Lead Conversion

Understand the hidden insights behind leads not getting converted

Agenda

- Problem Statement & Target
- Approach
- Assumptions
- Univariate Analysis
- Bivariate Analysis
- Model Building Criteria
- Model Output
- Model Evaluation
- Conclusion
- Recommendations

Problem statement & Target

- Our company X Education is an EduTech company that aims in providing courses to customers belonging to any occupation
- Our company identifies the initial pool of leads through various marketing channels
- Although the company is able to identify potential leads in the business the lead conversion rate has been poor since a long time
- Lower lead conversion leads to lower revenue generation
- Hence, the overall aim is to identify potential leads also known as "Hot Leads"

80% Target

Approach

- Data importing
- Data Preprocessing and cleaning
  - Handling outliers, missing data, etc.
- Exploratory Data Analysis
  - Univariate, Bivariate Analysis
- Model Building
  - Feature scaling, feature selection, etc
- Model Evaluation
- Results

Assumptions

- Missing value imputation criteria
  - Remove columns if %(data) missing in the respective column > 30%
  - Remove rows if %(data) missing in the respective column < 5%
- Since Prospect ID and Lead Number are unique to a customer we have considered Lead Number as our index and dropped the Prospect ID during model building
- Remove the columns that have only one distinct category ex: Yes or No
- Data conversion : Yes → 1 and No → 0

Exploratory Data Analysis (EDA) – Univariate Analysis

- Data Imbalance Check
  - Lead conversion rate is low
  - Data imbalance absent as the contribution of both the categories are sufficient to carry out machine learning
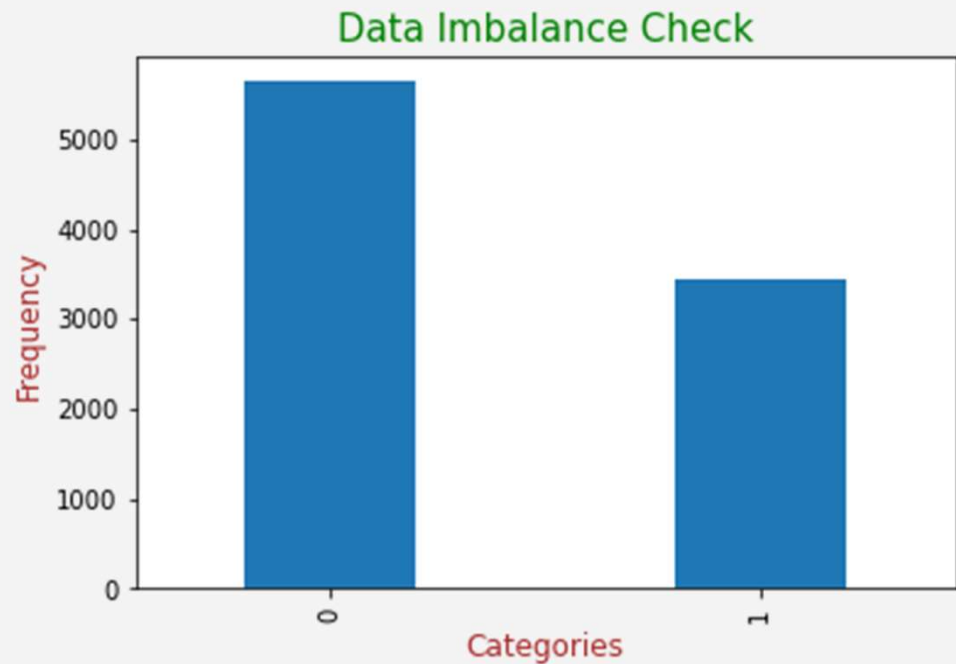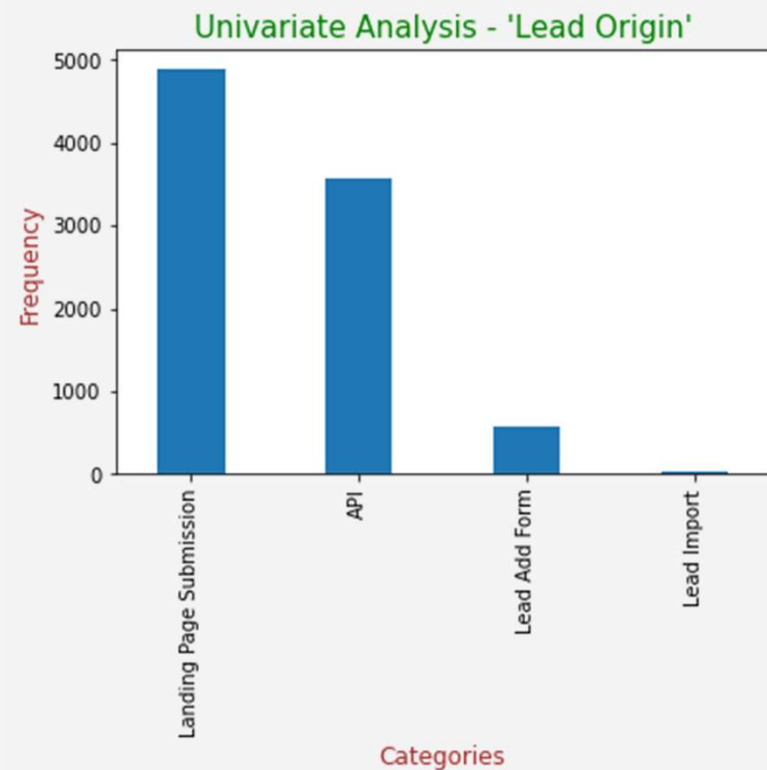
EDA – Univariate Analysis

