

LEAD SCORE CASE STUDY

Group Members

- MEGHA SUGUNAN
- SHRISH SURYA
- SUPRAJA ANNEPU

PROBLEM STATEMENT

- X Education is an online education company that sells online courses to industry professionals.
- X Education gets a lot of leads, its lead conversion rate is very poor. For example, they acquire 100 leads in a day, only about 30 of them are converted.
- To make this process the company to find the potential leads, 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

- To help X Education to select the most promising leads.
- To make this process the company to find the potential leads, known as 'Hot leads'.
- Deployment of the model for future use.

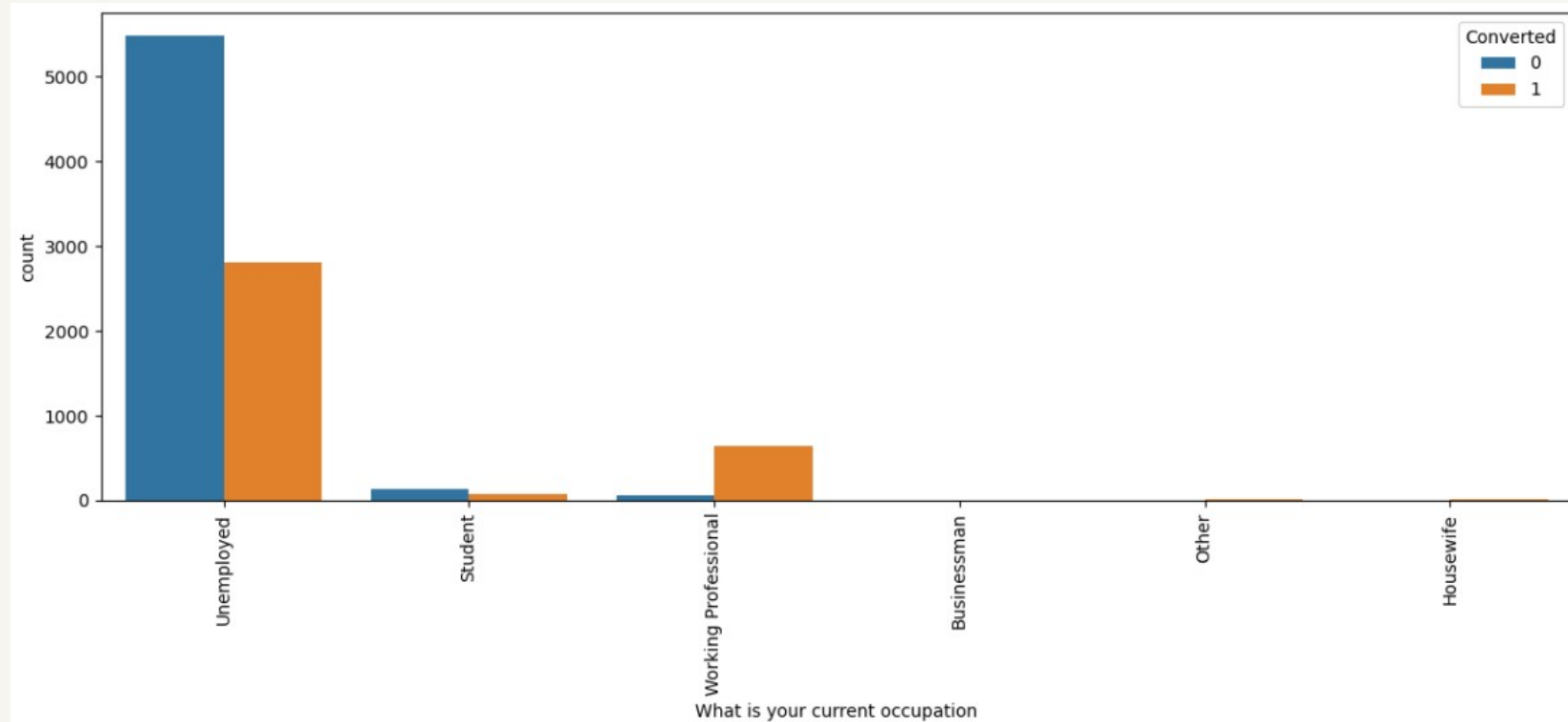
SOLUTION METHODOLOGY

- Reading and Understanding the data
- Data Cleaning
 - Check and handle duplicate data, NA values
 - Dropping variables having unique values
 - Dropping columns having more than 45% of null values
- EDA
 - Categorical variable analysis
 - Numerical variable analysis
- Dummy variable creation
- Test train split
- Model Building
- Conclusion

