

# Lead scoring case study

Anush Pakki

Abbireddy Praveen Reddy

Nimma Vamsi Krishna Reddy

# Introduction

- 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.
- The company markets its courses on several websites and search engines like Google. Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos.
- X Education gets a lot of leads, however 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'.

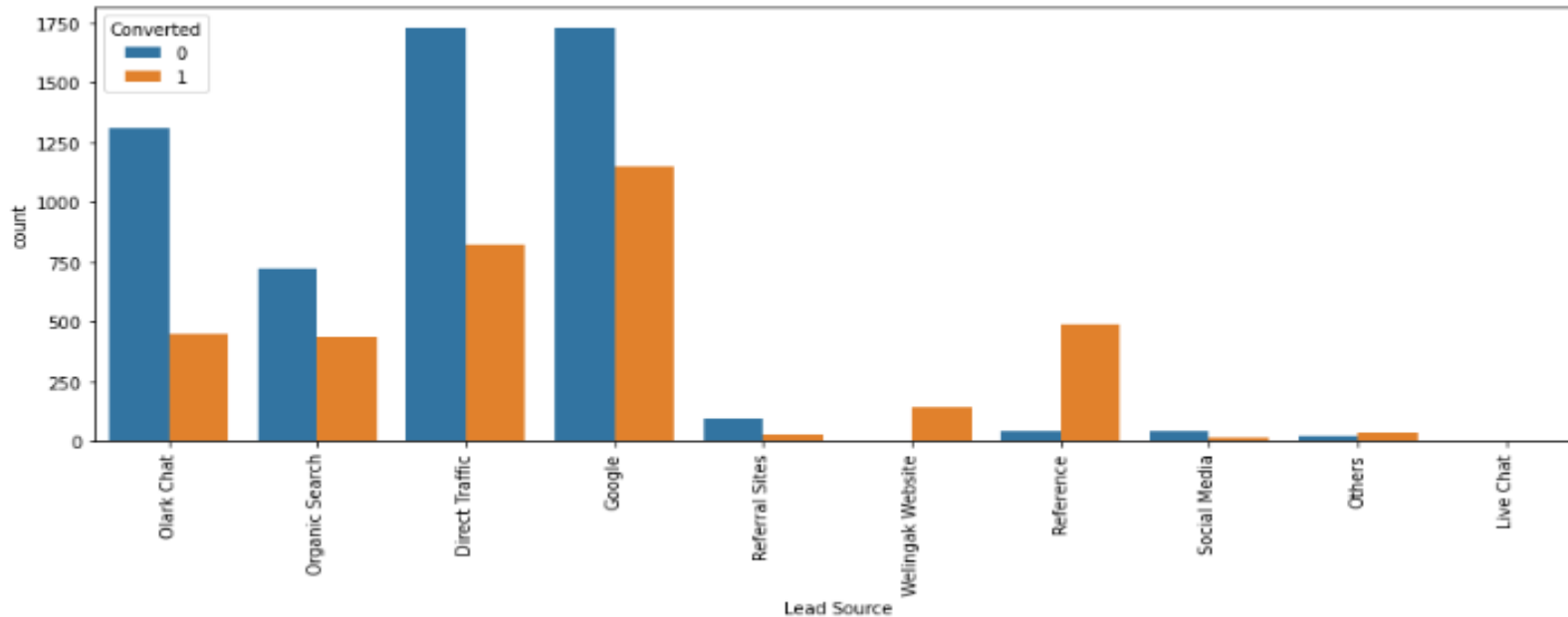
# Objective

- The objective of this case study is to identify the most promising leads, i.e. the leads that are most likely to convert into paying customers.
- To build a model that assign a lead score to each of the leads such that the customers with a higher lead score have a higher conversion chance and the customers with a lower lead score have a lower conversion chance.

# Methodology

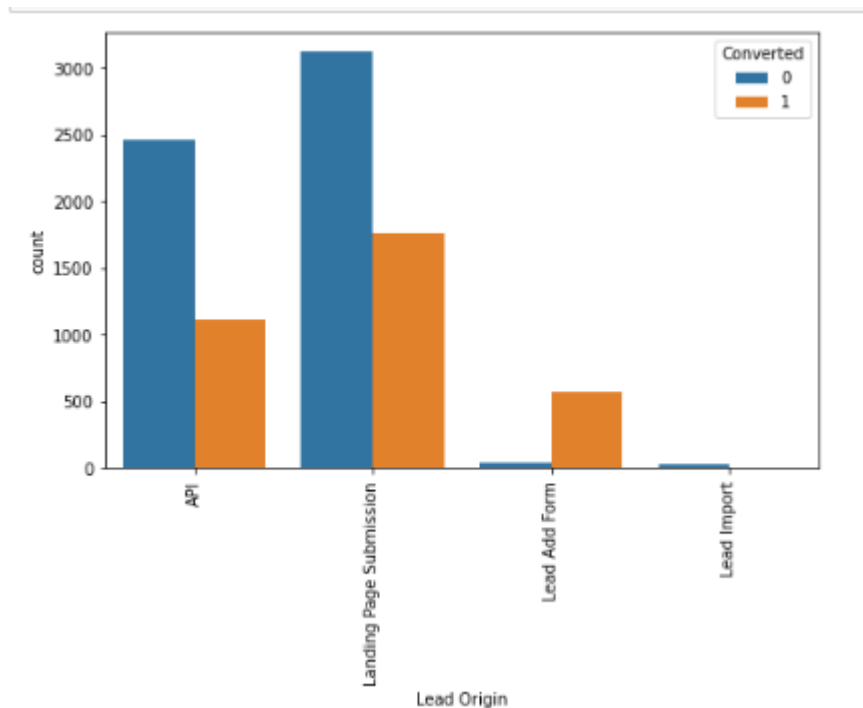
- Data understanding, preparation and EDA
  1. Ensuring Null values are removed as per threshold or imputed accordingly
  2. Handling duplicate values
  3. Handling outliers
  4. Univariate Analysis
  5. Bivariate Analysis
  6. Feature Scaling and adding Dummy variables and encoding of the data
  7. Model Building: Classification using Logistic Regression
  8. Validation
  9. Conclusion and Recommendations

