



Lead Score Case Study

Problem Statement

- **X Education sells online courses to industry professionals.**
- **The company receives so many leads, but the lead conversion rate is very poor.**
- **The aim is that the company has to identify the most potential leads, also known as 'Hot Leads', so that the focus can be made on such leads, and thus avoiding making calls to non-materializing leads.**

Objectives

- **Build a model Identify promising leads**
- **Reuse the model for future use**

Approach

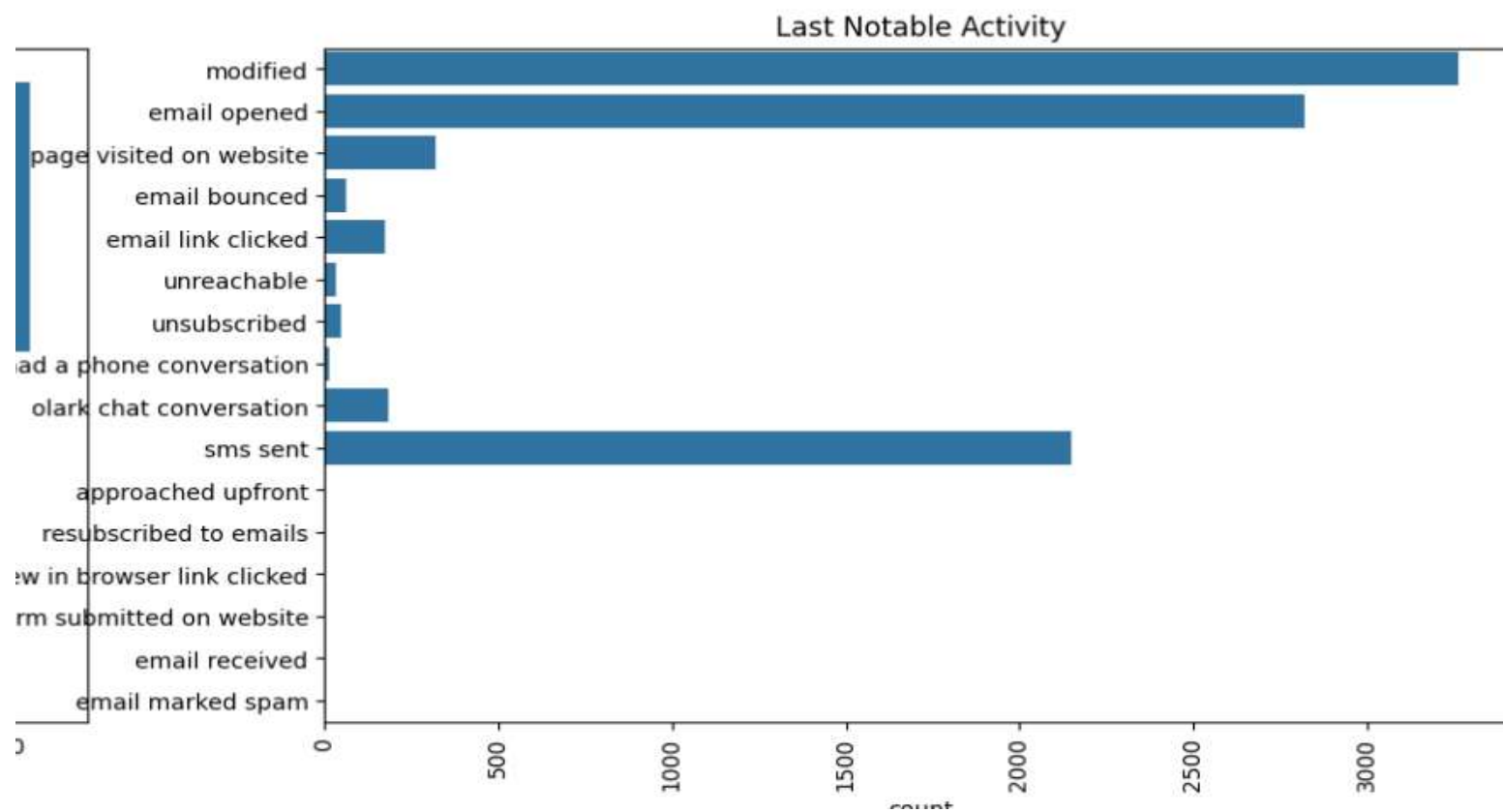
- **Load & Cleanse the data**
- **Handle duplicate rows**
- **Perform null value handling**
- **Identify and drop irrelevant columns**
- **Identify outliers**

