

# XEDUCATION LEAD SCORING DATA ANALYSIS AND APPROACH

- NIKHIL TEJ O



# PROBLEM STATEMENT

- An education company named X Education sells online courses to industry professionals. Many professionals who are interested in the courses land on their website and browse for courses.
- Once these leads are acquired, employees from the sales team start making calls, writing emails, etc.
- There are a lot of leads generated in the initial stage but only a few of them come out as paying customers. In the middle stage, you need to nurture the potential leads well in order to get a higher lead conversion.
- The typical lead conversion rate is around 30%.
- The company requires you to build a model wherein you need to assign a lead score to each of the leads such that the customers with a higher lead score have a higher conversion chance. The CEO has given a ballpark of the target lead conversion rate to be around 80%.



# 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.
- 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.



# APPROACH

- Study the dataset provided by company.
- Perform Data cleaning which includes imputing null values, conversion of categorical column values into binary 1/0 and removal of columns that do not add value to the model.
- Columns containing more than 40% null values are dropped.
- Perform Uni-variate and Bi-variate analysis.
- Check for outliers and cap them to a certain level so that it doesn't affect the results.
- Check levels present in the categorical variables and create dummies.
- Plot a heat map to check correlation between attributes of dataset and drop highly correlated columns if exist.

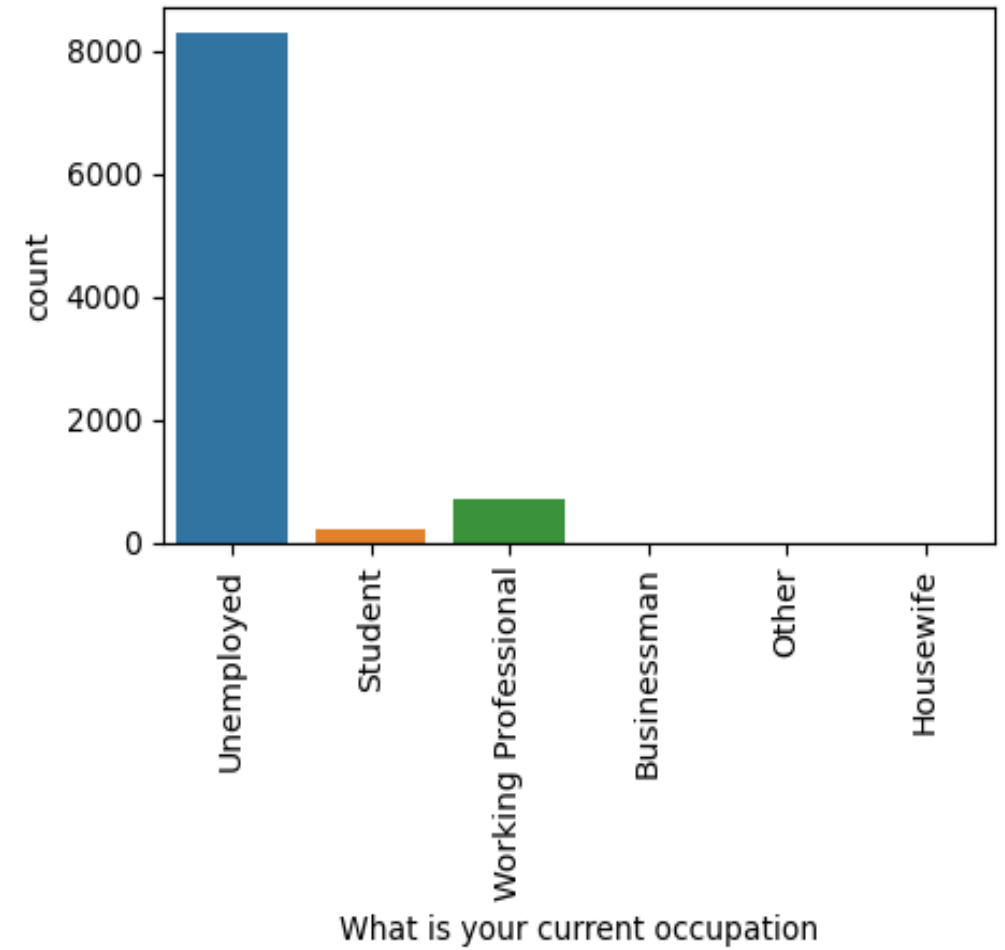
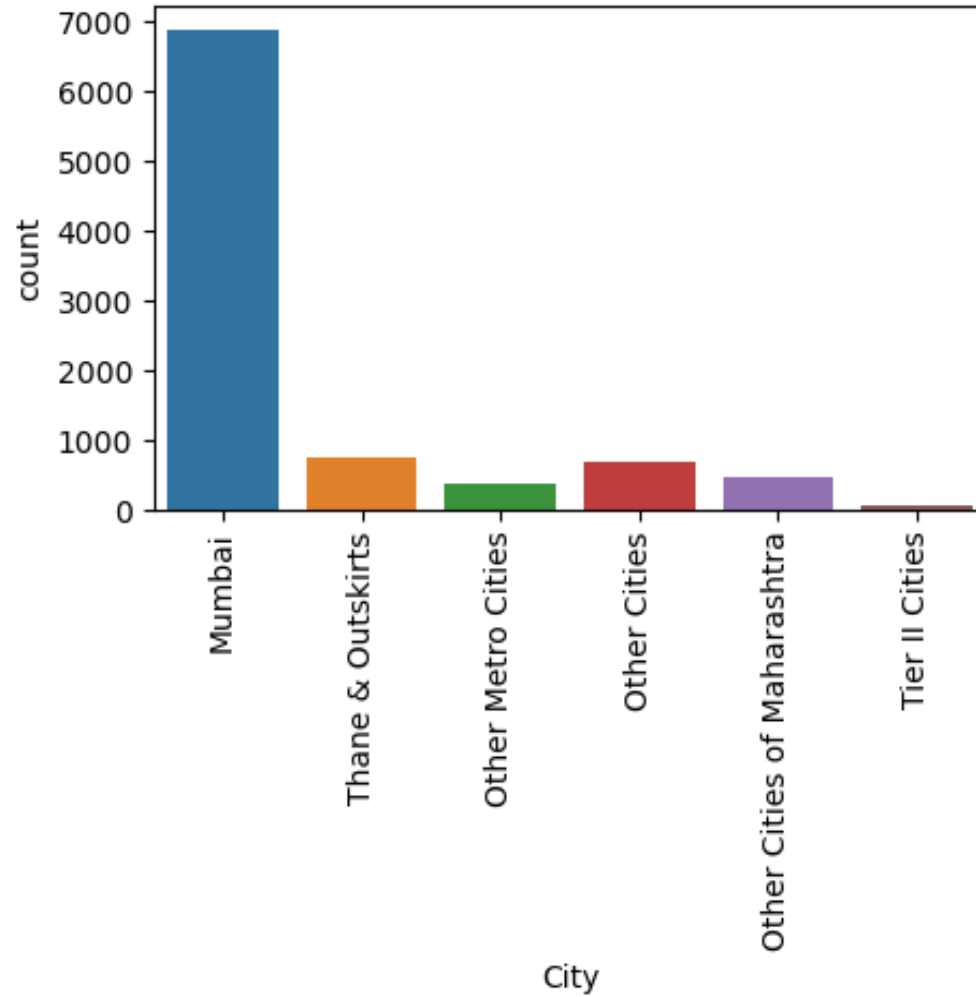