# EDA



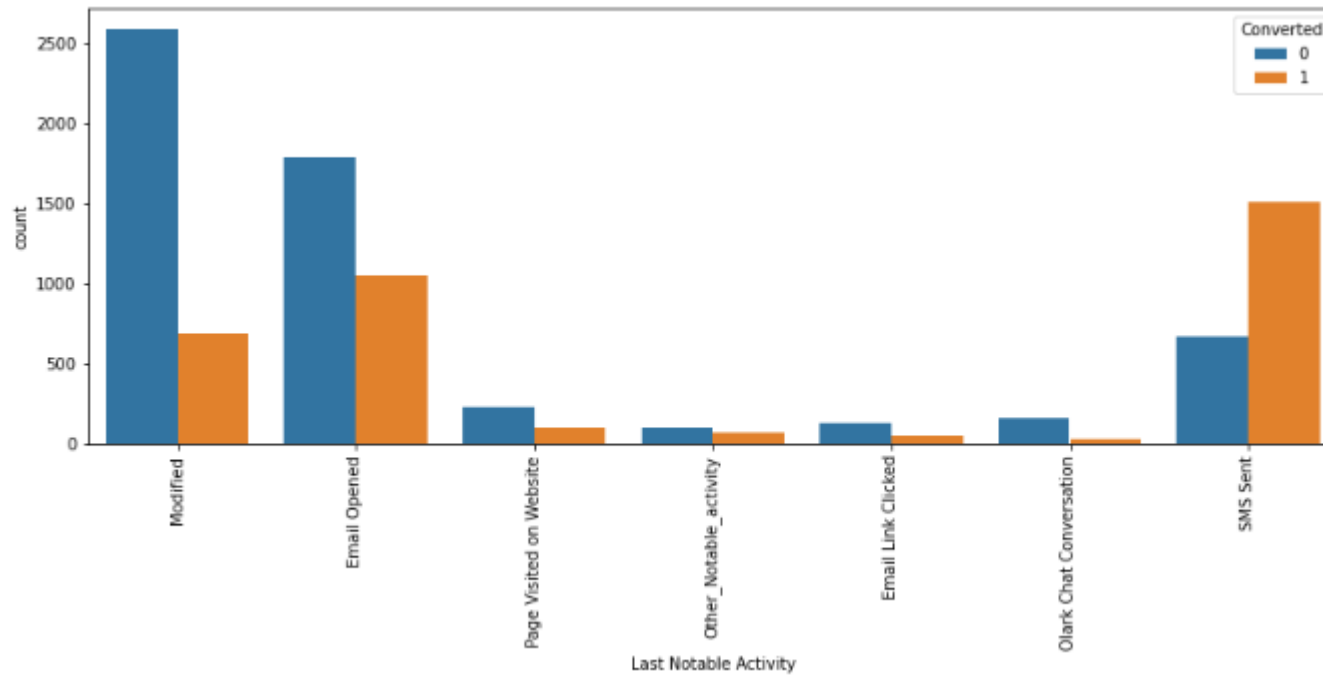
- Majority of leads are generated by Google and Direct traffic.
- welingak website is high in terms of reference leads.
- There should be more focus on improving lead conversion of olark chat, organic search, direct traffic, and google leads

