Case Study – Lead Scoring

Case Statement

- An education company sells online courses to industry professionals.
- Even though they acquire a lot of leads through different sources, their conversion rate is significantly low, around 30%.
- 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.

Business Objective

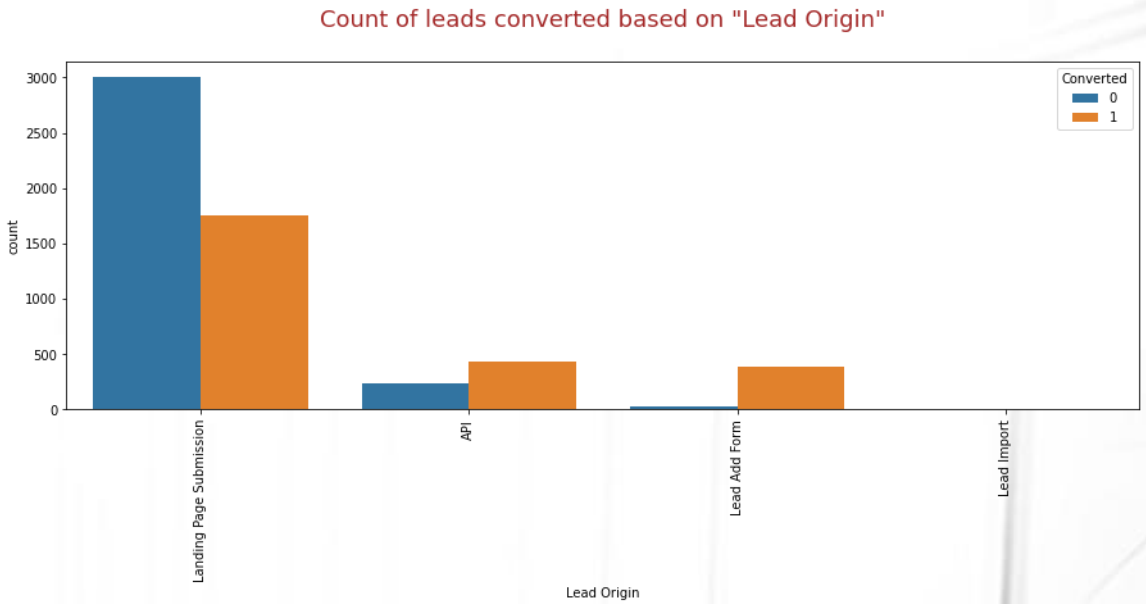
- The education company wants to identify the most promising leads.
- In order, they request to build a predictive model.
- The model will be deployed in future.

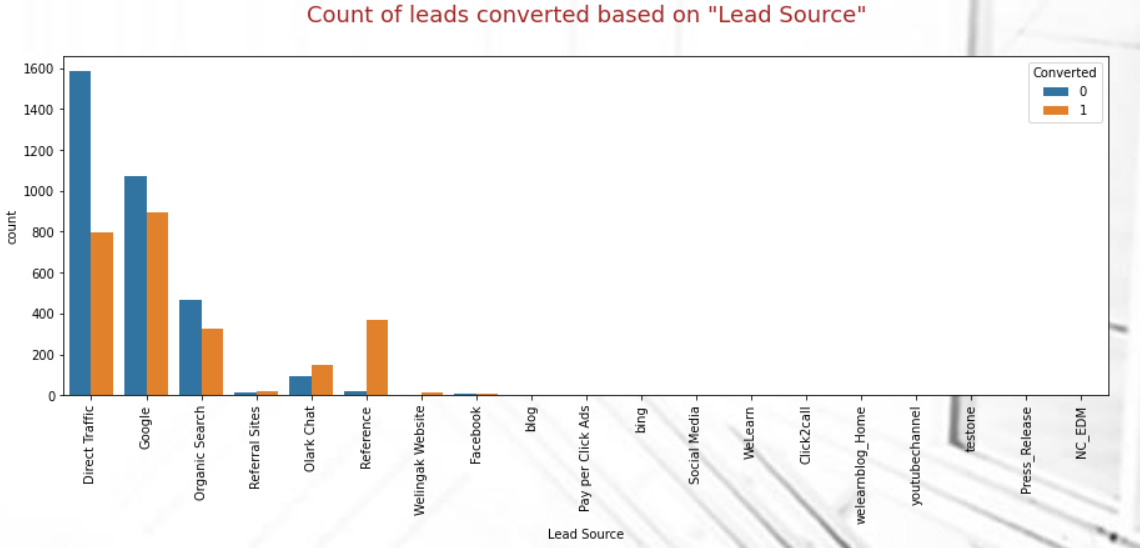
Approach

- Data cleaning and data manipulation.
  1. Check and handle duplicate data.
  2. Check and handle NA values and missing values.
  3. Drop columns, if it contains large amount of missing values and not useful for the analysis.
  4. Imputation of the values, if necessary.
  5. Check and handle outliers in data.
- EDA
  1. Analysis of categorical variables and checking their conversion count.
  2. Analysis of numerical variables, correlations.
- 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.