- A higher percentage of people clicked on advertisements leading to a higher contribution of "Landing Page Submission" followed by API
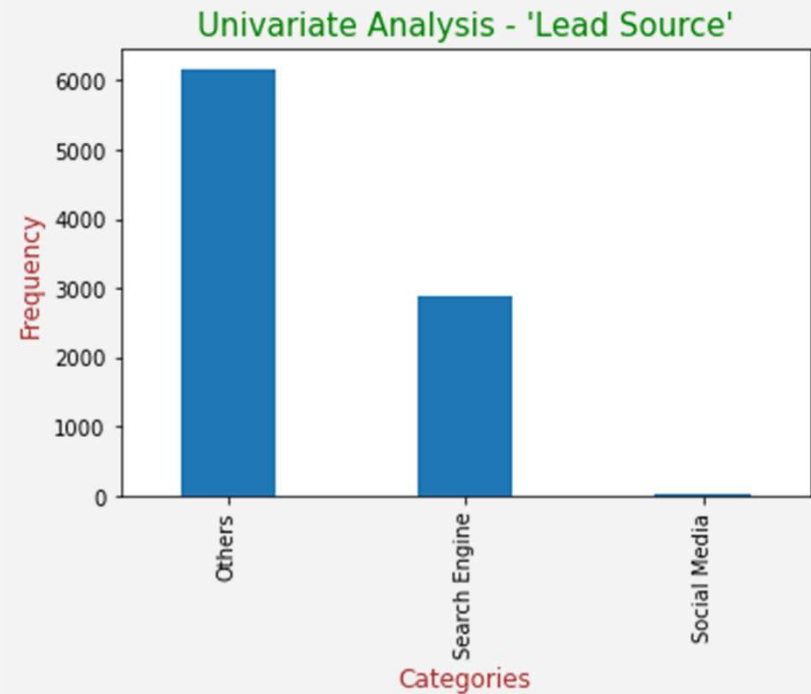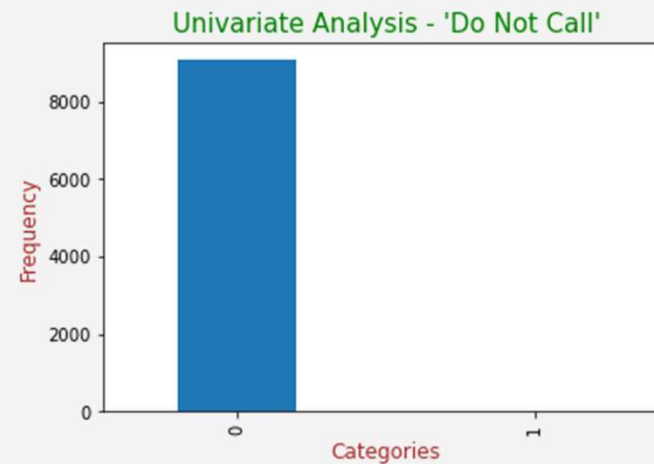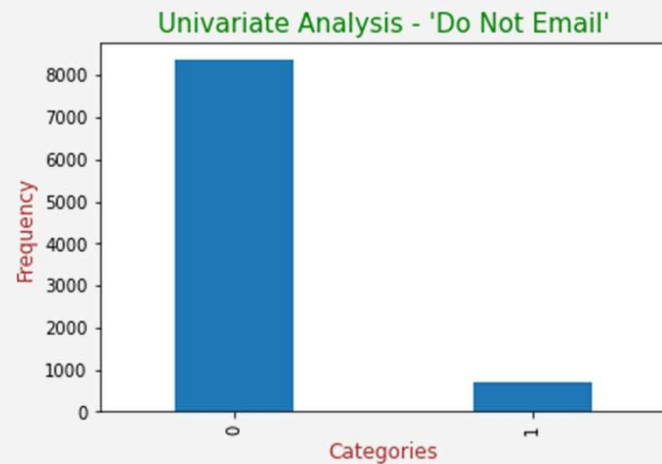
EDA – Univariate Analysis