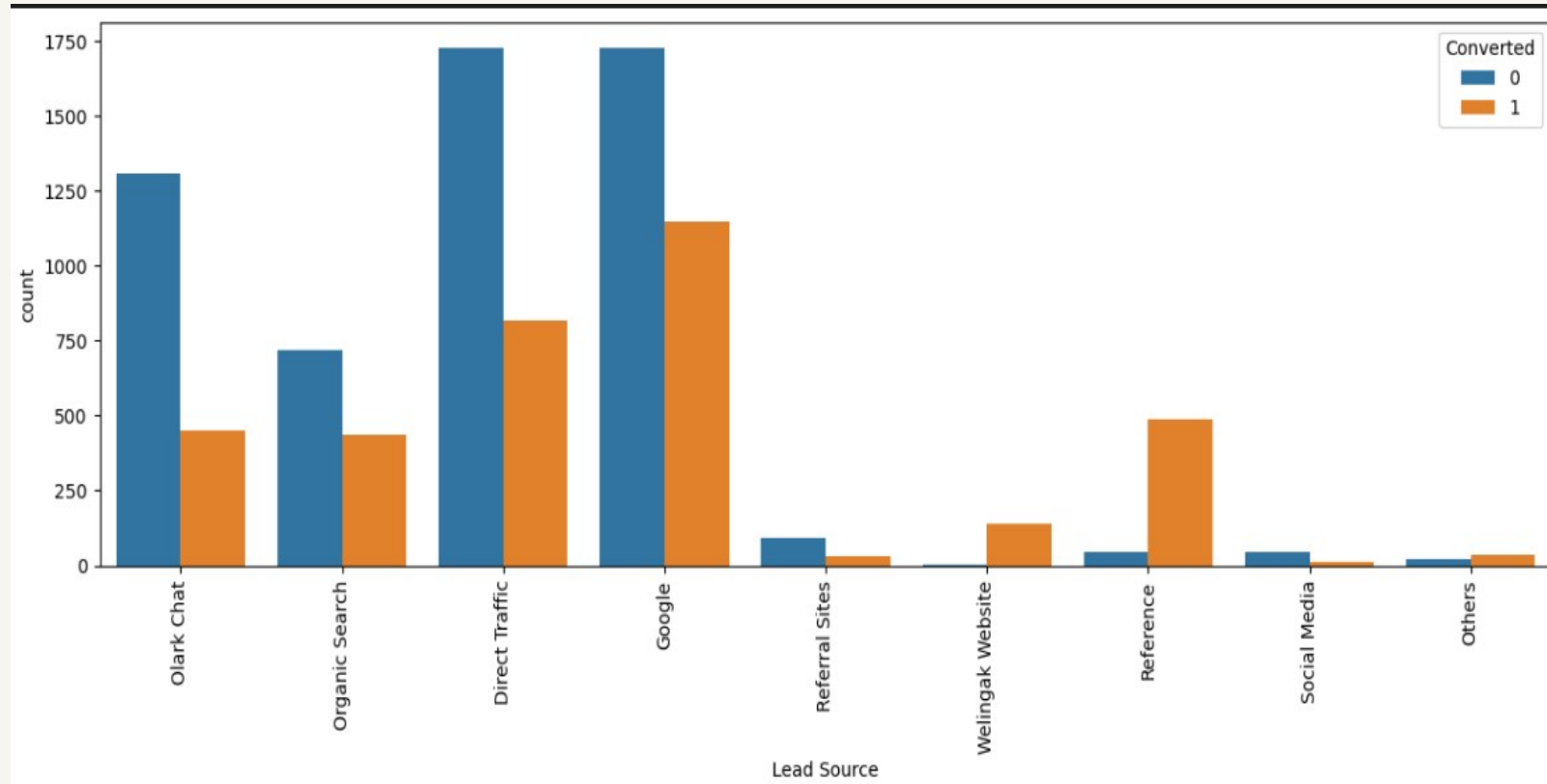
DATA MANIPULATION

- Total number of rows= 37, total number of columns= 9240
- Single values like 'Magazine', 'Receive more updates about our courses'
- Removing the 'Prospect ID' and Lead number' which does not contain much relevance in the analysis.
- Handling null values: The columns with more than 45% of null values where dropped like 'How did you hear about X Education', 'Lead profile', 'Lead quality', etc.,.