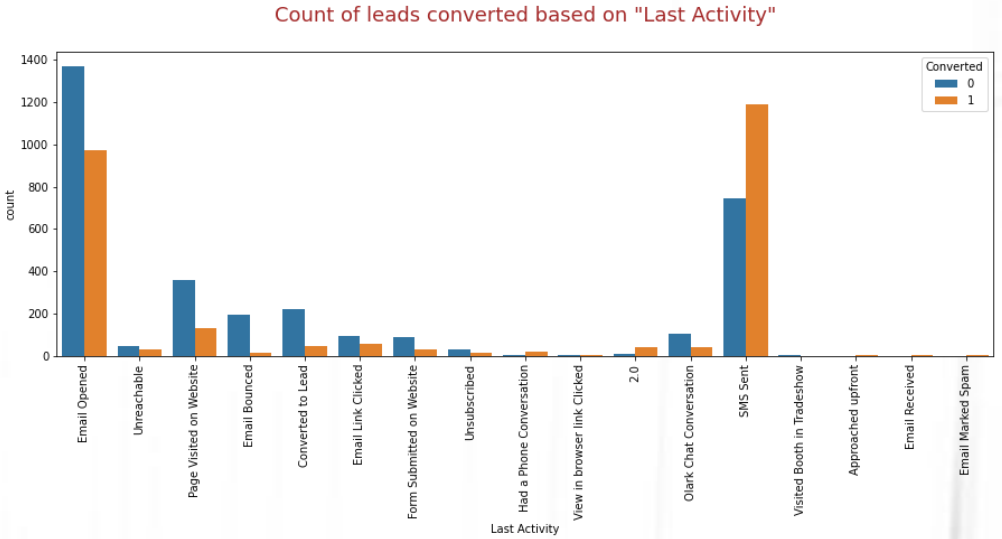
EDA

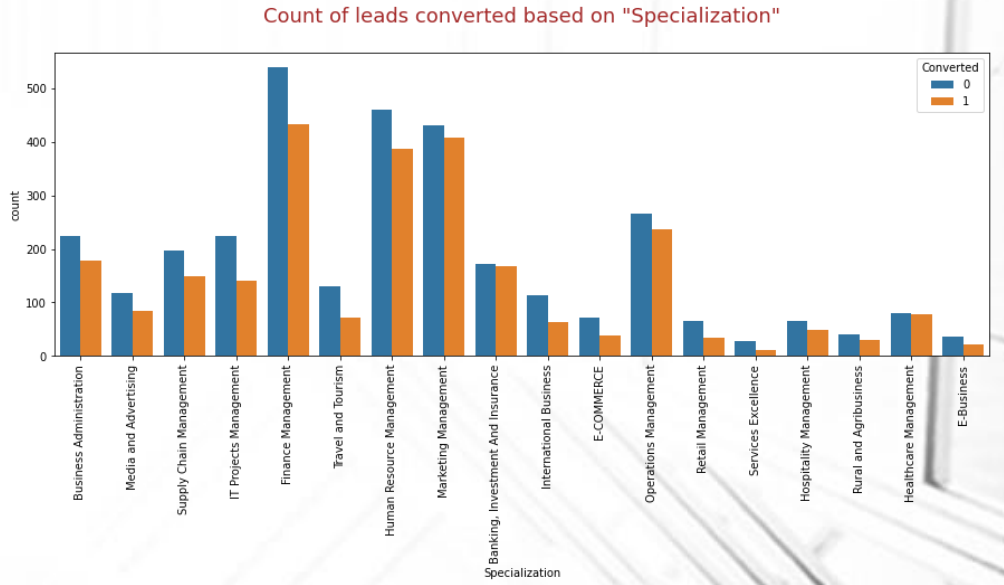
- Most of the leads identified are from 'Landing page submission`.
- Leads identified from `API`, `Lead Add Form` has high conversion rate.
- There are no leads identified from `Lead Import`.
- Highest leads are obtained from `Direct Traffic` & `Google`.
- Leads obtained from `Reference` has highest conversion rate.



