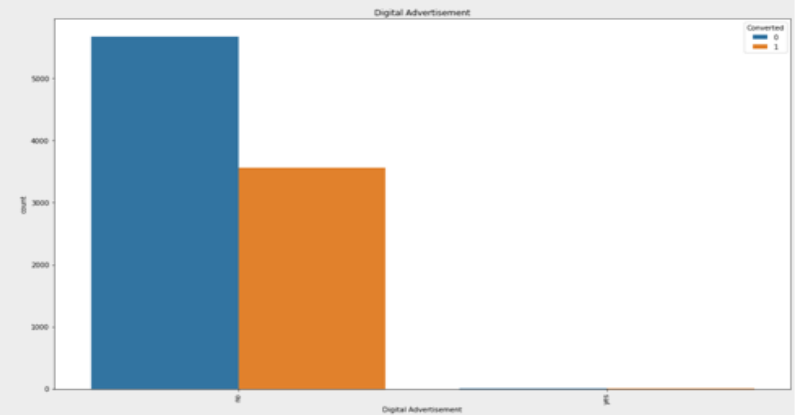
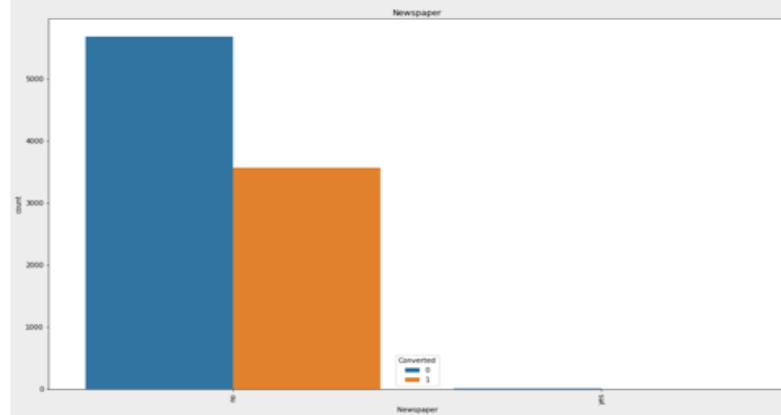
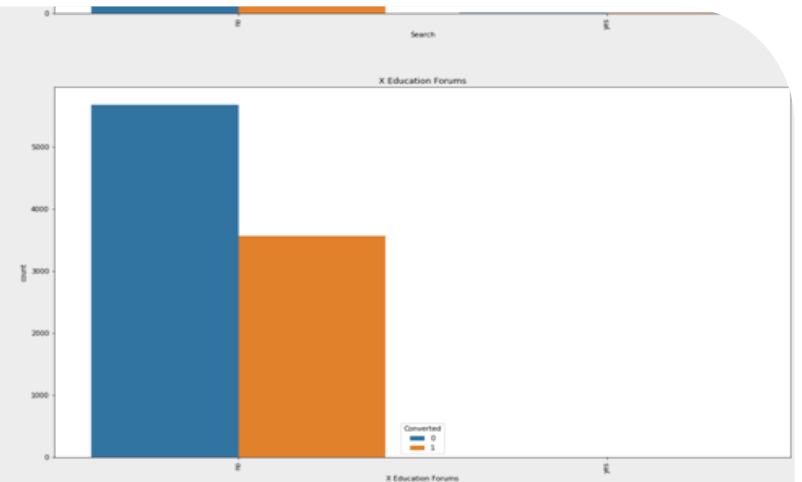
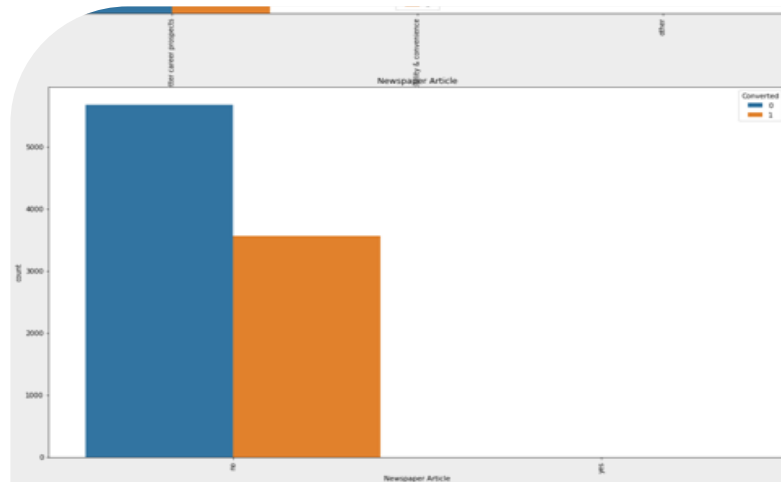
- The Count of Not converted lead is around 6000.