# APPROACH

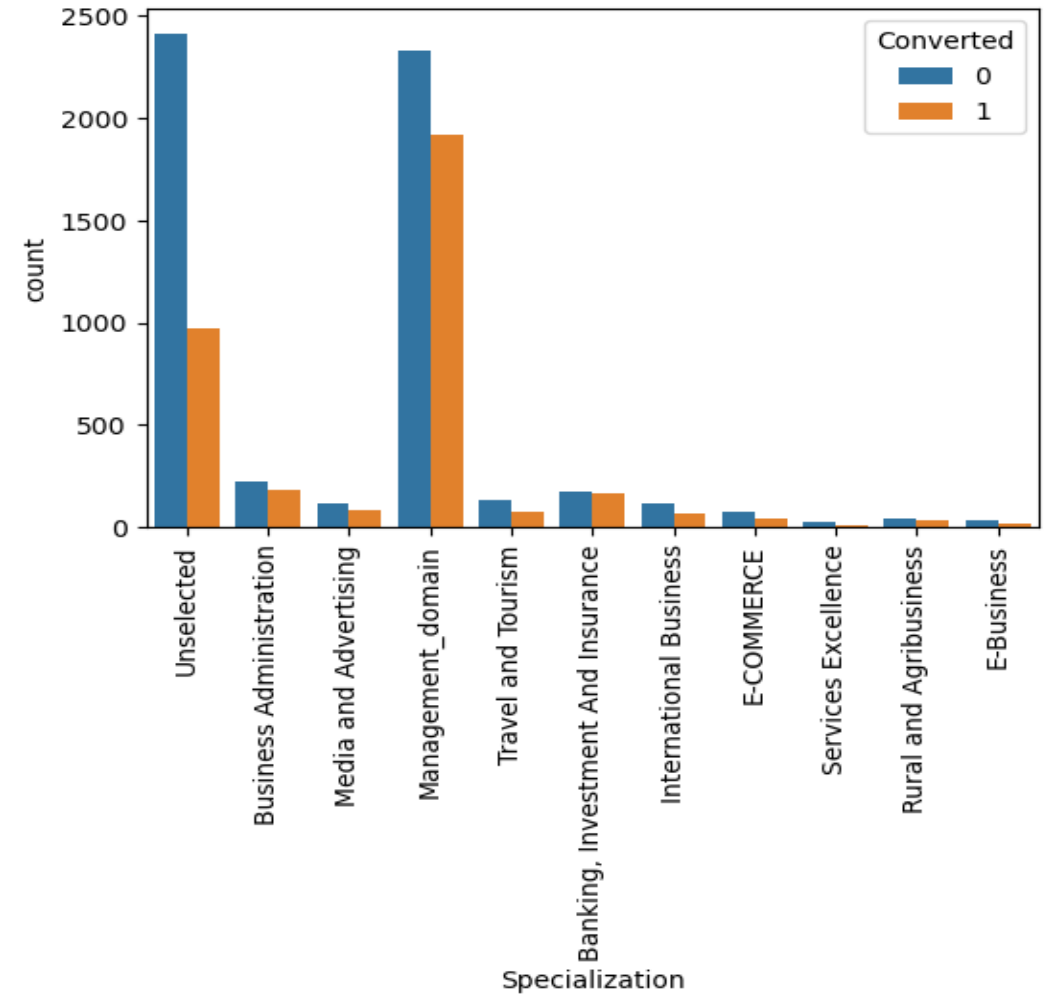
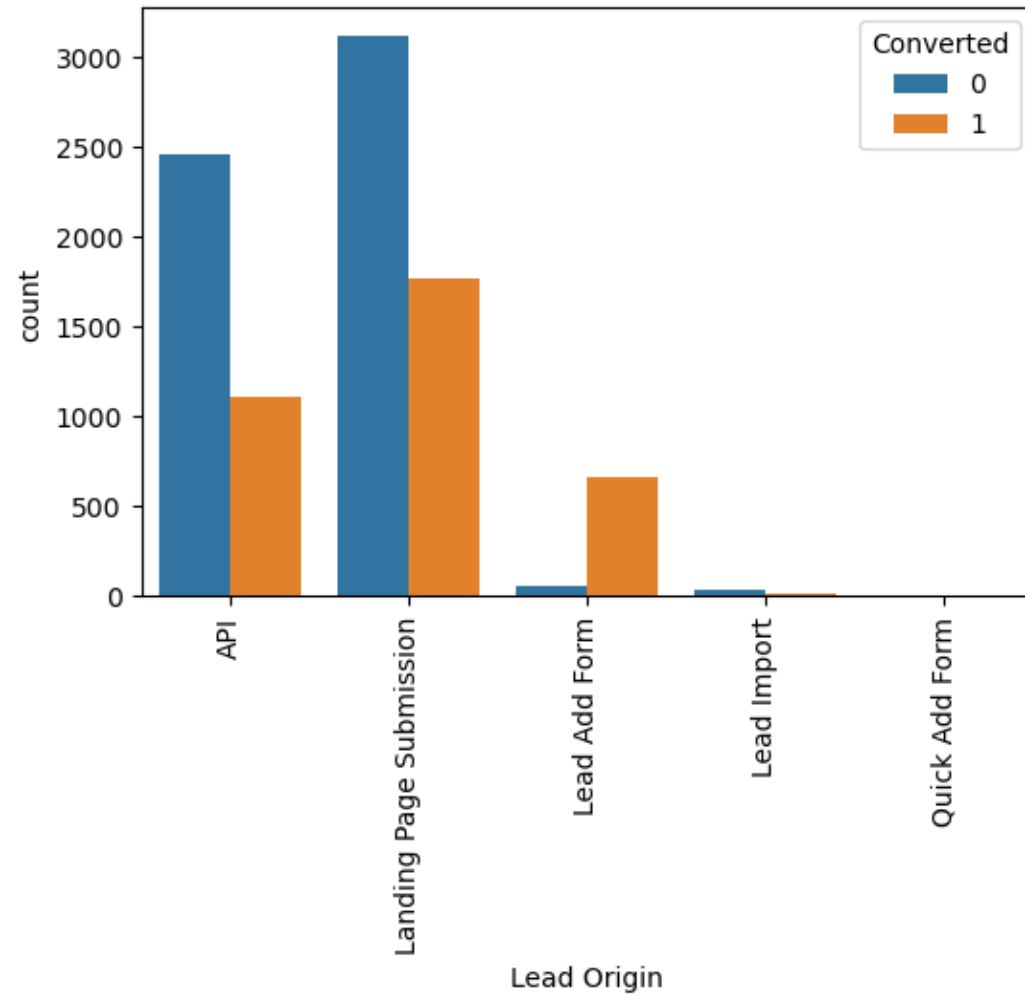
- When the data is cleaned completely and do not have irregularities, we split the data into train and test data in the ratio of 70-30 respectively.
- Scale the numerical variables using MinMaxScaler or StandardScaler depending on the data. For this model building, we have used MinMaxScaler.
- We build the Linear Regression model using Statsmodels. Then do the automated feature selection using RFE.
- Find Threshold value by plotting the metrics i.e. Accuracy, Sensitivity, Specificity and range of threshold values.
- Assess the model, check for metrics.
- when you are satisfied with the results, Feed test data to the model and assess it. It should lie according to the specified conditions.



