LEAD SCORING CASE STUDY

Kavala Sudarshan Sampath Kumar

Problem Statement

- An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses.
- In this way, Company is getting many leads but conversion rate is poor(30 out of 100). In order to make this process more efficient, Company identifies most potential customers as ‘Hot Leads’.
- Identifying the the hot lead, and extra focus on the potential individuals may raise the conversion rate rather than calling everyone.

Business Objective:

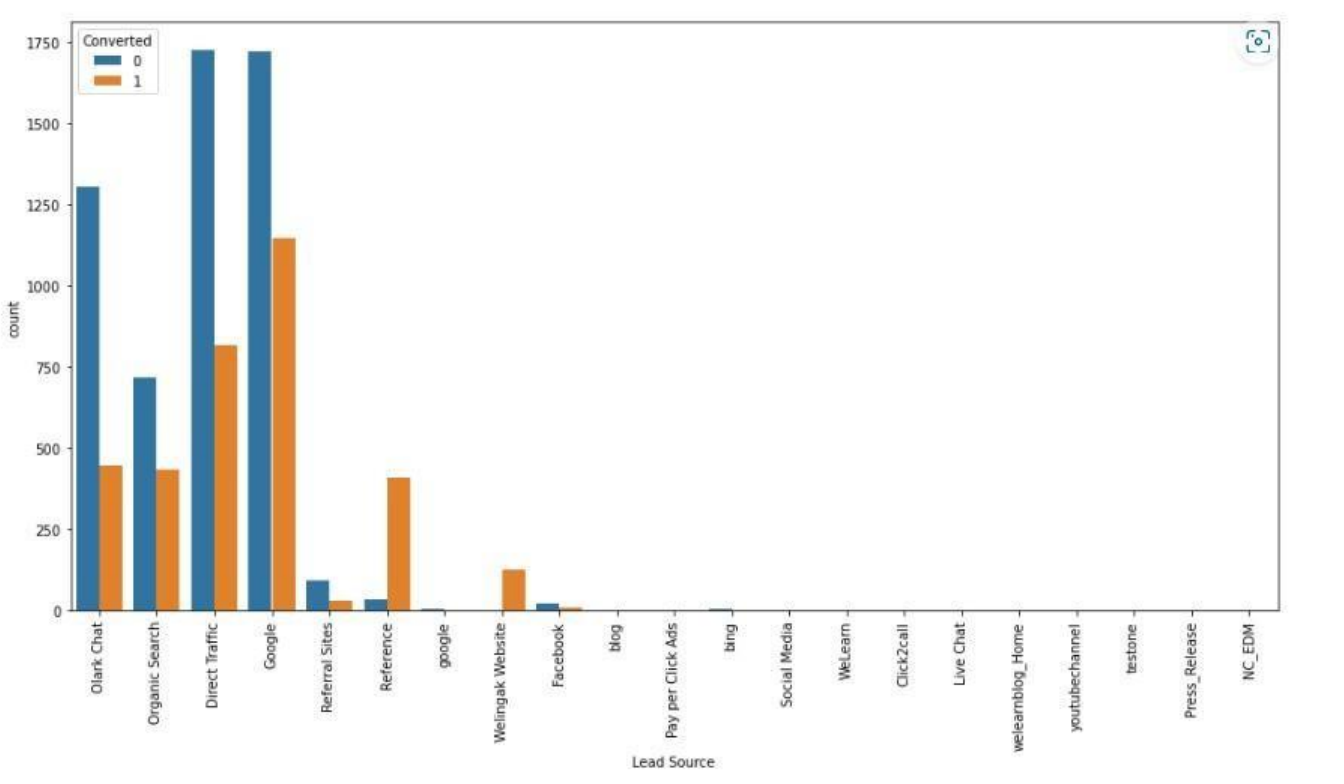
- Build a logistic regression model to assign a lead score between 0 and 100 to each of the leads which can be used by the company to target potential leads. A higher score would mean that the lead is hot, i.e. is most likely to convert whereas a lower score would mean that the lead is cold and will mostly not get converted.
- Conversion rate target is 80%
- Deployment of Model for future requirements.

