

LEAD SCORING CASE STUDY

BY:

SHREEVATSA HEGDE, SOUBHAGYA PANDA, CHANDRASHEKAR SAVADATTI

PROBLEM STATEMENT

- X Education sells online courses to industry professionals.
- X Education gets a lot of leads, 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'.
- 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

- X education wants to know most promising leads.
- For that they want to build a Model which identifies the hot leads.
- Deployment of the model for the future use.

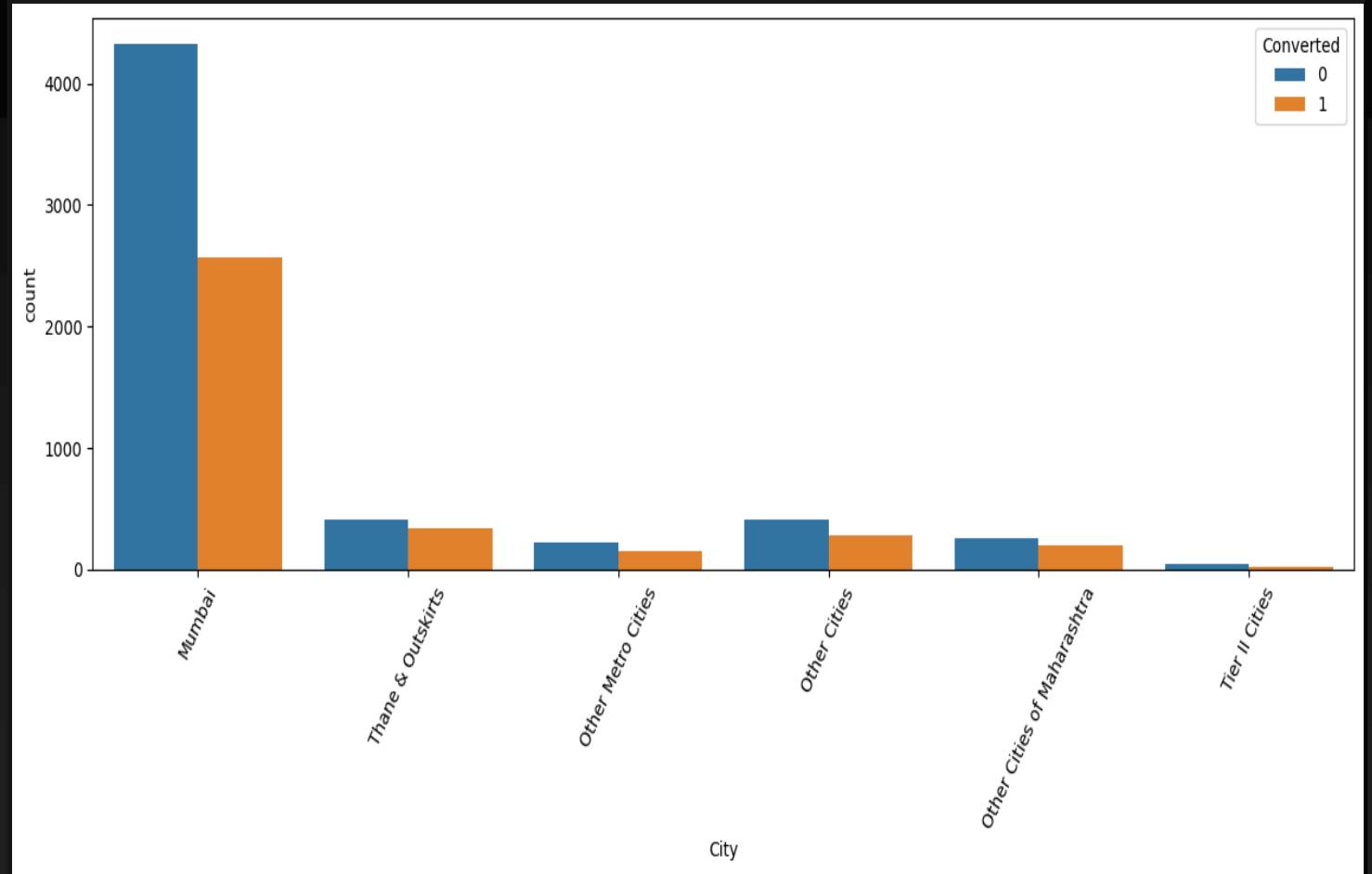
SOLUTION METHODOLOGY

1. Data cleaning and data manipulation.
 1. Check and handle duplicate data.
 2. Check and handle NA values and missing values.
 3. Drop columns, if it contains large amount of missing values and not useful for the analysis.
 4. Imputation of the values, if necessary.
 5. Check and handle outliers in data.
2. EDA
 1. Univariate data analysis: value count, distribution of variable etc.
 2. Bivariate data analysis: correlation coefficients and pattern between the variables etc.
3. Feature Scaling & Dummy Variables and encoding of the data.
4. Classification technique: logistic regression used for the model making and prediction.
5. Validation of the model.
6. Model presentation.
7. Conclusions and recommendations.