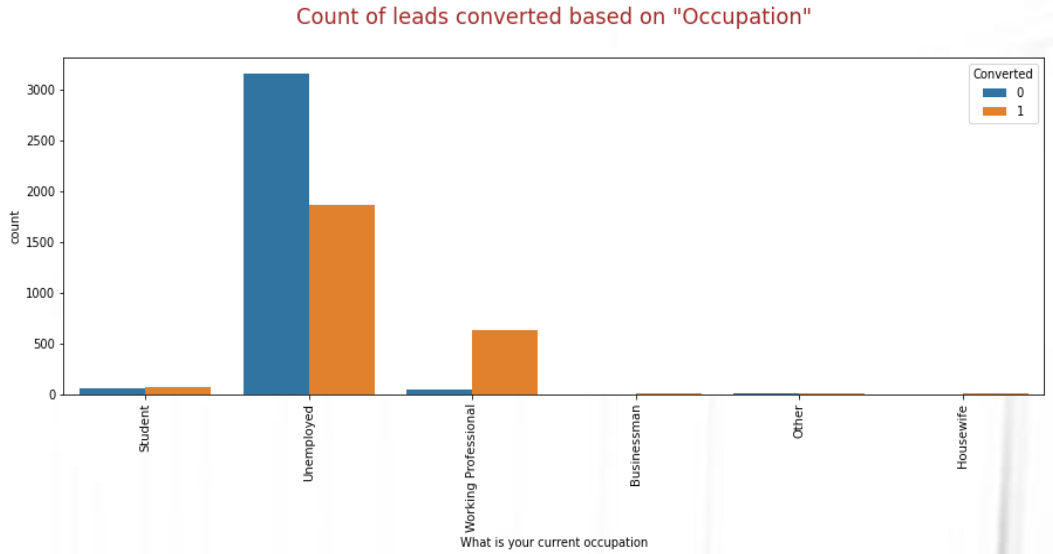


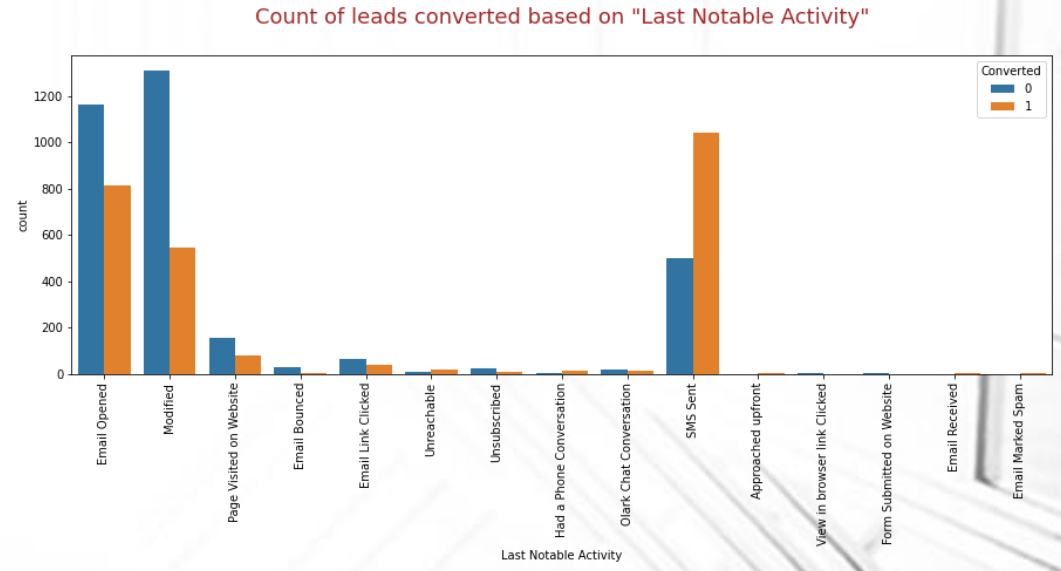
- It's seen that `Email Opened` & `SMS Sent` are most performed last activities.
- `SMS
- It can be seen that most leads are from `x-Management` sectors.
- There is no clear insight of conversion rate though based on Sectors in which leads worked before. Sent` has highest conversion rate.