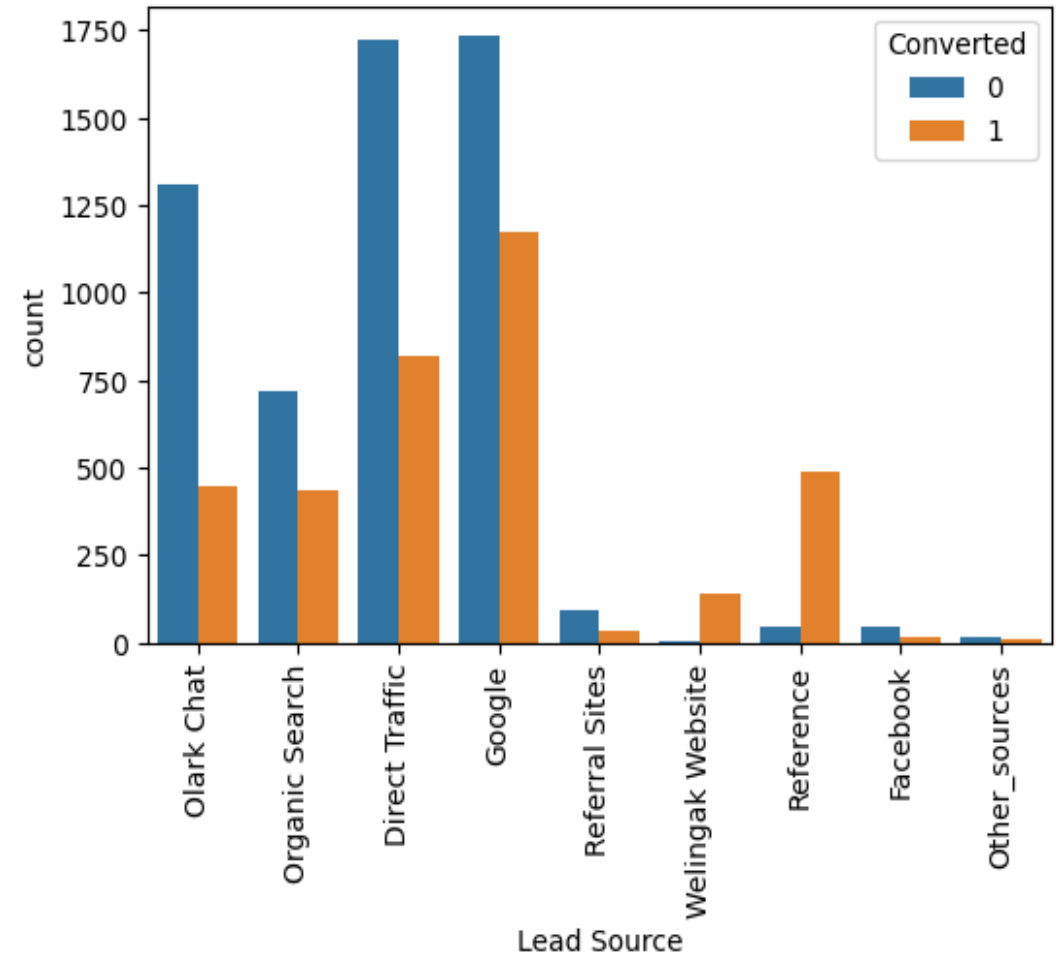
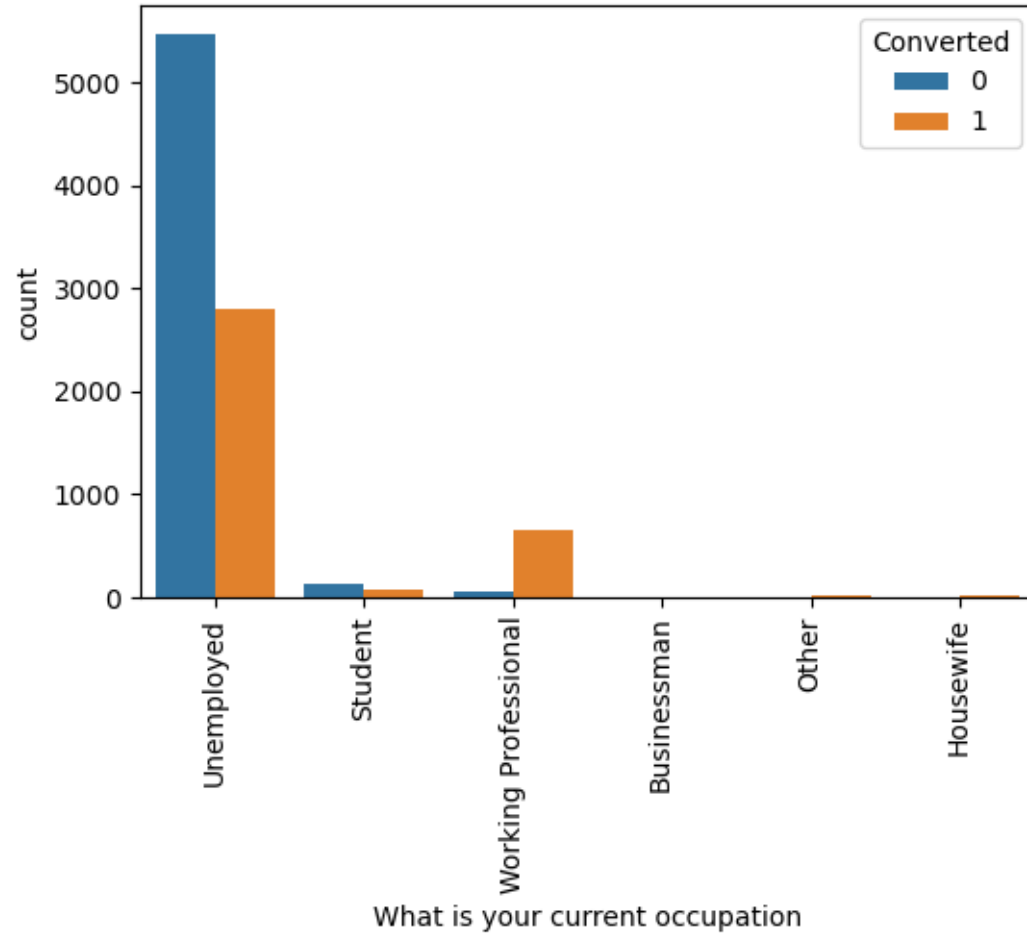
# DATA ANALYSIS