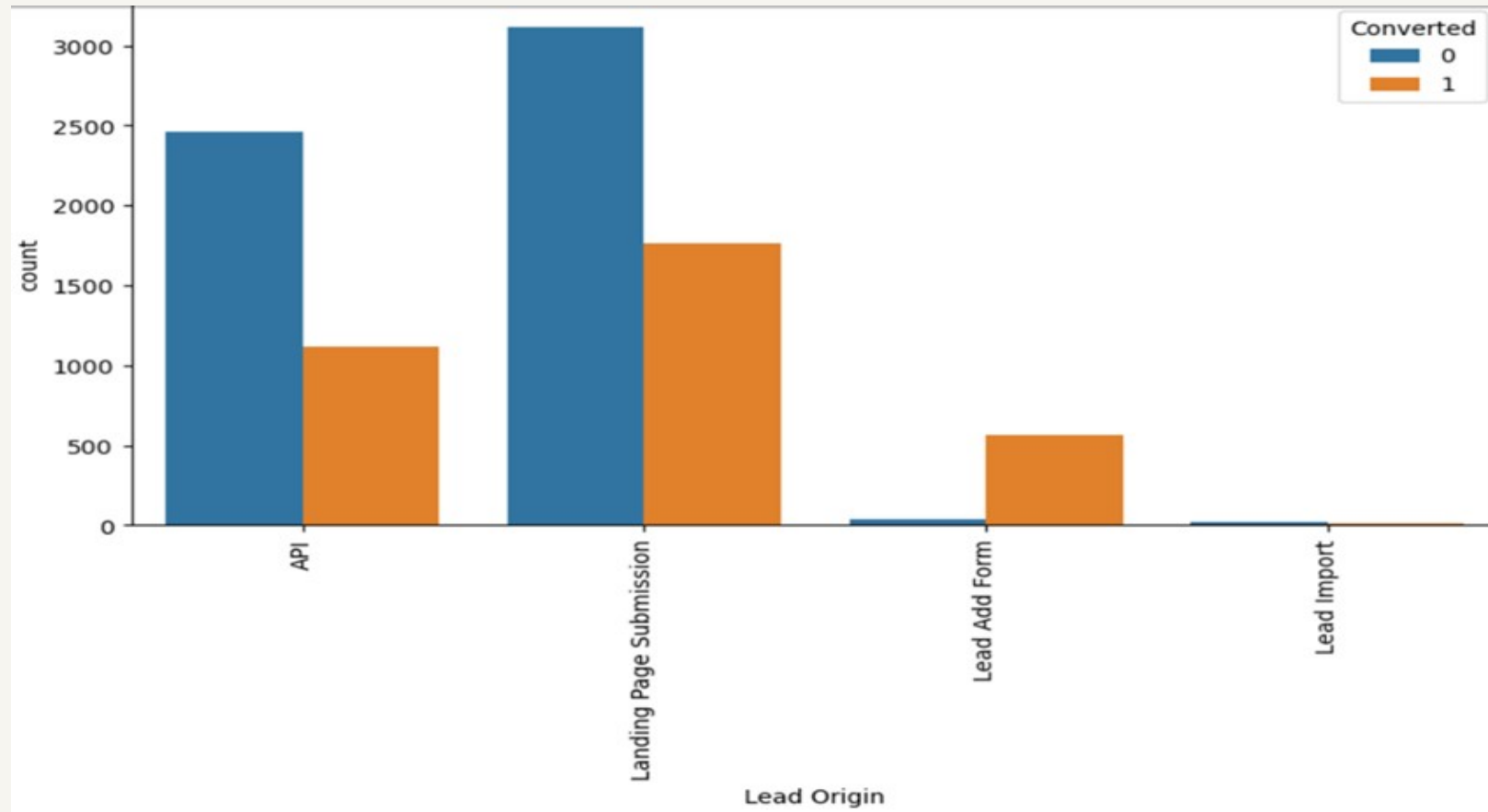
CATEGORICAL VARIABLE ANALYSIS



From the graph, we can see that most of the working professional are interested were converted; whereas most of the people who are unemployed are not interested and did not wish to enroll in the course. We can also see that most of the leads were from the unemployed section.



From the above graph, we can see that while Google and Direct traffic have the most leads, the conversion rate is low. So focus must be on improving lead conversion