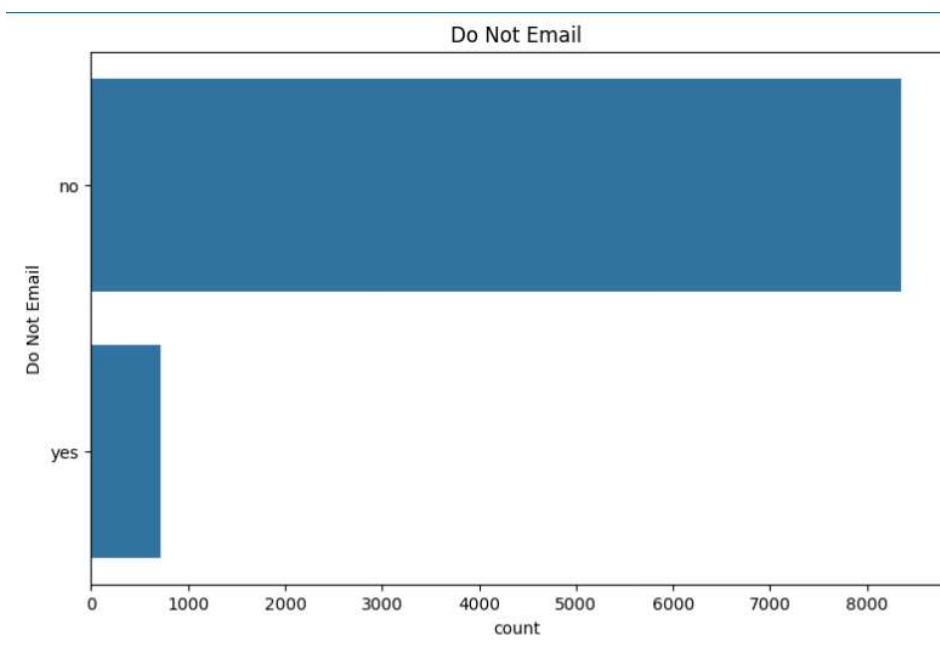
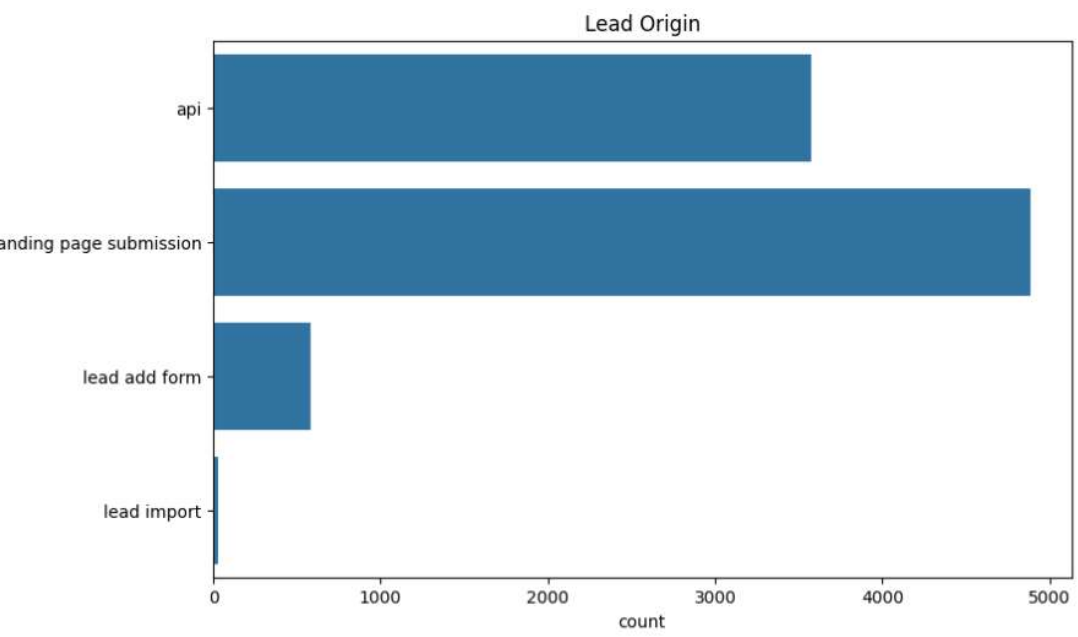
Approach