# Bivariate Analysis





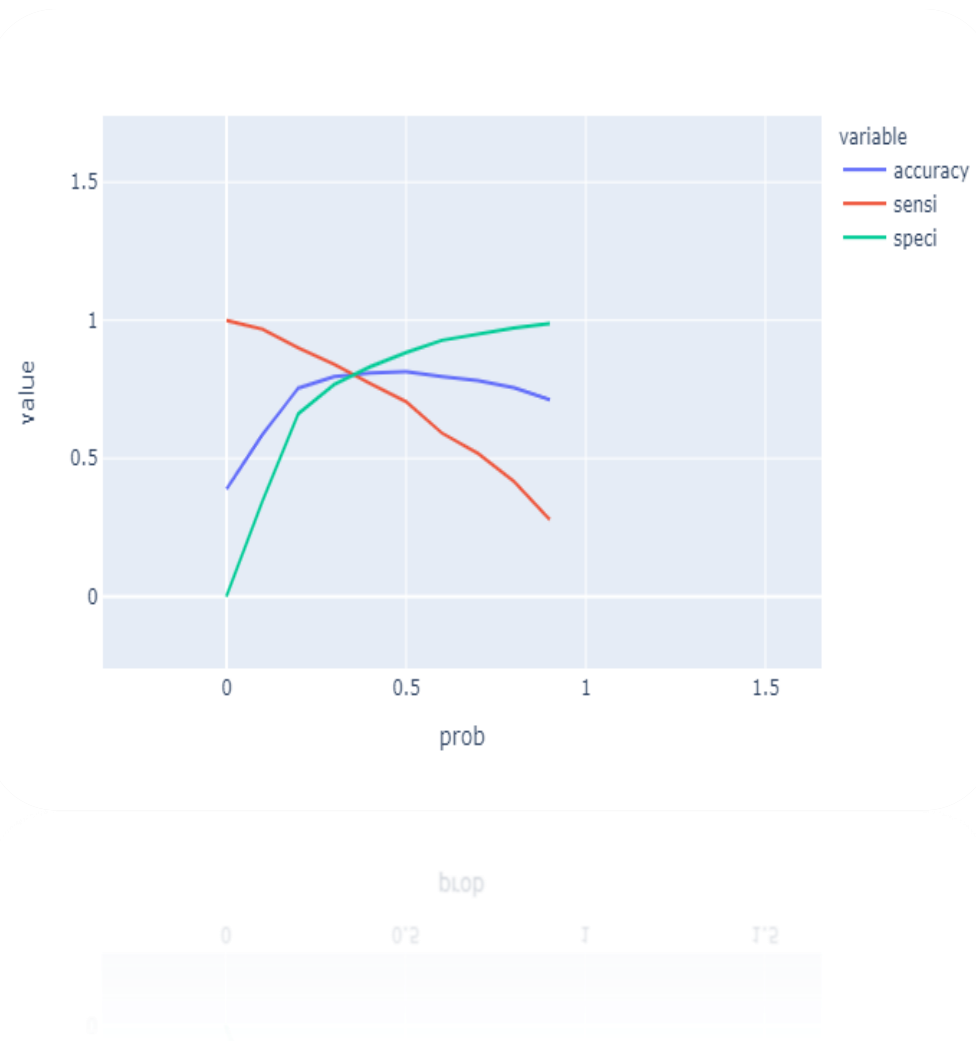
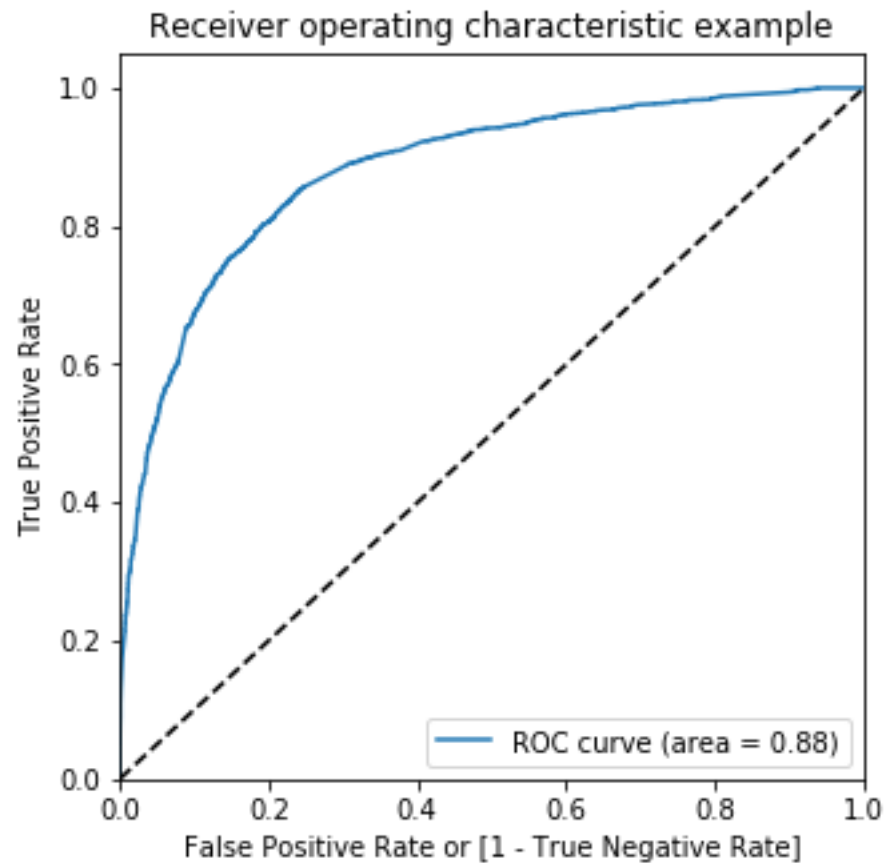




# Model Building

- Splitting the Data into Training and Testing Sets
- The first basic step for regression is performing a train-test split, we have chosen 70:30 ratio.
- Use RFE for Feature Selection
- Running RFE with 15 variables as output
- Building Model by removing the variable whose p-value is greater than 0.05 and vif value is greater than 5
- Predictions on test data set
- Overall accuracy 81%

# ROC CURVE & OPTIMAL CUTOFF



# For Lead Conversion X Company should focus on :

- Last activity performed by the customer includes Olark Chat Conversation.
- The last notable activity performed by the student would be modified, email opened, olark chat conversation, page visited on website and email link clicked.
- The total time spent by the customer on the website should be more.
- The total number of visits made by the customer on the website.
- The source of the lead were welingak website, Olark Chat.
- The origin identifier with which the customer was identified to be a lead includes only lead add form.
- Customer who is Working Professional
- "Do Not Mail is Yes" means customer want to be mail about the course.



THANK YOU