

**upGrad**



International  
Institute of Information  
Technology Bangalore

# Lead Scoring Case Study

**upGrad & IIITB | Data Science Program - June 2022**

Group Members:

- Prashant Kumar Amit
- Madhulika Sahni
- Swathikiran Sattenapalli



# Business Objective

## ◆ **Goals of the Case Study**

◆ There are quite a few goals for this case study:

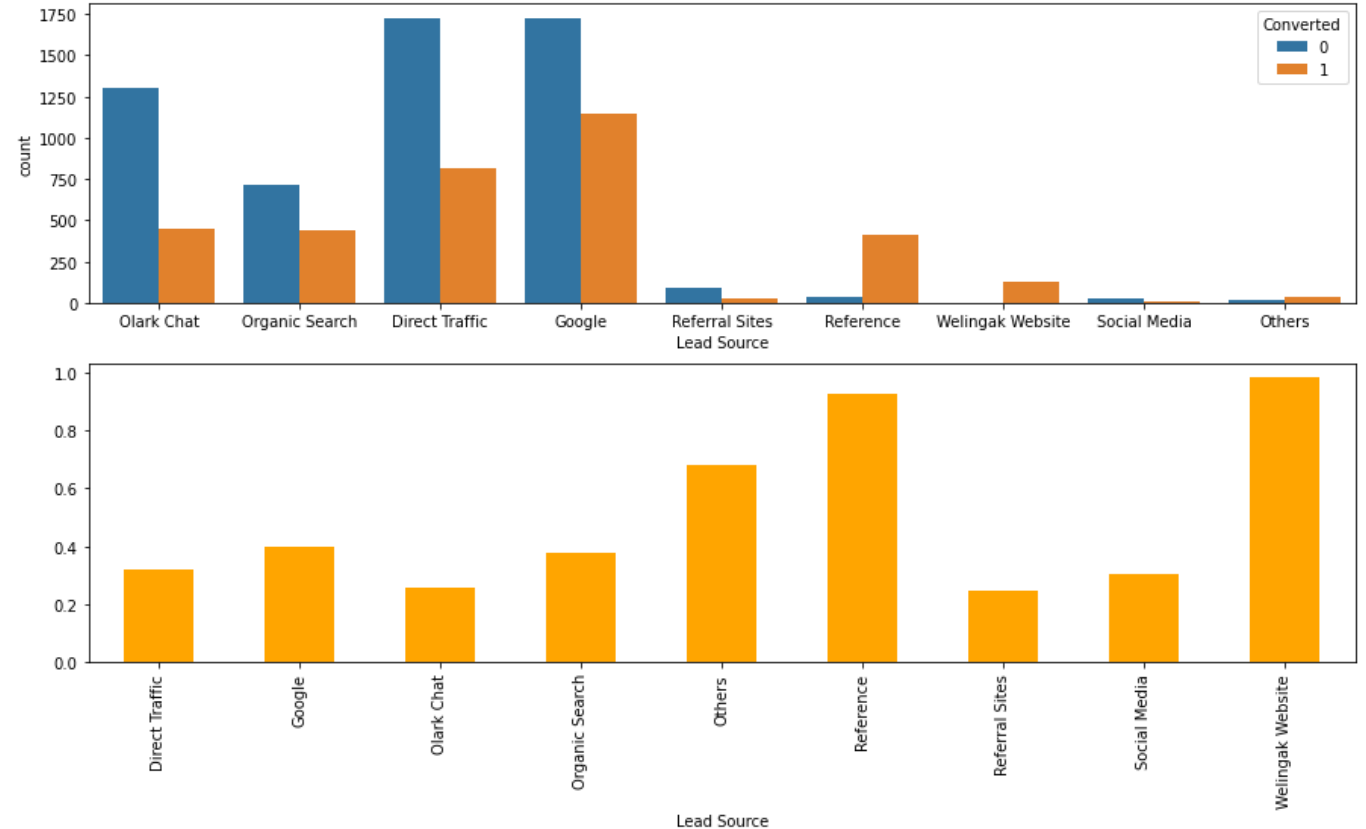
1. Build a logistic regression model to assign a lead score between 0 and 100 to each of the leads which can be used by the company to target potential leads. A higher score would mean that the lead is hot, i.e. is most likely to convert whereas a lower score would mean that the lead is cold and will mostly not get converted.
2. There are some more problems presented by the company which your model should be able to adjust to if the company's requirement changes in the future so you will need to handle these as well. These problems are provided in a separate doc file. Please fill it based on the logistic regression model you got in the first step. Also, make sure you include this in your final PPT where you'll make recommendations.