in both these areas. The company must also focus on improving lead conversions in both Olark Chat and Organic Search. References yield a good number of leads.

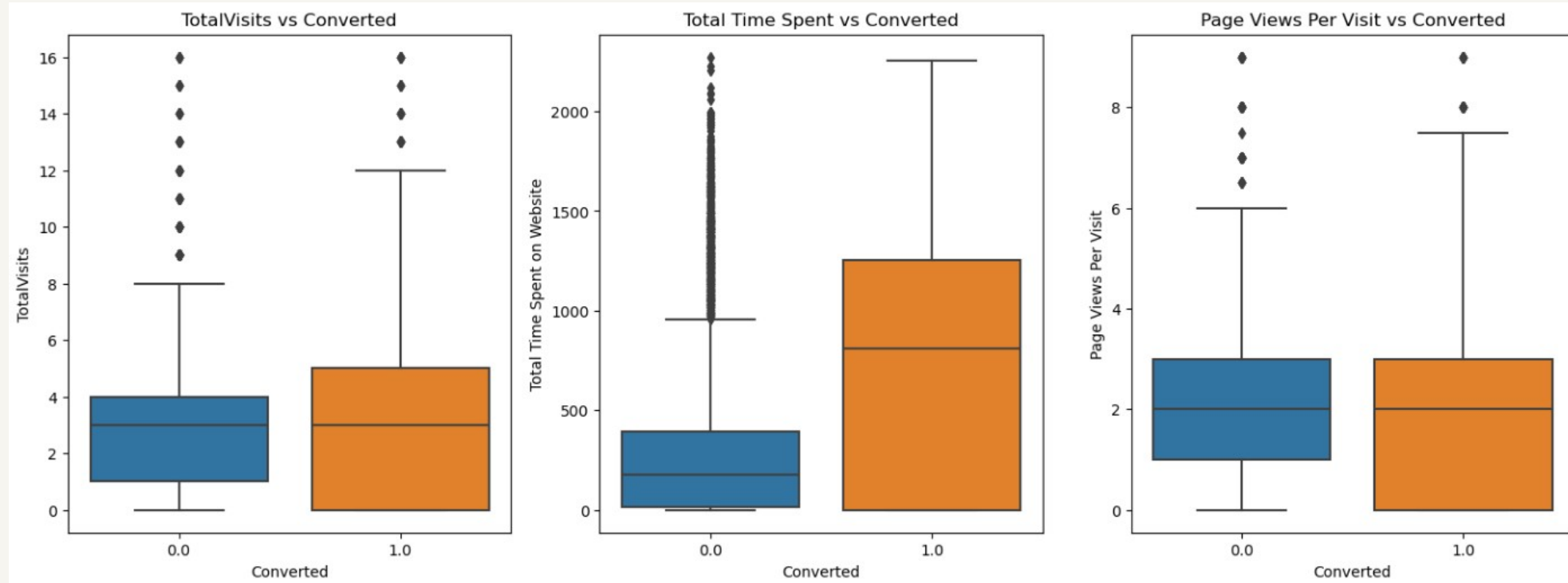


From the graph, we can see that both API and Landing Page Submission (LPS) bring most of the leads and also have a significant lead conversion rate. While the lead conversion rate for Lead Add Form (LAF) is very high, the total no of leads are low. So the organization must focus on the LAF to bring in more leads.