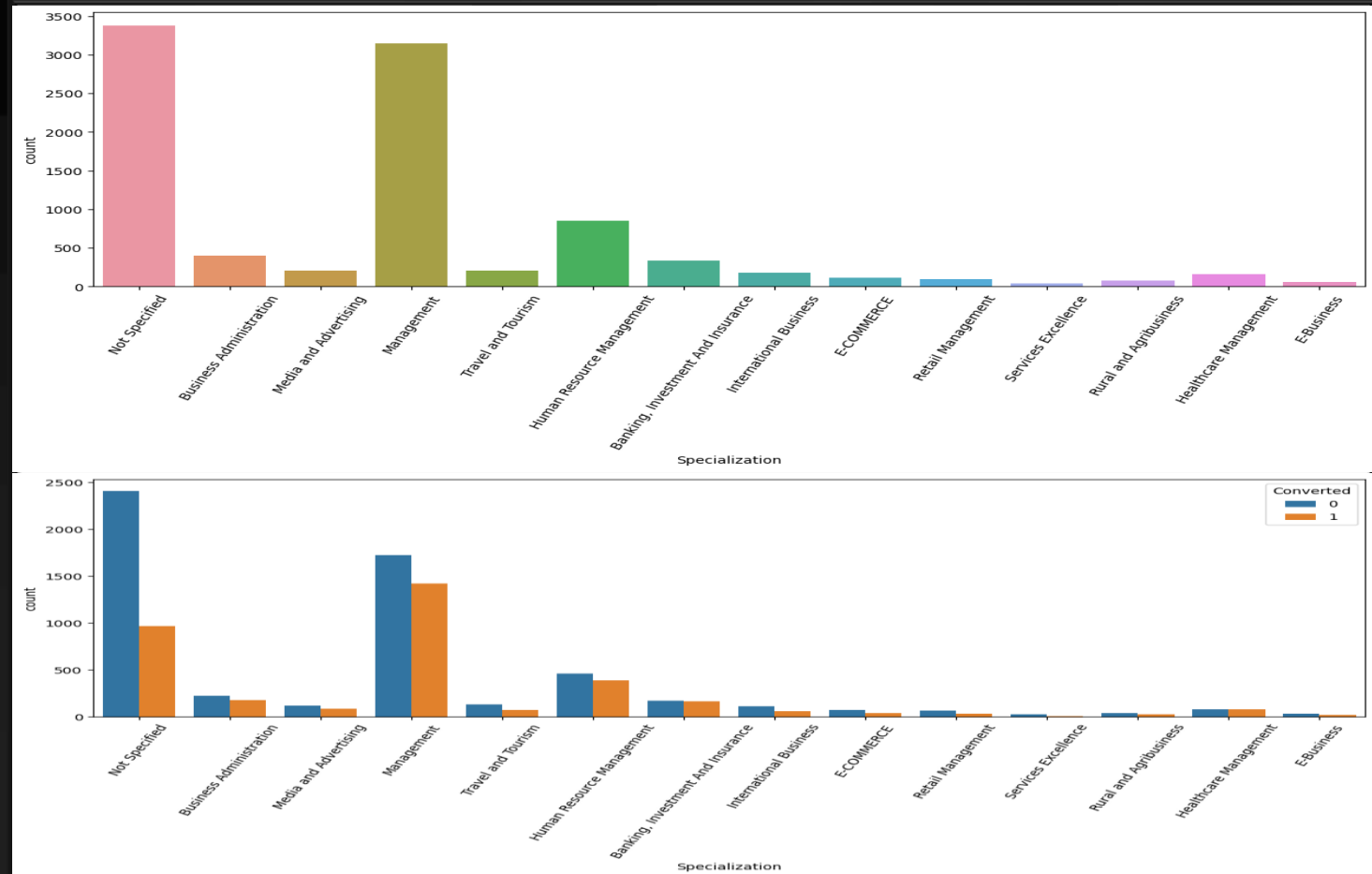


EXPLETORY DATA ANALYSIS

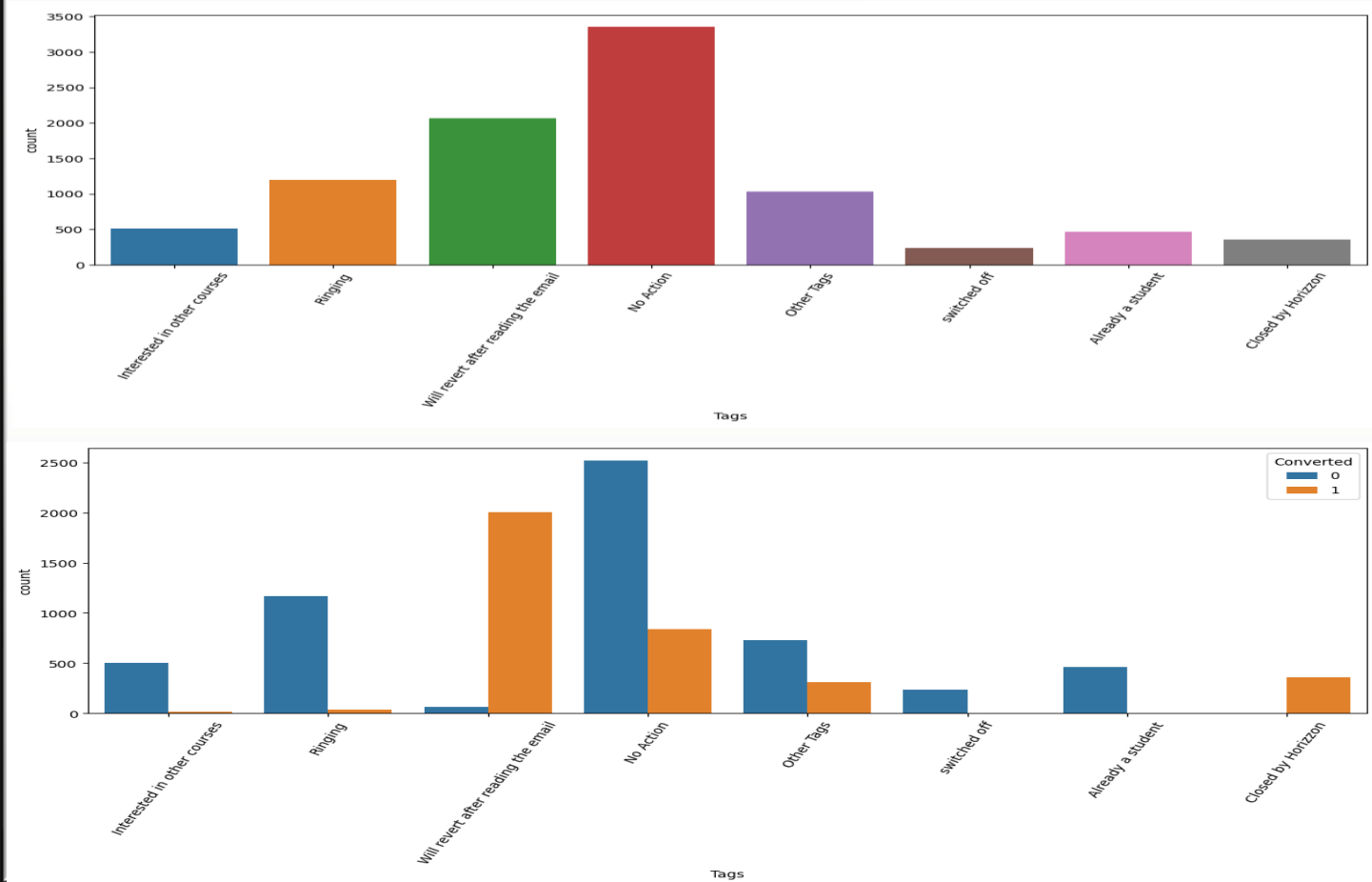
- We can see that the Mumbai City has the Highest number of Lead generation and Converted Leads