# Solution Methodology

- Cleaning Data
  - a. The data was partially cleaned at first step by removing some of the null values and then some of the null values were generalised by replacing it with “Not Provided”
- EDA
  - a. Using categorical analysis very few categorical values were actually related to lead conversion and was relevant but rest were irrelevant. Numerical analysis had few cases with outliers but overall seems good.
- Dummy Variables
- Train test Split
- Model Building
- Model Evaluation
- Prediction
- Assign Lead score to find Hot leads

# EDA - Categorical

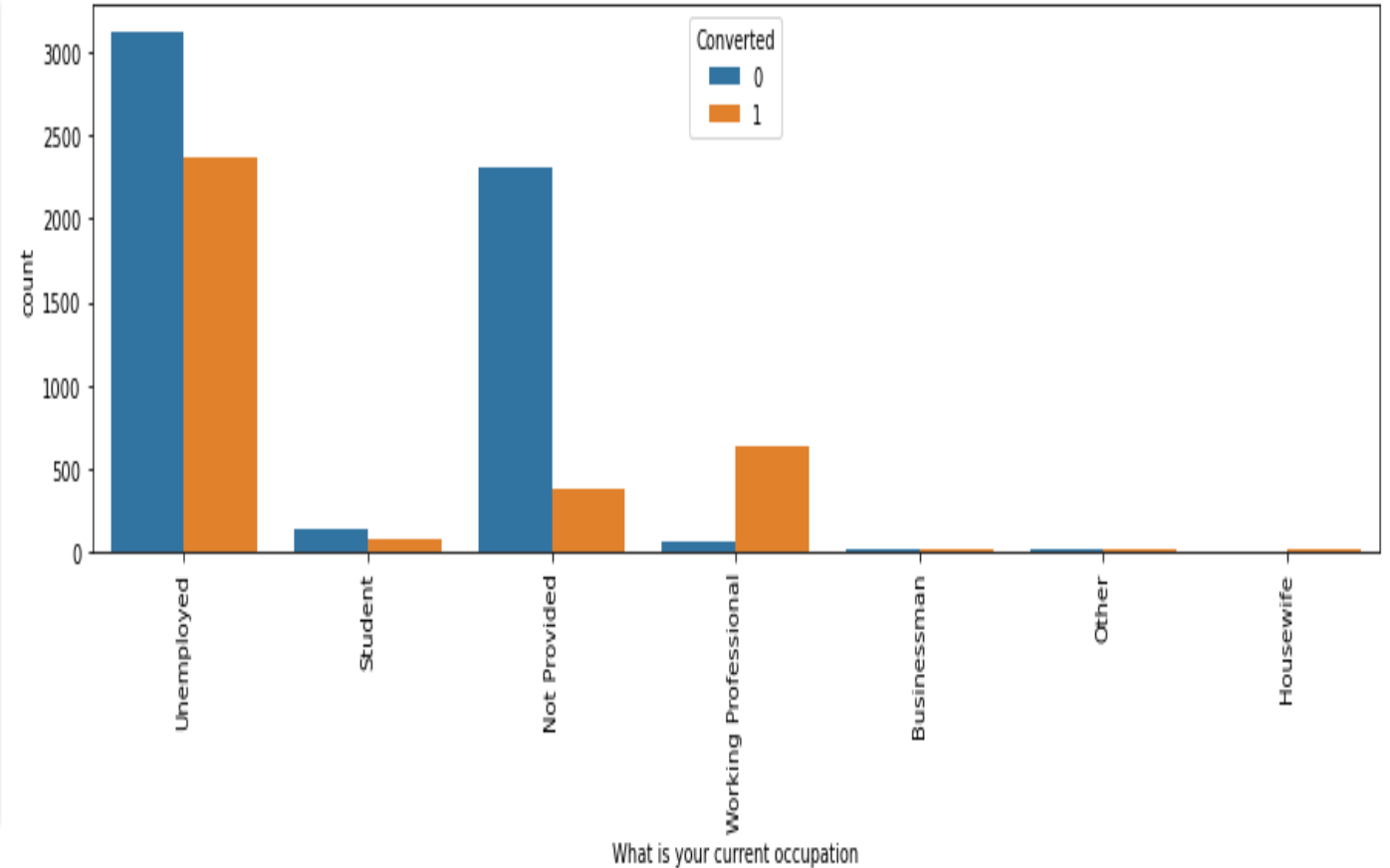
- ◇ Maximum Leads are generated by Google and Direct Traffic
- ◇ Conversion Rate for Reference Leads and Welinkgak Website Leads are Very High