Stepwise Strategy

- Importing and inspecting Dataset
- Read, Clean and preparing the Data for further Analysis
- Exploratory Data Analysis for getting most helpful features to raise the conversion rate
- Feature Scaling
- Preparing the dataset for Model Building
- Building the Logistic Regression Model
- Assigning lead scores for all leads
- Testing of built model on train data set
- Evaluating the model on the basis of different measures and metrics
- Test the model on test data set
- Measuring the accuracy of final model and other metrics for evaluation

Exploratory Data Analysis

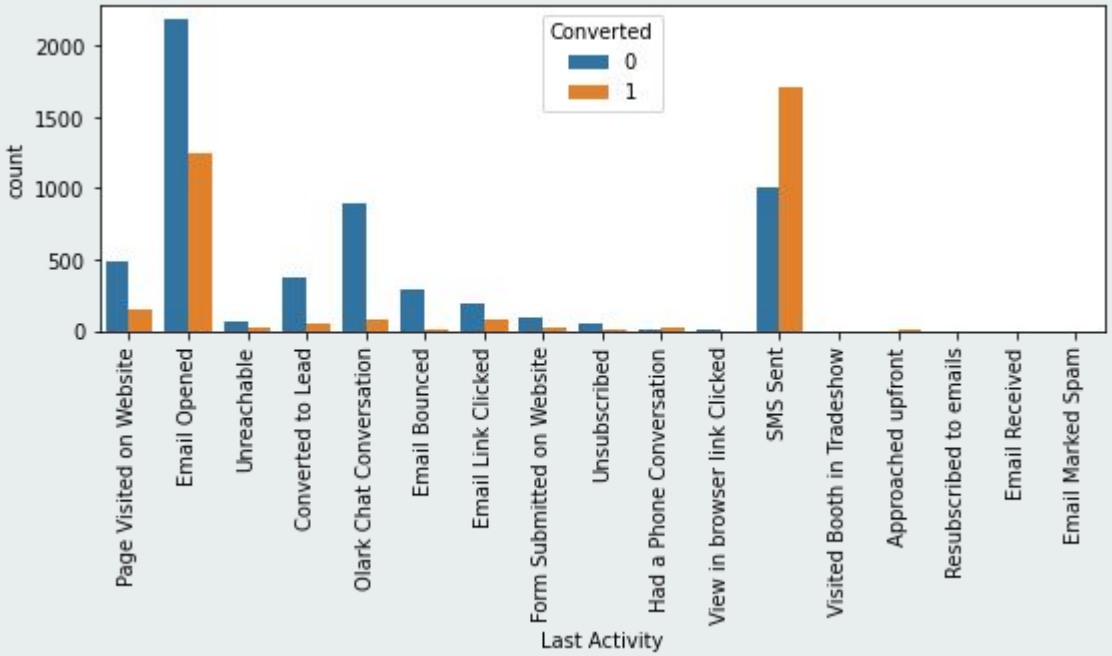
Lead Source VS Converted



As a lead source **Google Search** has high conversion counts compared to other sources whilst Reference has highest **conversion rate**