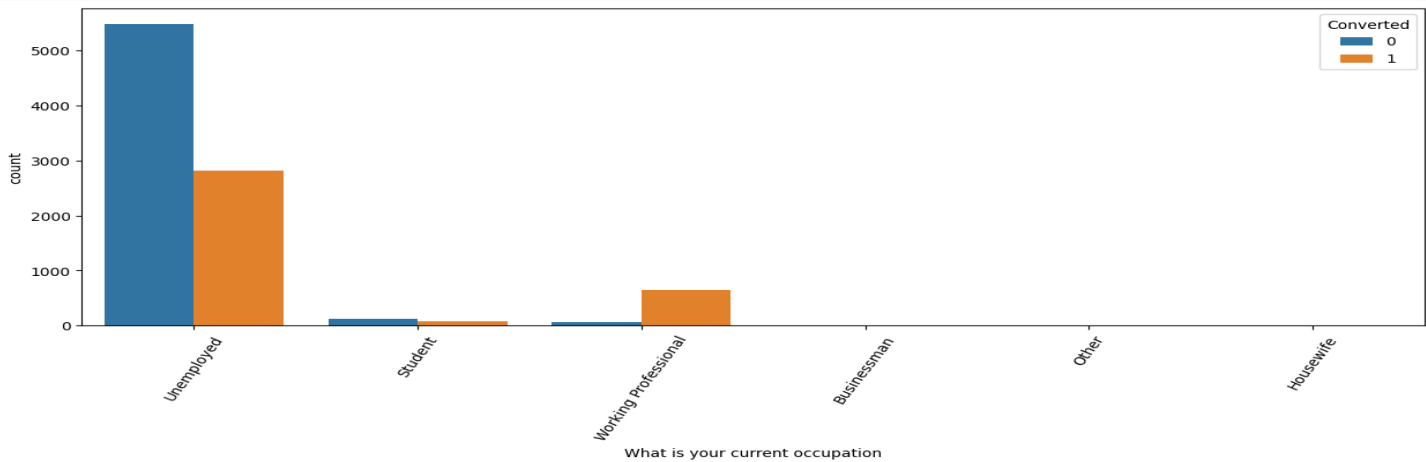
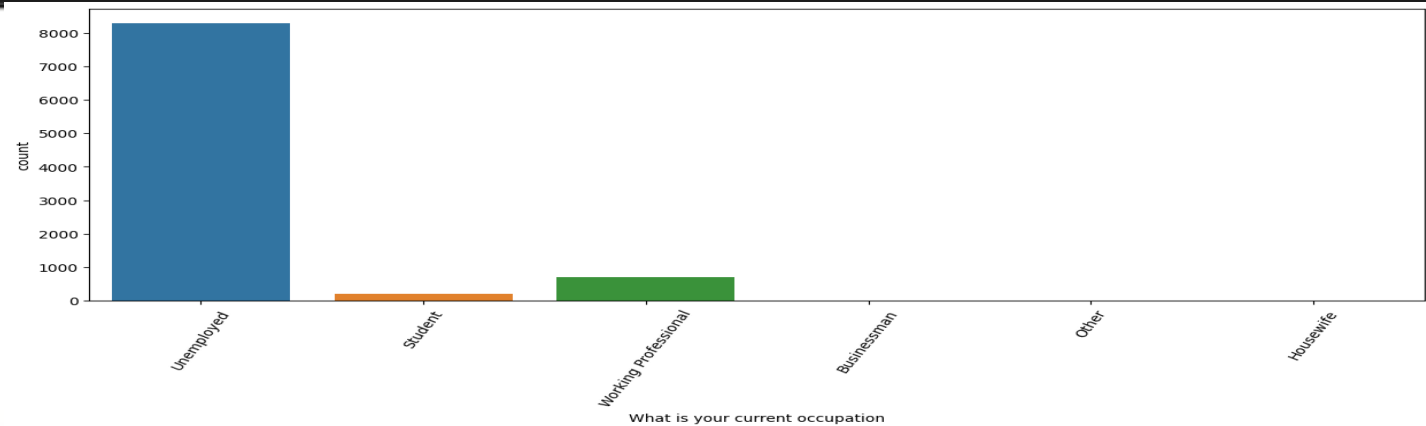
- As we can see that Management has highest Leads conversation
- High number of Leads are generated where specialization is not specified