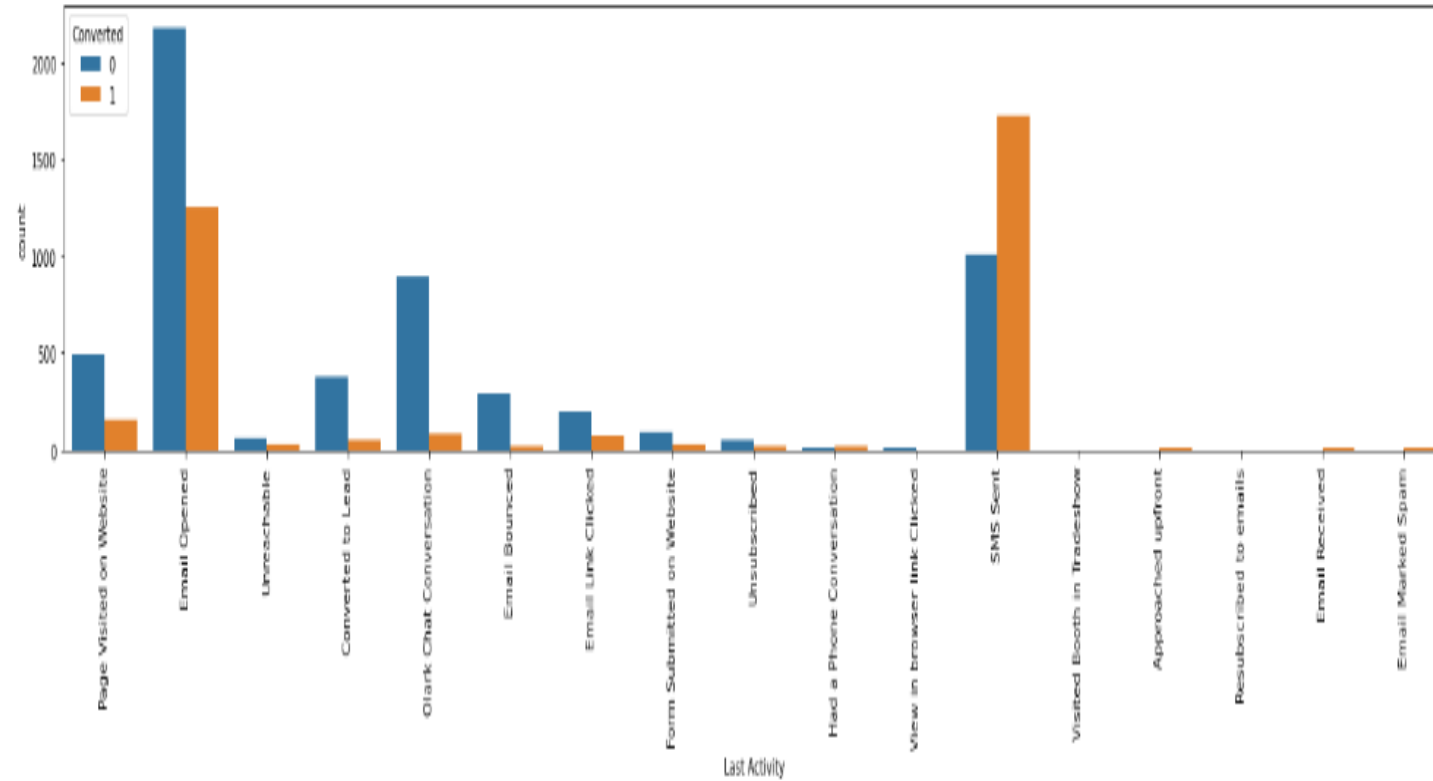
# EDA - Categorical

- ◇ Maximum Leads are generated by Unemployed and have more than 50% of conversion rate
- ◇ Conversion Rate for working professionals are Very High