- **Load & Cleanse the data**
- **Handle duplicate rows**
- **Perform null value handling**
- **Identify and drop irrelevant columns**
- **Identify outliers**

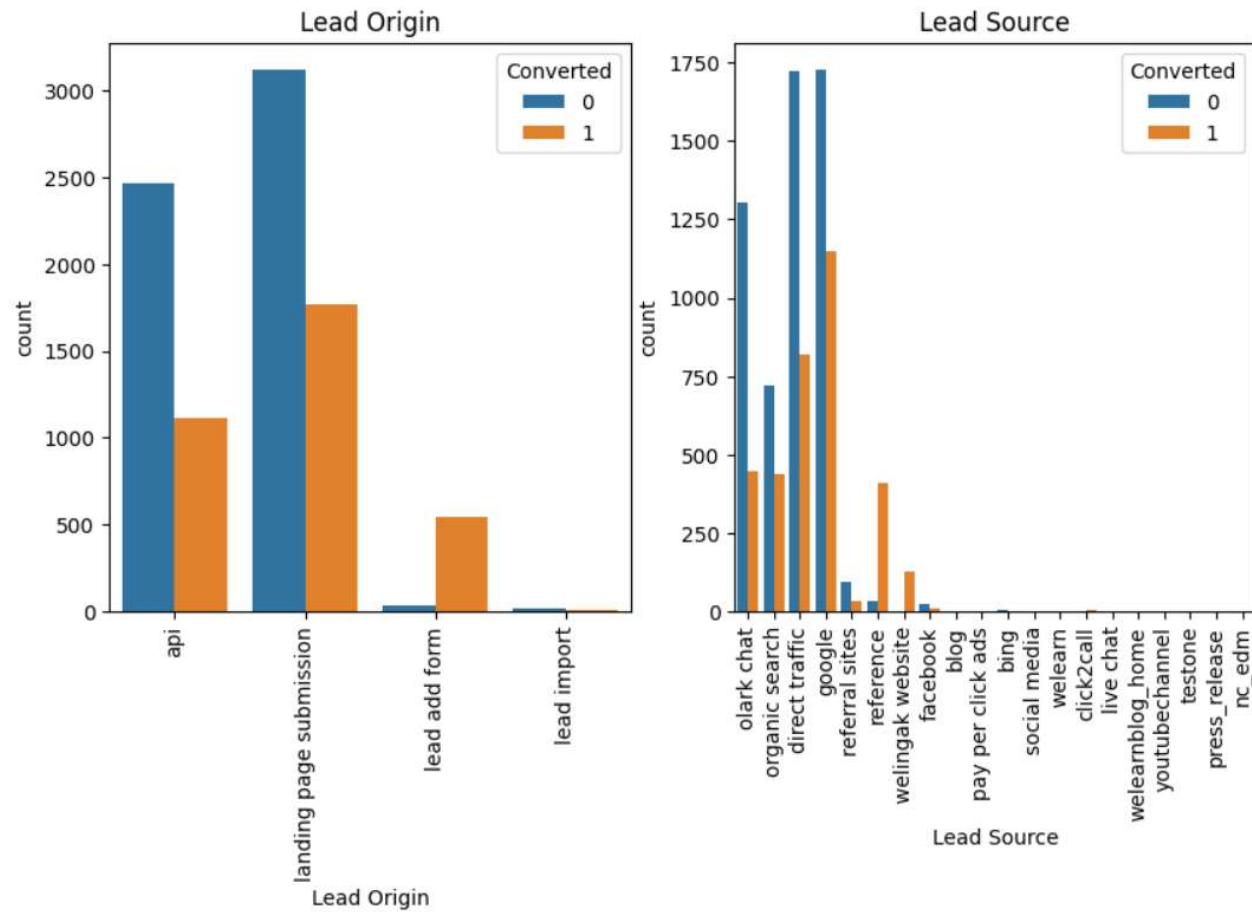
- **EDA – Univariate & Bivariate analysis**
- **Feature Scaling**
- **Logistic Regression**
- **Model Validation**
- **Model Presentation**