- Very less percentage of people have been identified as lead through social media and web-search engine (Google, Bing)
- This suggests that the firm's website isn't listed in the top rankings in a user search history

Univariate Analysis - 'Lead Source'

EDA – Univariate Analysis

- Majority of the customers/prospects are interested in receiving calls and emails about the courses offered by the firm. This is a positive sign
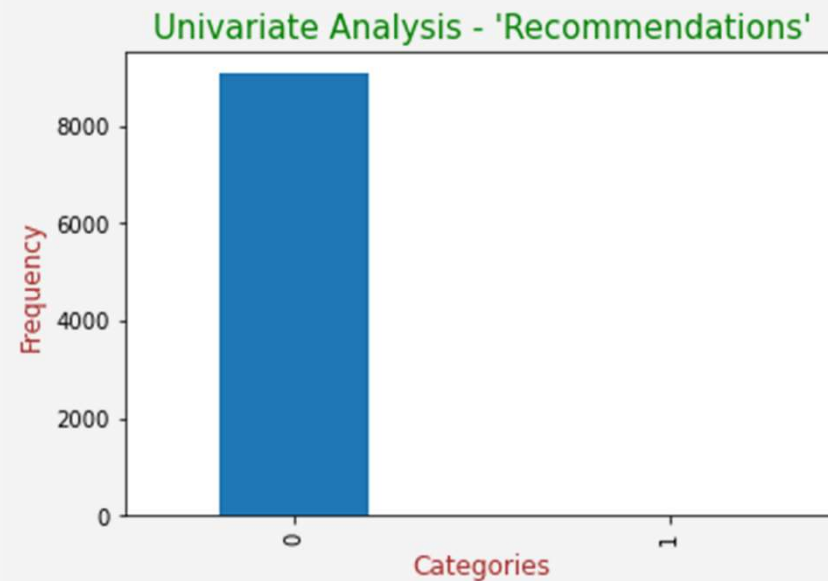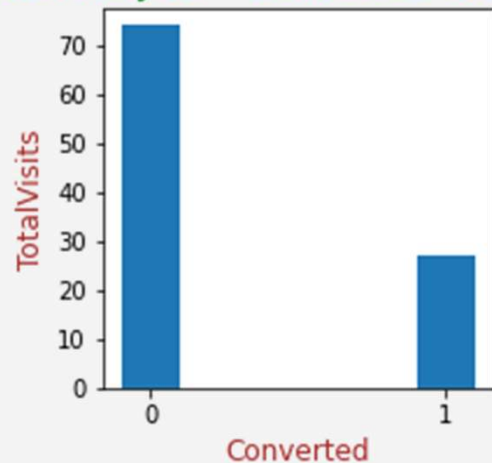
EDA – Univariate Analysis