# EDA



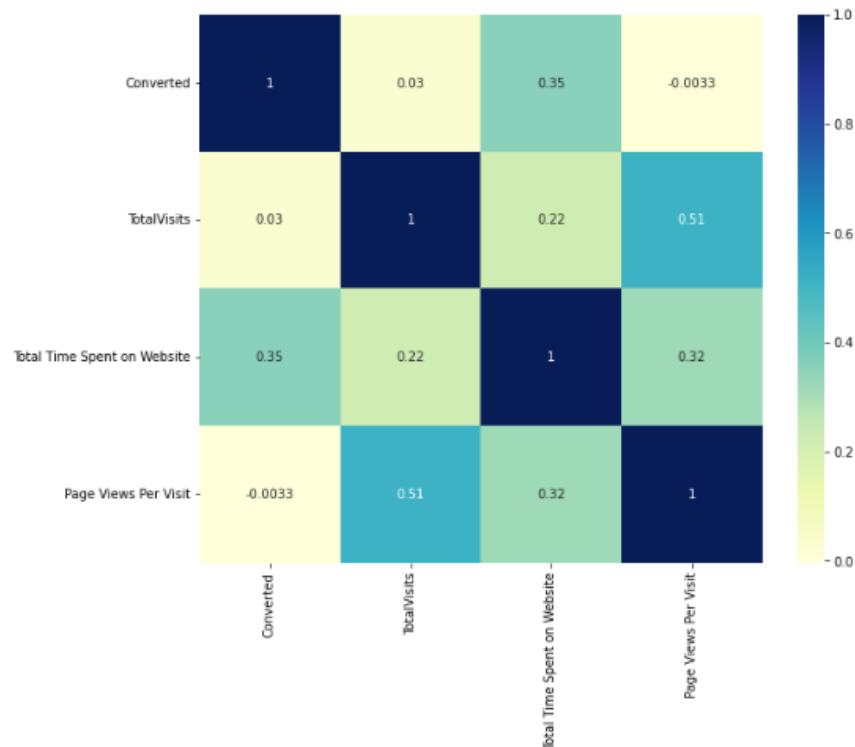
- API and Landing Page Submission lead to higher number of leads and conversion.
- Lead Add Form shows high conversion rate however count of leads are not very high.
- Lead Import and Quick Add Form may not result in more leads.
- To improve overall lead conversion rate, there should be more focus on improving lead conversion of API and Landing Page Submission origin and generate more leads from Lead Add Form.

# EDA



- Modified, Email Opened and SMS Sent are contributing to majority categories in last notable activity

# Correlation - Numerical variables



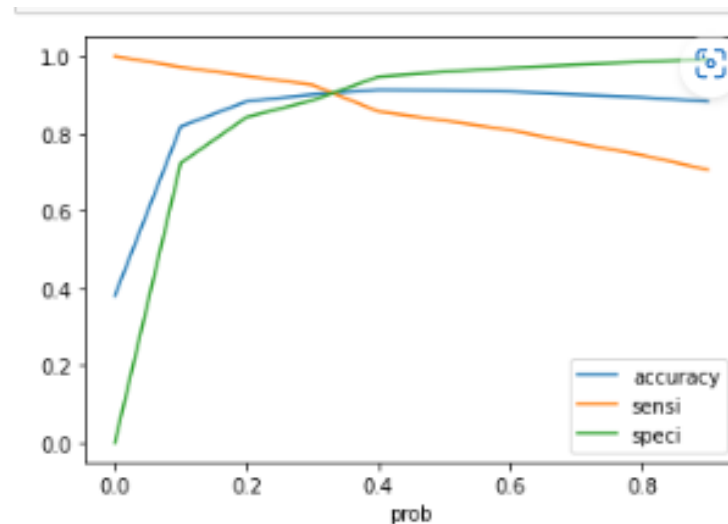
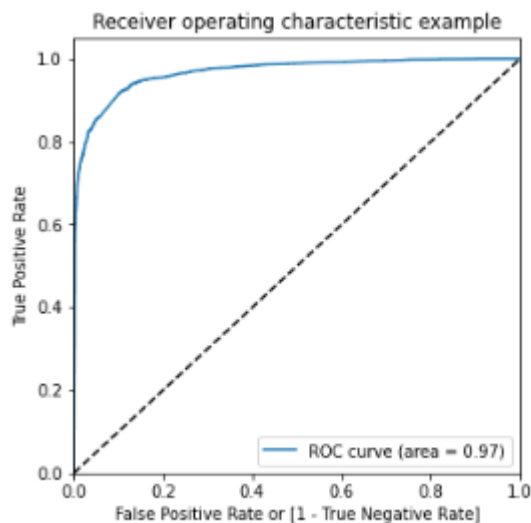
Total time spent on website have direct correlation with converted variable, indicating the users spending more time is likely to get converted



# Model Building

- Splitting the data into train and test
- Scaling the data
- Model building using stats model and RFE
- Running RFE with 15 variables as output
- Model building by removing variables whose p values is greater than 0.05
- Predictions on test data set

# ROC Curve and Optimal Probability



- The ROC Curve should be a value close to 1. We got 0.97 indicating a good predictive model.
- 0.38 is the optimum point to take it as a cutoff probability.

# Conclusion and Recommendations

- The key variables that are affecting the conversion are as follows:
  - Total time spent on website
  - Working professionals
  - Lead source was Google, Direct traffic, Organic search and Welingak website
  - Last activity was SMS and Olark Chat conversion

*This indicates that the company could focus on these attributes or variables to get more leads and conversions*

*Make the website more engaging so that the users will get attracted*

*Use SEO techniques to improve the company's promotions on Google*

*Target Working Professionals so that the company can avoid unnecessary cost and time for other categories*