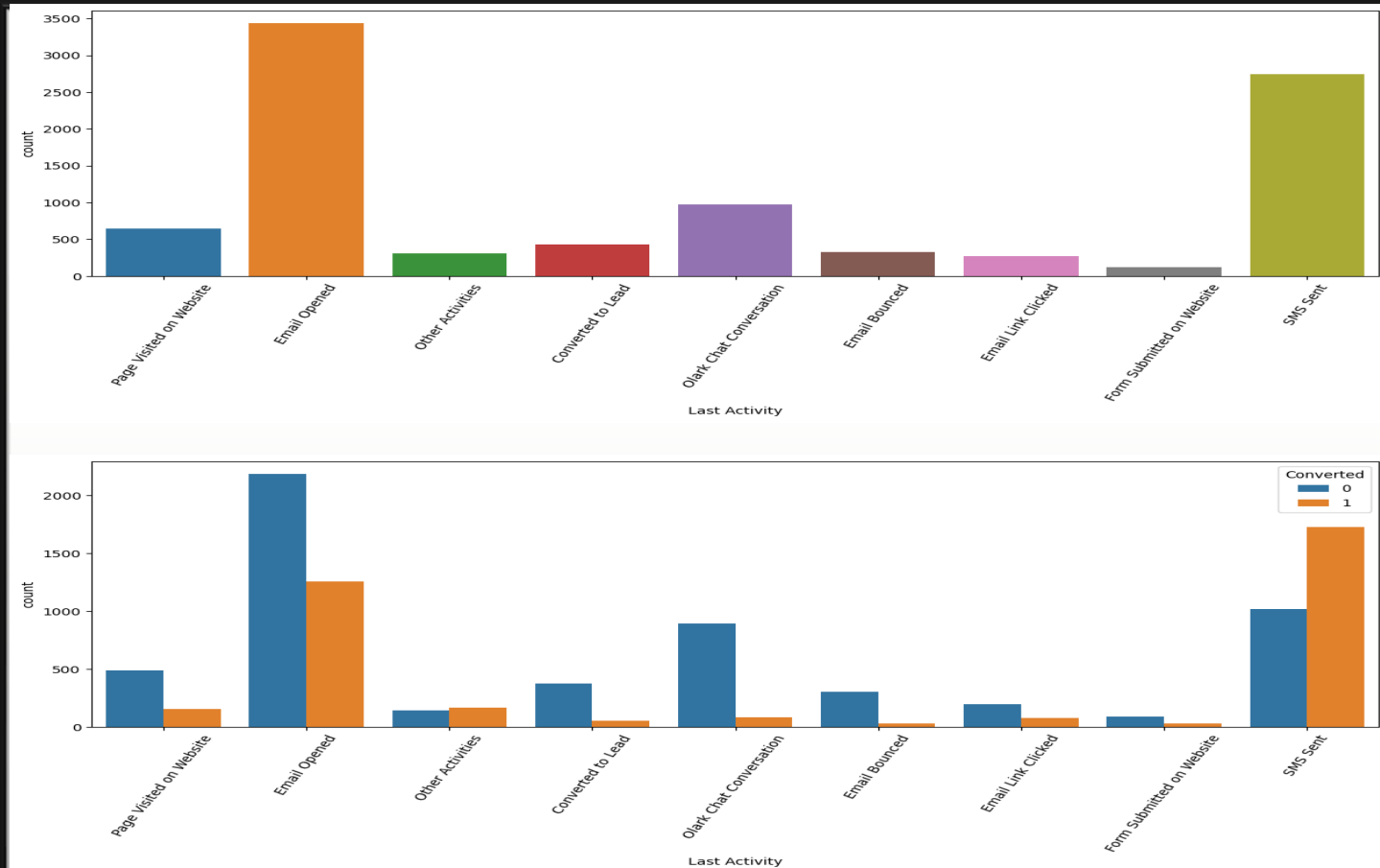
- Highest number of Leads are tagged as No Action
- Highest number of converted Leads are Tagged as will revert after reading the mail
- Closed by Horizzon tags have high percentage of Lead conversation