- `Working Professional` are the most potential leads.
- Most leads are `Unemployed` but conversion rate is low.
- `Student` are equally likely converted and non-converted.
- It's seen that `Email Opened`, `Modified` & `SMS Sent` are most notable last activities performed.
- `SMS Sent` has highest conversion rate.
- Though not sure about `Modified` activity, still it has lowest conversion rate.

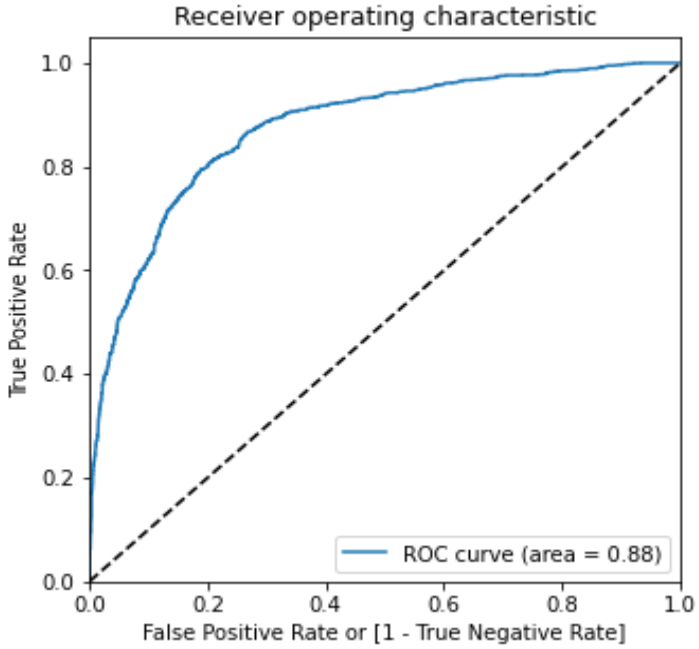


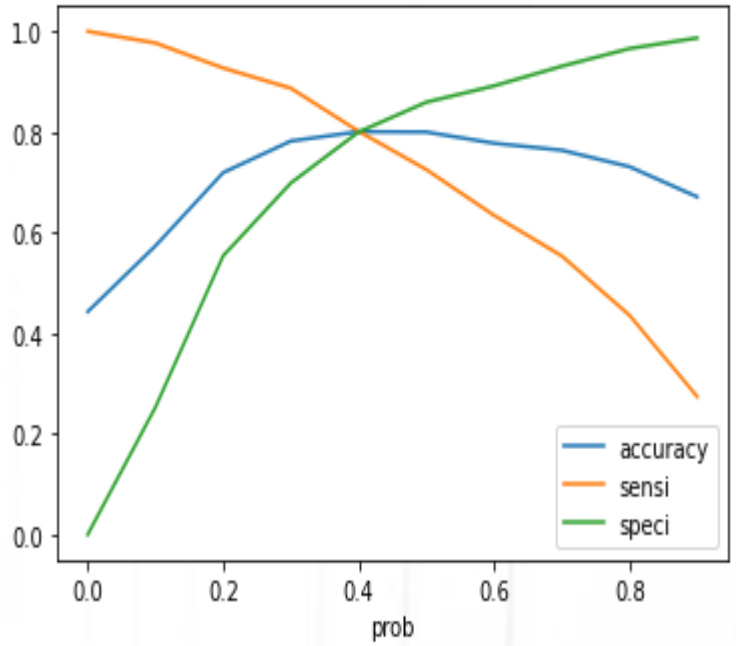


Model Building

- 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 .
- 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 80%

Optimization





- From the above left ROC curve it's evident that the model is stable with good predicting ability.
- The right plot gives the optimal threshold for predictions which for us is 0.4.

Observations

- The following are variables in descending order of importance which influence our business problem and model:

Last Notable Activity	Had a Phone Conversation
Lead Origin	Lead Add Form
What is your current occupation	Working Professional
Last Activity	Unsubscribed
Last Notable Activity	Unreachable
Do Not Email	
Last Activity	SMS Sent
Total Time Spent on Website	
Last Notable Activity	Modified
Last Activity	Email Opened

Train data Metrics

Accuracy	80 %
Sensitivity	80 %
Specificity	80 %

Test data Metrics

Accuracy	80 %
Sensitivity	80 %
Specificity	80 %

Recommendations

- Based on the observations the recommended strategy is:
  - Customers feel convinced after a call. Hence, preference must be allotted to conversations on phone .
  - Now these conversations must be held with the leads identified from **Lead Add from.**
  - Also, the targeted customers must be **working professionals**.