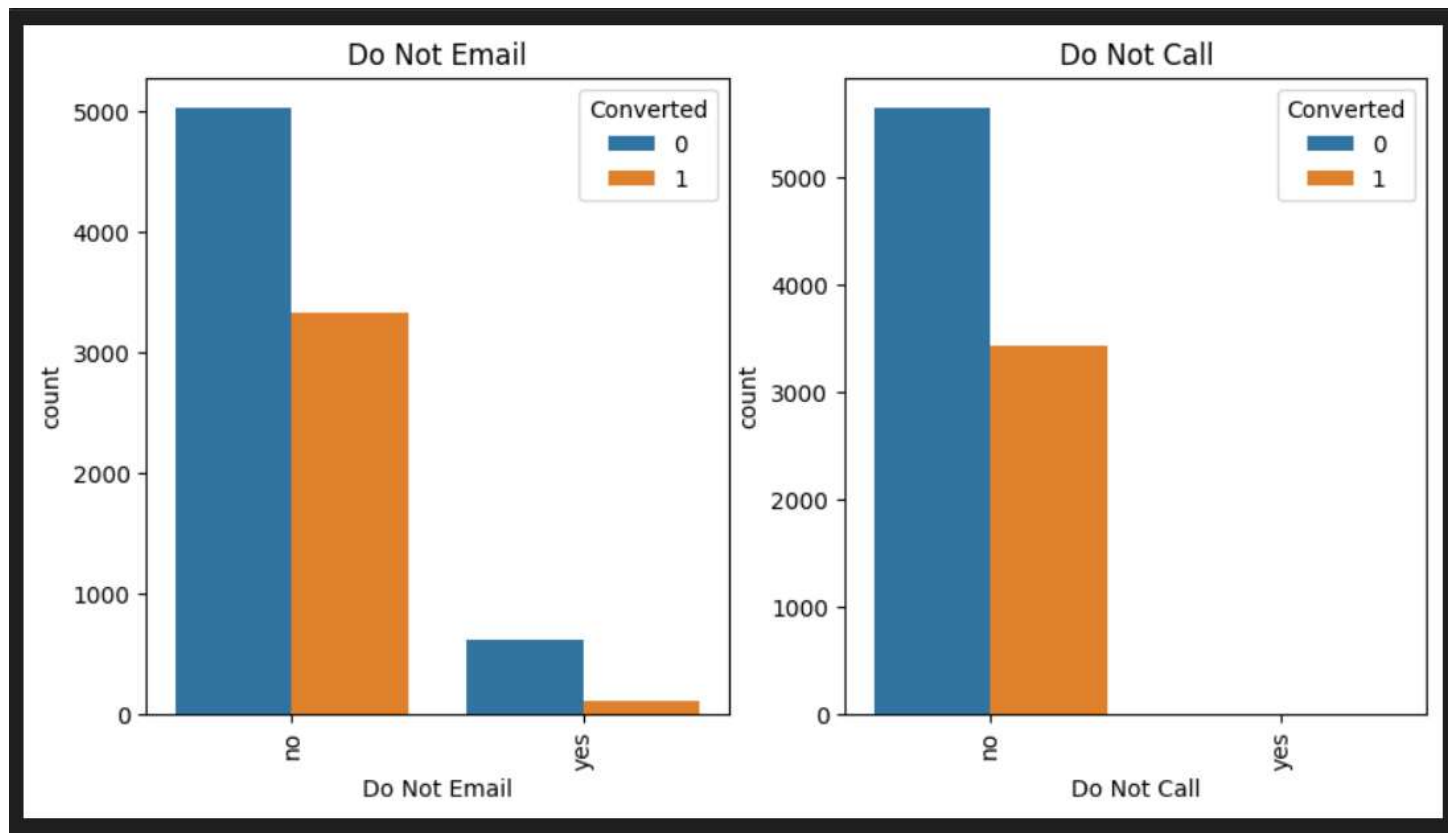
EDA





Relating Categorical Variables

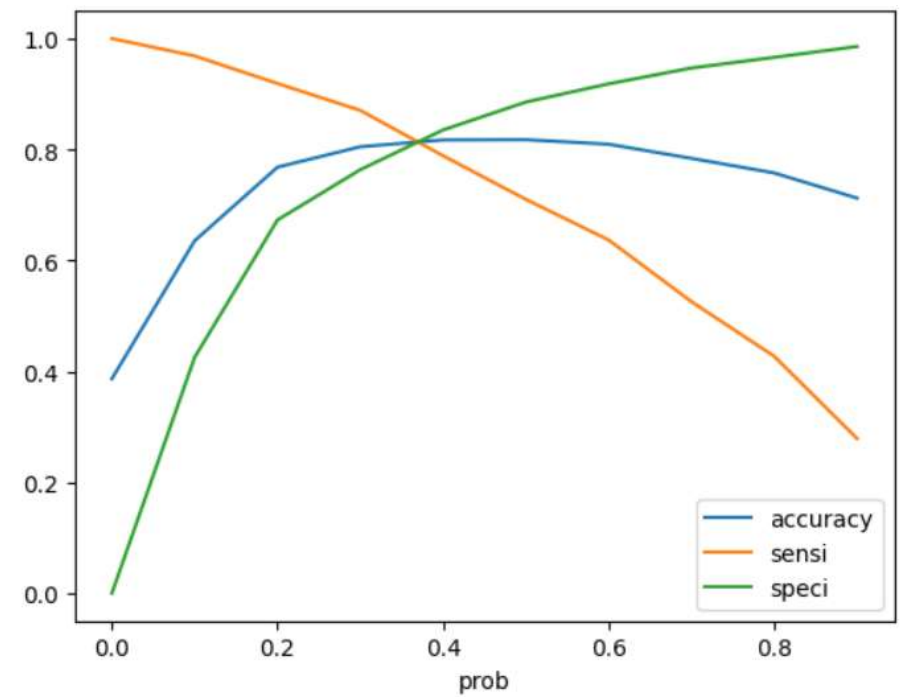
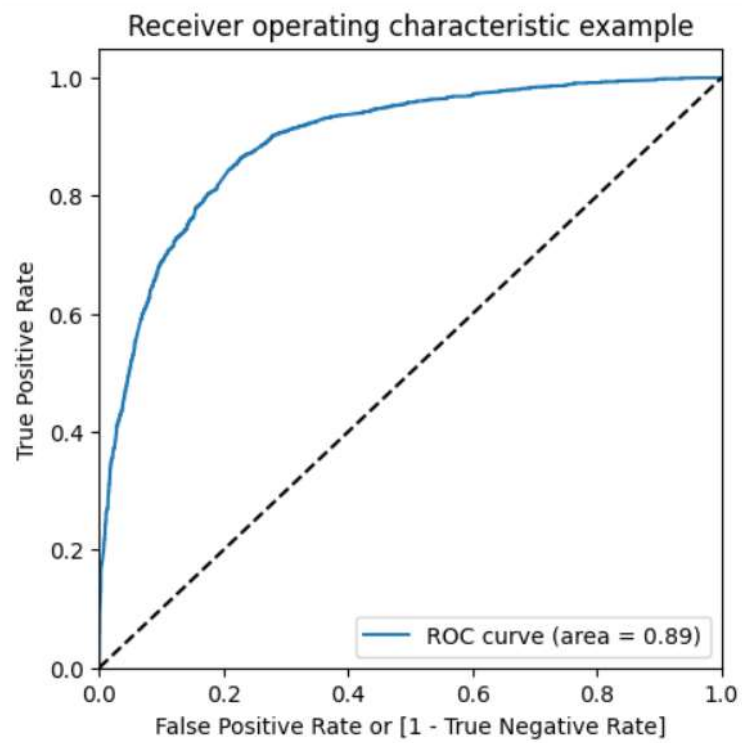




Building the Model

- Split the data into Test and Training sets
- The ratio of the above split was 70:30 %
- RFE was used for feature selection
- Removed the variables that were irrelevant
- TestData set prediction showed an accuracy of 82%

ROC



Conclusion

- **Most significant variables were**
 - **Total time spent on the website**
 - **Total number of visits**
 - **Lead source were**
 - **Organic Search**
 - **Google**
 - **Direct Traffic**
 - **Last activity SMS/Chat Conversation**
 - **Working professionals**