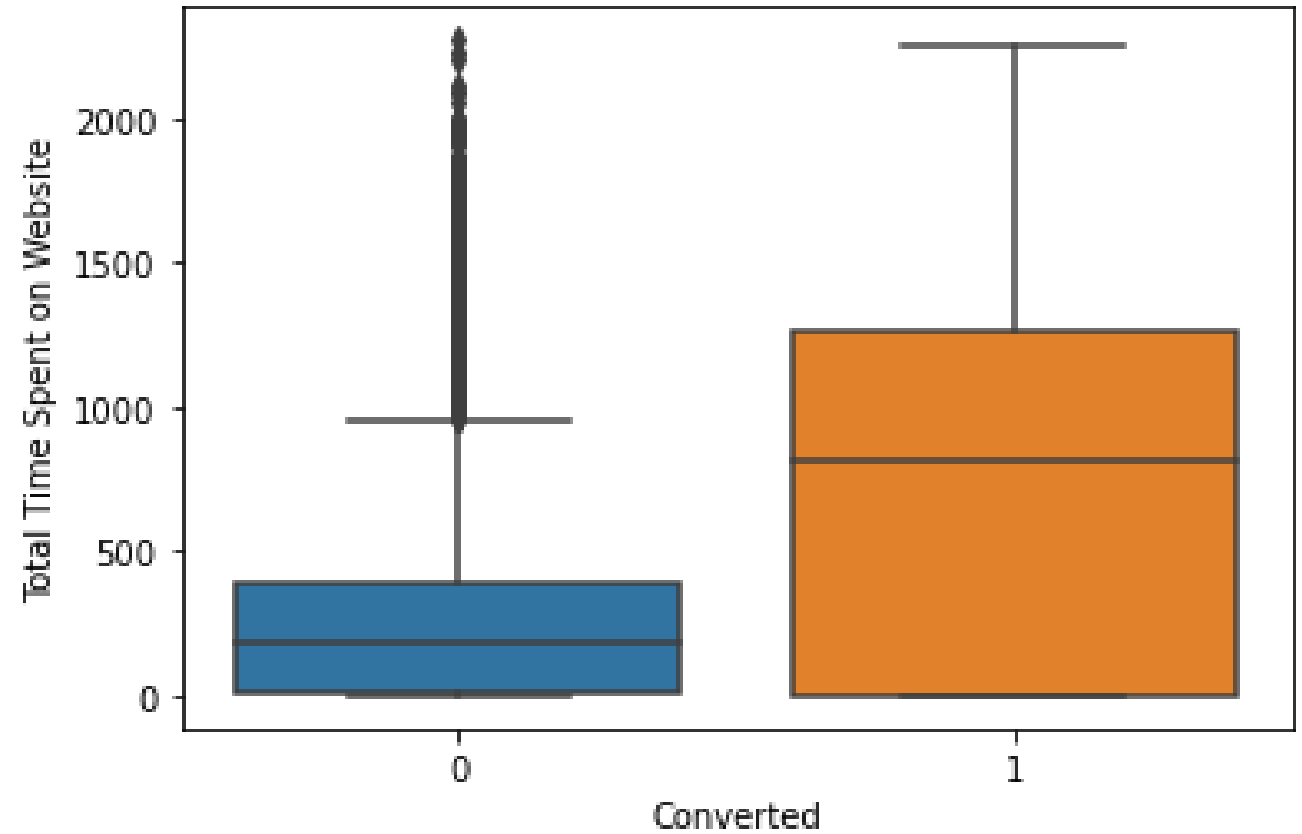
# EDA - Categorical

- Maximum Leads are generated having last activity Email opened but conversion rate is not good.
- Conversion Rate for last activity SMS sent are Very High



# EDA - Numerical

- Leads spending more time on website have more conversion Rate



# Conclusion

1. It was found that variables that to be considered as potential leads are

1. Total Time spent on the website
2. Total Number of visits
3. When lead source are coming from
  1. Google, Direct Traffic etc
4. Current occupation with unemployed and working professionals
5. Last activity is SMS
6. When the lead origin is Lead add format.

Keeping these parameters and using the prediction model, X Education can have maximum Leads converted into business.