INFERENCES- CATEGORICAL VARIABLE ANALYSIS

- Lead origin:- The majority of leads are brought in by both API and Landing Page Submission (LPS), and both have a sizable lead conversion rate.
 - Conversion rate is also high.
- Lead source: - Google and Direct traffic have the maximum leads.
 - The Conversion rate is low.
- Country: - No. of users are low when compared to India.
- City: - Most of the leads are from Mumbai, it has high conversion rate.
- Specialization:- Most of the leads have management specialization.
- Occupation: - Most of the leads where from unemployed section
- Tags :- Maximum no. of users are most likely to enrol after reading the email.

NUMERICAL VARIABLE ANALYSIS



From the above plots, we can see that:

- The median for converted and not converted leads is the same in the "Total Visits" and "Page Views Per Visit" column. Therefore no conclusion can be drawn from this
- The lead conversion in the "Total Time Spent on Website" is high, and therefore the organization must focus on improving their website to attract more customers

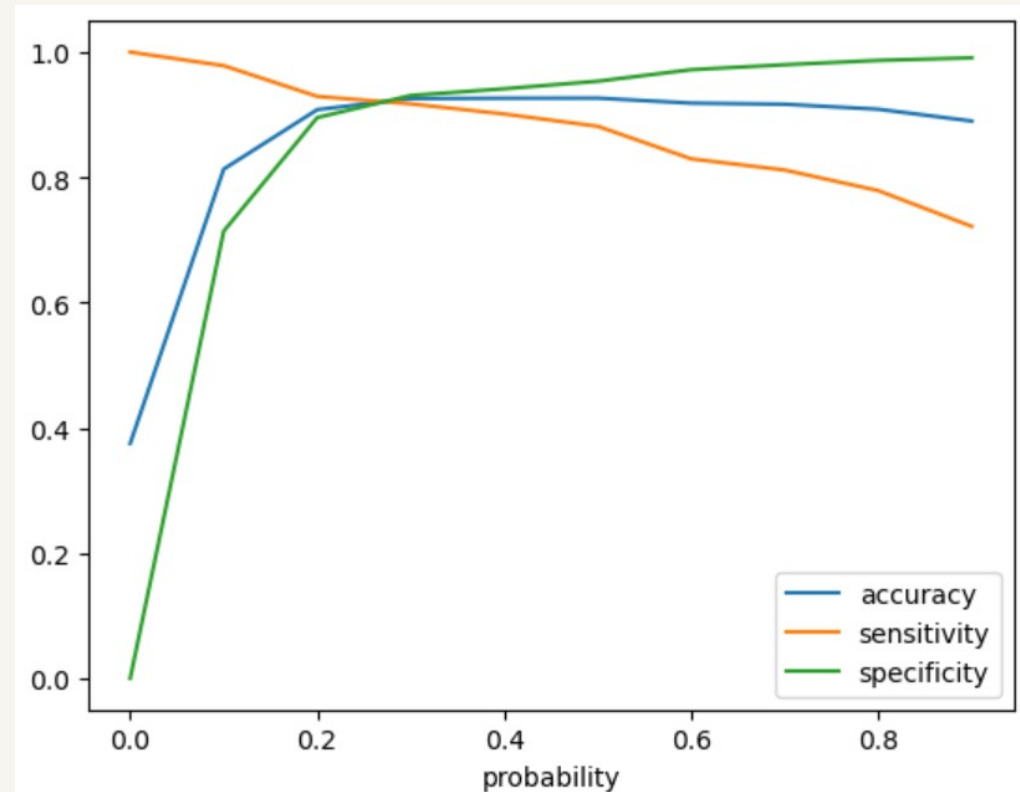
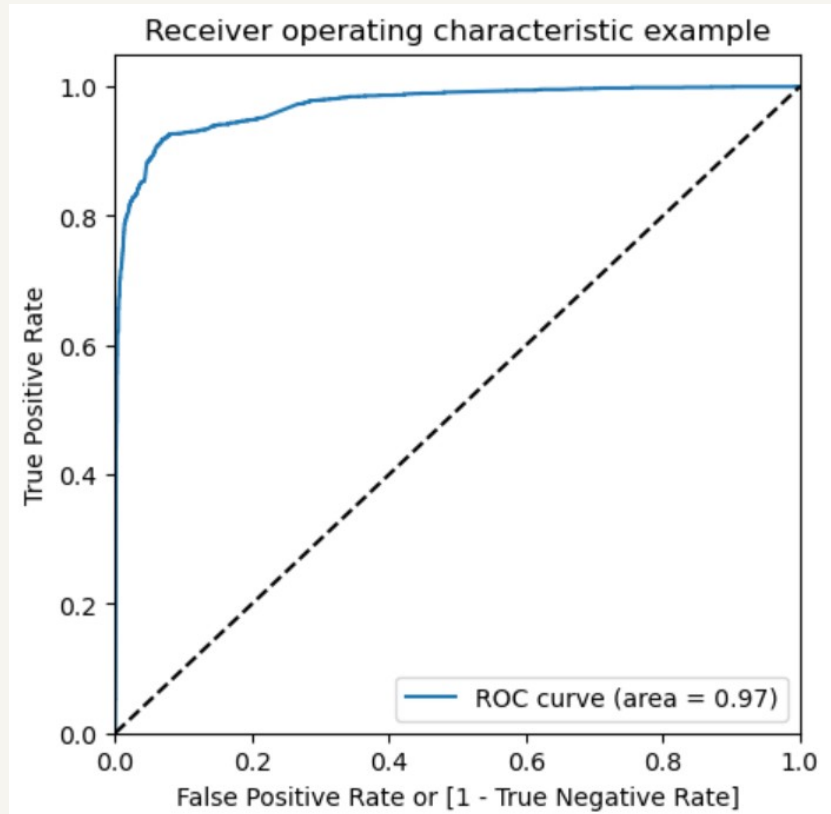
INFERENCES-NUMERICAL VARIABLE ANALYSIS