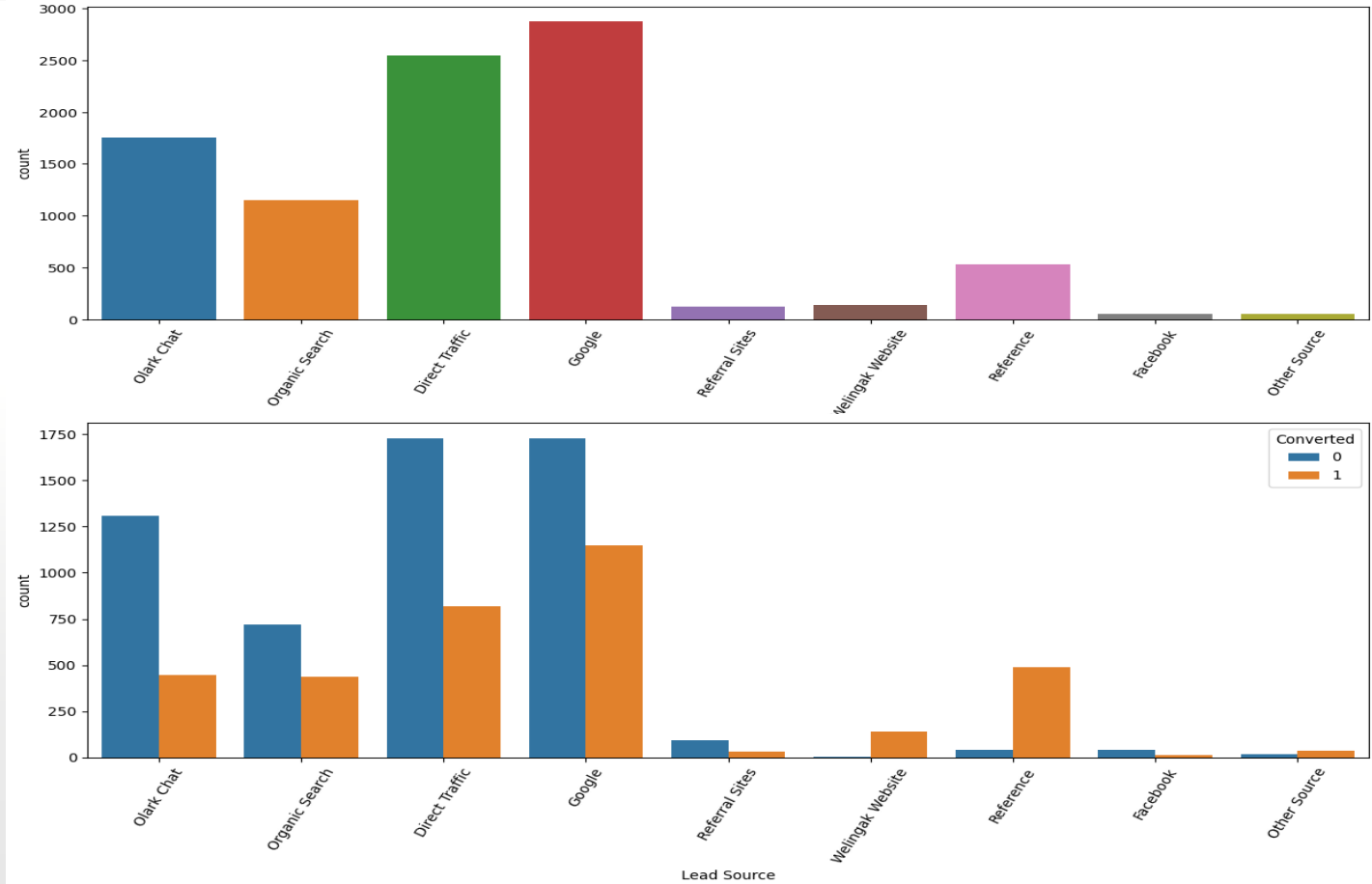
- As we can see here
Unemployed Customers
has more number of
Leads
- Working Professional
Customers has the high
Lead conversation