- Majority of the identified prospects are new and have no background about the company and the courses offered.
- A possible scope of improvement is to provide identified prospects the experiences of other individuals
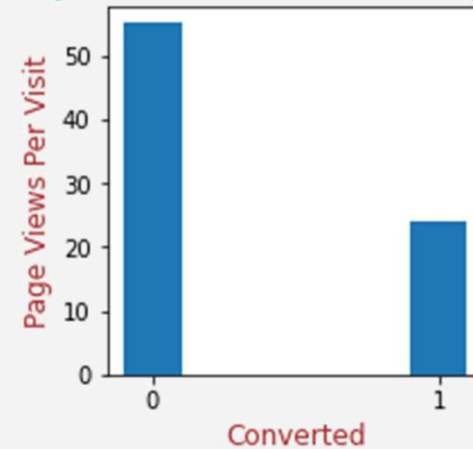
EDA – Bivariate Analysis

Bivariate Analysis - 'Converted' v/s 'TotalVisits'

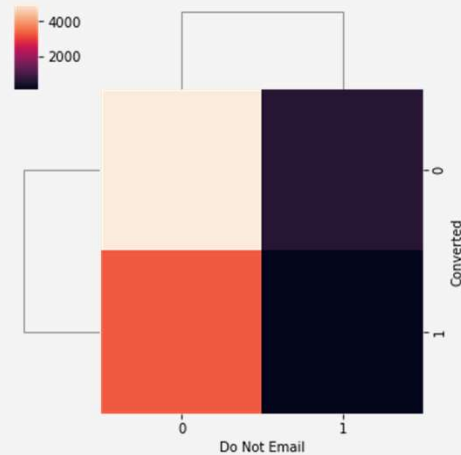Bivariate Analysis - 'Converted' v/s 'Page Views Per Visit'

- A higher number of visits to the website and number of page viewed suggests that the customer/prospect is not finding the suitable course and hence the conversion rate is lower
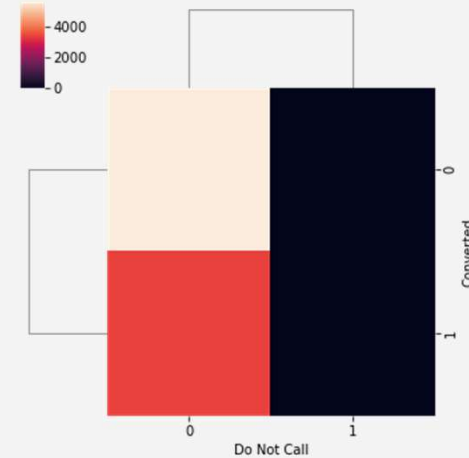
EDA – Bivariate Analysis

- Prospects who opted for receiving phone call or emails were majorly not converted. Probable cause – they were not satisfied

Model Building Criteria

- 70% data → training the model & 30% data → testing the model
- The concept of Standard Scaler was applied for feature scaling
- Columns that have correlation above 80% were dropped as the variance in one was explained well by other
- Feature selection was carried out by combining automation and manual techniques
- Feature elimination process was carried out by considering the p-value and Variance Inflation Factor (VIF)
  - VIF < 5.5 and p-value < 0.05 were fixed as acceptable thresholds