- 'page views per visit' and 'Total visits' are similar, having outliers skewed.
- The median for converted and not converted leads are same in total visits and page views per visit
- The lead conversion in the 'Total time spent on website' is high.
- The company must focus on website to attract more customers.

MODEL BUILDING

- After Data conversion Numerical variables are normalised.
- Dummy variables are created for object type variables.
- In Data Building, Splitting the Data into Training and Testing sets
- By performing a train-test split, we chosen 7:3
- Use RFE for feature selection
- Running RFE with 15 variables as output
- MODEL BUILDING by removing variable whose p-value is greater than 0.05 and vif (variance inflation factor) value is greater than 5
- Predicting values of train data set
- We have accuracy of over 90% for train dataset.

ROC CURVES



- Optimal cut off probability is that probability where we get balanced sensitivity and specificity.
- From the second graph we can see 0.3 is the most optimal cut off having the accuracy of 0.9256.

MODEL METRICS- TRAIN DATASET

Confusion matrix	
3727(TN)	182(FN)
278(FN)	2068(TP)

Observations:

- Accuracy = 92%
- Sensitivity = 91%
- Specificity = 93%

MODEL METRICS- TEST DATASET

Confusion matrix	
1515(TN)	120(FN)
79(FN)	968(TP)

Observations:

- Accuracy = 92%
- Sensitivity = 92%
- Specificity = 92%

We can see that the model is able to predict the conversion rate with an accuracy of over 90 %.

CONCLUSION

- The leads having high 'lead score' can be focused on more for better conversion rate.
- Since the number of leads is high in Mumbai as compared to other major cities, the company can increase marketing in other cities as well to achieve more leads.
- The unemployed category can be focused on more and having management as specialization.
- The lead score calculated in the test set of data shows the conversion rate of over 90% on the final prediction model which clearly meets the expectation of CEO has given a ballpark of target lead conversion rate to be around 80%.
- Good value of sensitivity of our model will help to select the most promising leads.
- The main features that have greater impact on probability that a lead will convert are:
 - Lead origin, What is your current occupation, Total time spend on website.

THANK YOU