Exploratory Data Analysis

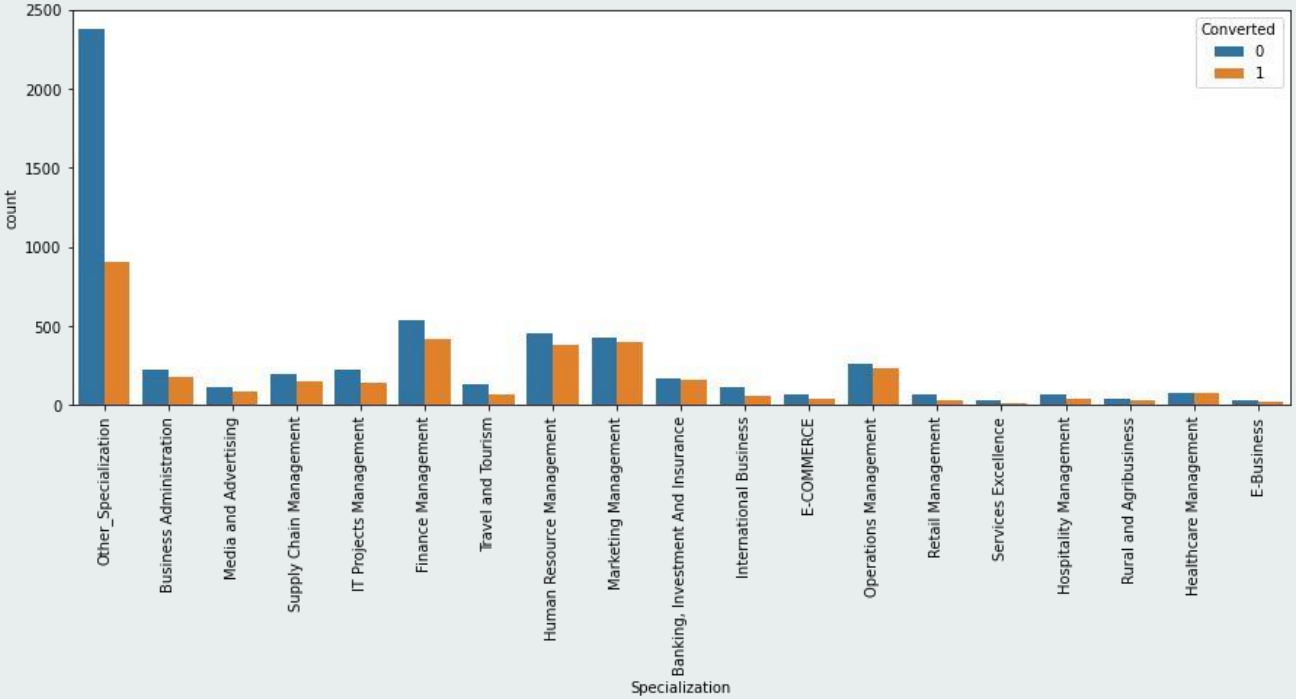
Last Activity VS Converted



SMS are proved the best promising method for getting higher converted leads followed by Email opened among all activities.

Exploratory Data Analysis

Specialization VS Converted

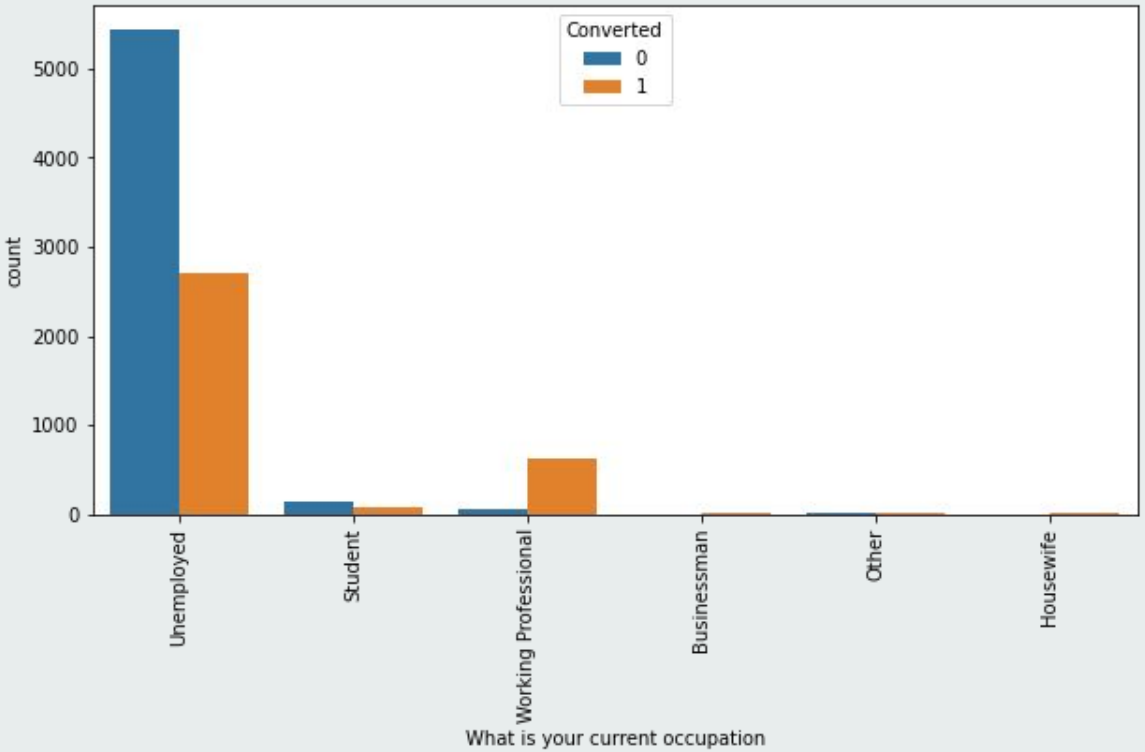


Enough Values are not present for specialization feature.