# DATA ANALYSIS

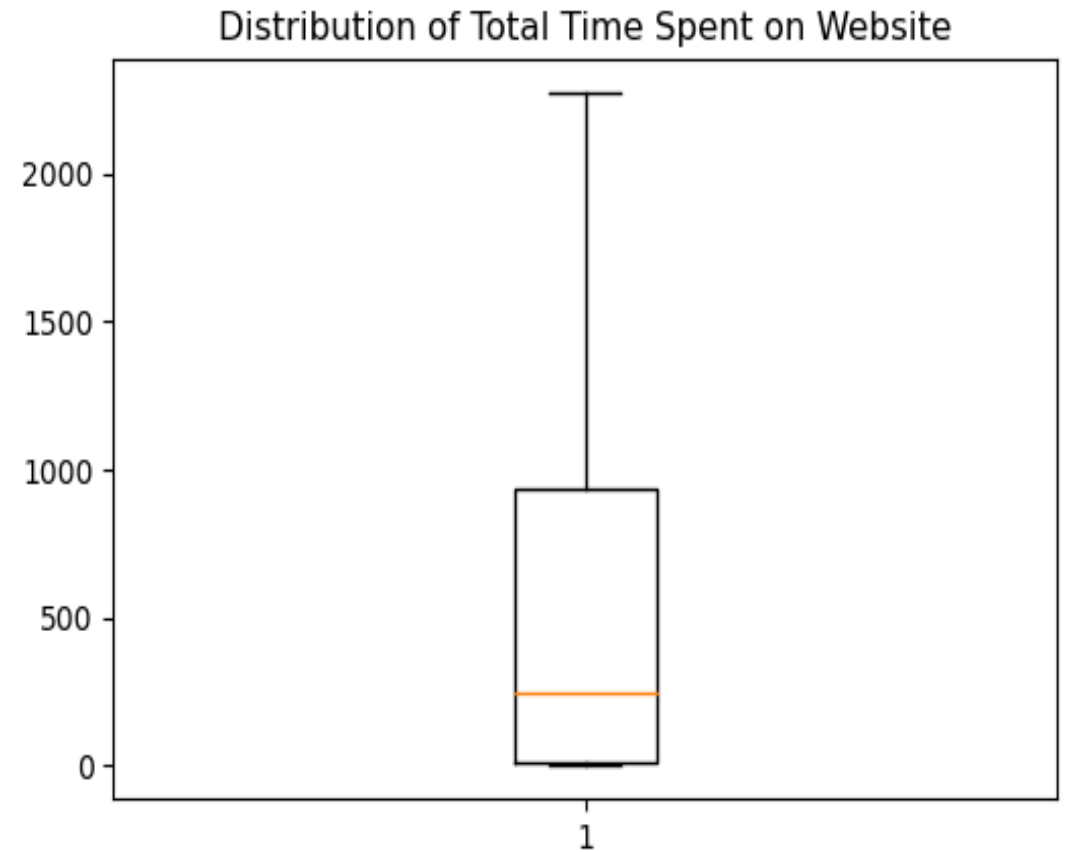
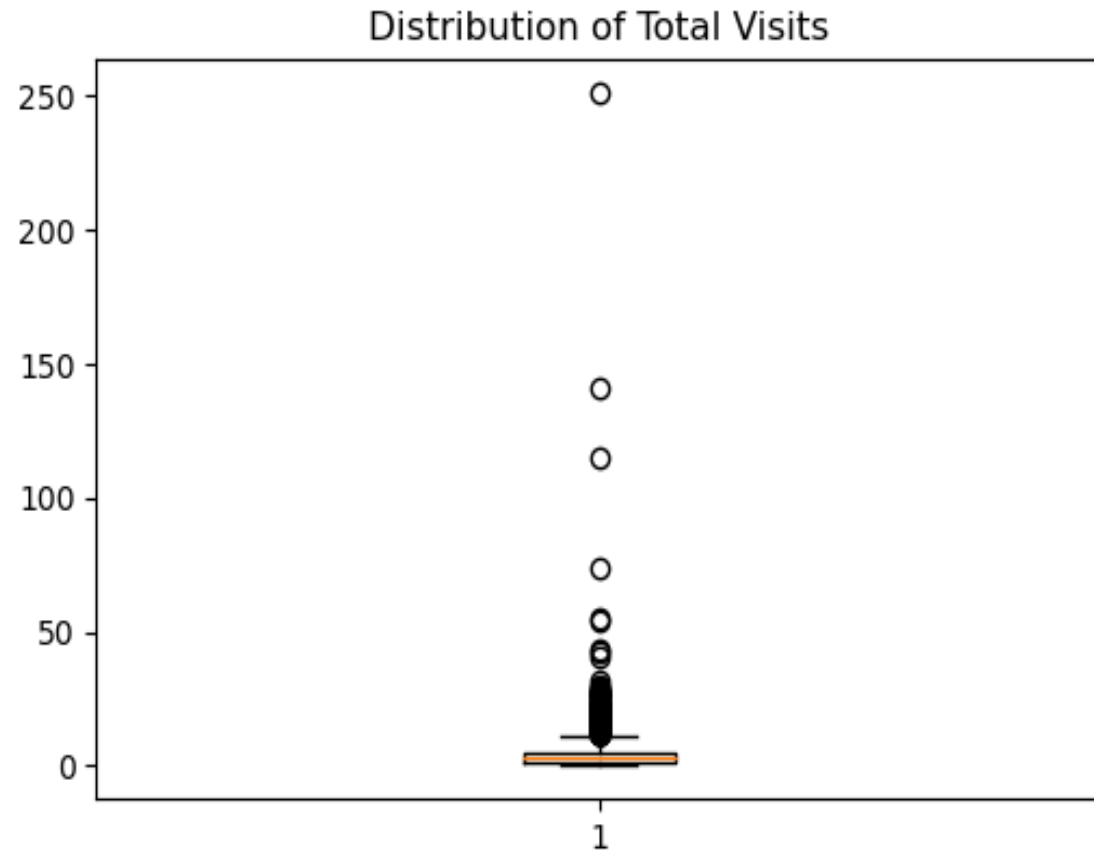


# DATA ANALYSIS





# OUTLIERS CHECK



# RESULTS

- Model has achieved an accuracy of 87% through training data.

- **Train Data Metrics**

Train Accuracy score: 0.869

Train Sensitivity: 0.862

Train Specificity: 0.872

- When we fed the test data to model, It has achieved an accuracy of 88%.

- **Test Data Metrics**

Test Accuracy score: 0.88

Test Sensitivity: 0.862

Test Specificity: 0.892



# RESULTS

- Few attributes that play an important role in determining the behaviour of the customer.

TotalVisits

Total Time Spent on Website

Tags\_Will revert after reading the email

Lead\_origin\_Lead Add Form

Lead\_source\_Welingak Website

Current\_occupation Working Professional

- Assessing the customer based on Total website visits, Total time spent on website, Tags\_Graduation in progress etc., and nurturing the lead before he moves away will help us in Lead conversion.
- Customers who are assigned Tags\_Will revert after reading the email, the candidates from source Welingak website and the ones who are working professionals must be paid attention as these leads are most likely to convert as customers.