Model Output
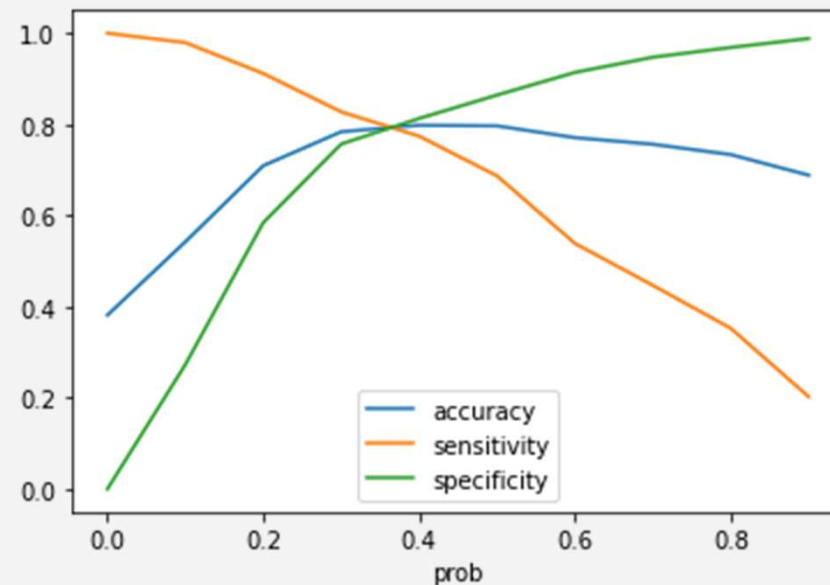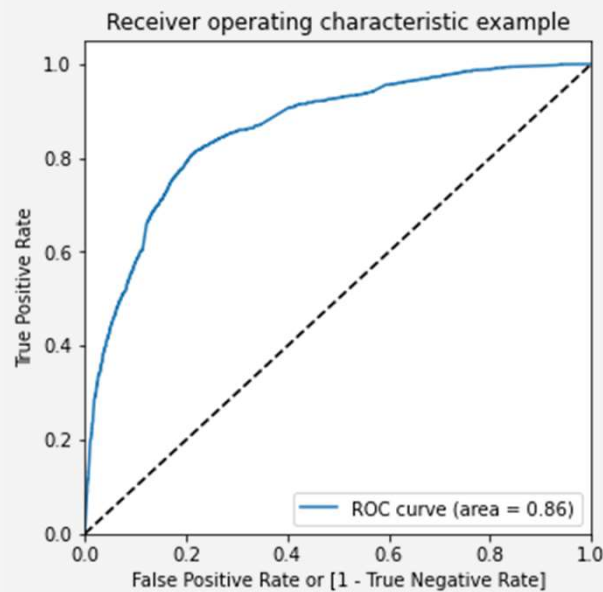
- The area covered under the ROC curve was 0.86
- The optimal cut-off point was decided as 0.35

Model Evaluation

Measure	Training Data	Testing Data
Sensitivity	80.80%	81.22%
Specificity	78.71%	77.89%
Overall Accuracy	80%	80%
False Positive Rate	21.28%	22.10%

Conclusion

- The overall accuracy of the model remained the same through testing and training dataset and a much drastic difference in the accuracy was not observed. This ensures that the model is stable and reliable.

- The false positive rate is low which is a positive sign

- The sensitivity of the model is approximately 81%

- For future purpose the model can be fine-tuned by adding/removing features thereby increasing the accuracy to a higher level, but since the training and testing dataset are giving the same level of accuracy we finalized the model

Recommendations

- SEO optimization should be carried out as the ranking of the website is not at top rankings in the search history
- Maximum of the identified prospects have not come through recommendations. This is a possible scope of improvement. If the already converted leads are given some compensation on referrals that will increase the recommendations count
- Customers/prospects who are not interested in receiving emails or phone calls need to be contacted to know the exact reason and issue
- Company should focus on marketing using social media and search engines

X – Education Company

Thank You