Marketing management, human resources management has high conversion rates, and students related to these specializations can be focused for the conversion.

Exploratory Data Analysis

Current Occupation VS Converted



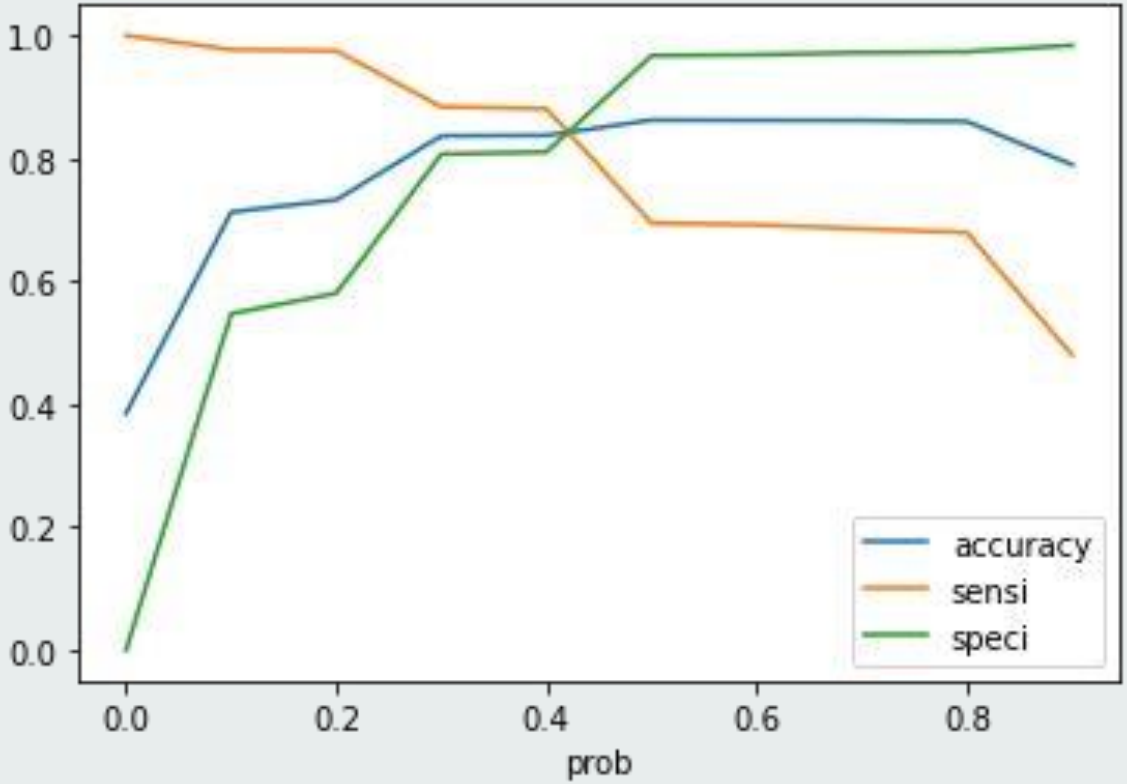
Unemployed People are the highest number of converted people but Working professional are having the highest conversion rate.

Model Building

- Splitting into train and test set
- Scaling of variables in train set
- Building the first model
- Eliminating less relevant variables using RFE
- Building the next model
- Eliminating the variables with the existing columns
- Prediction using train set
- Evaluating the accuracy and other important metrics of the model
- Prediction using test set
- Precision and recall analysis on test predictions

Model Evaluation

Accuracy, Sensitivity and Specificity



Conclusion

EDA: potential features on which people spend more time than average times are leads

- The total duration of Website browsing
- Total number of website visits
- Most of the lead source are:
  - **Direct traffic**
  - **Organic search**
  - **Google search**
- Last activity of the clients are **Olark Chat conversation | SMS**
- Building the next model
- Eliminating the variables with the existing columns

Logistic regression Model

- Accuracy, Sensitivity, specificity measures and precision help getting threshold.
- The outcome of model are ‘Hot Leads’ and the leads which have less chances of conversion.

Submitted by:

Kavala Sudarshan Sampath Kumar