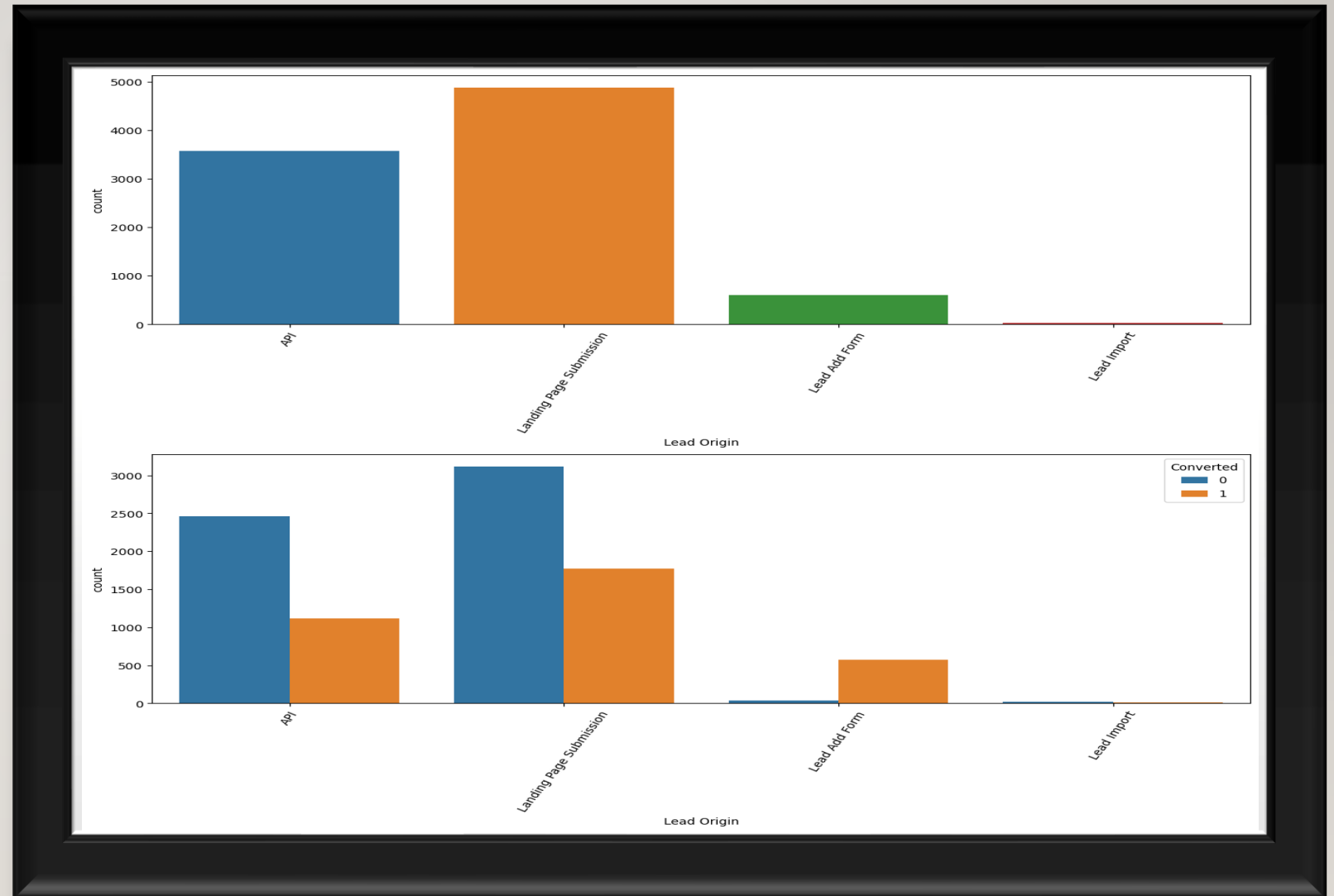
- As we can see here
Highest number of Leads
has Email opened as its
Last Activity
- we can see that when
the Last Activity is SMS
Sent then it has high
Lead conversion



- Most number of Leads are coming from Google and Direct Traffic
- Reference has the highest Percentage of Lead conversion
- Highest number of converted Leads are coming from Google



- Highest number of Leads are coming from Landing Page Submission and API
- Lead Add Form has highest percentage of Lead Conversion
- In order to improve overall lead conversion rate, we have to improve lead conversion of API and Landing Page Submission origin and generate more leads from Lead Add Form.

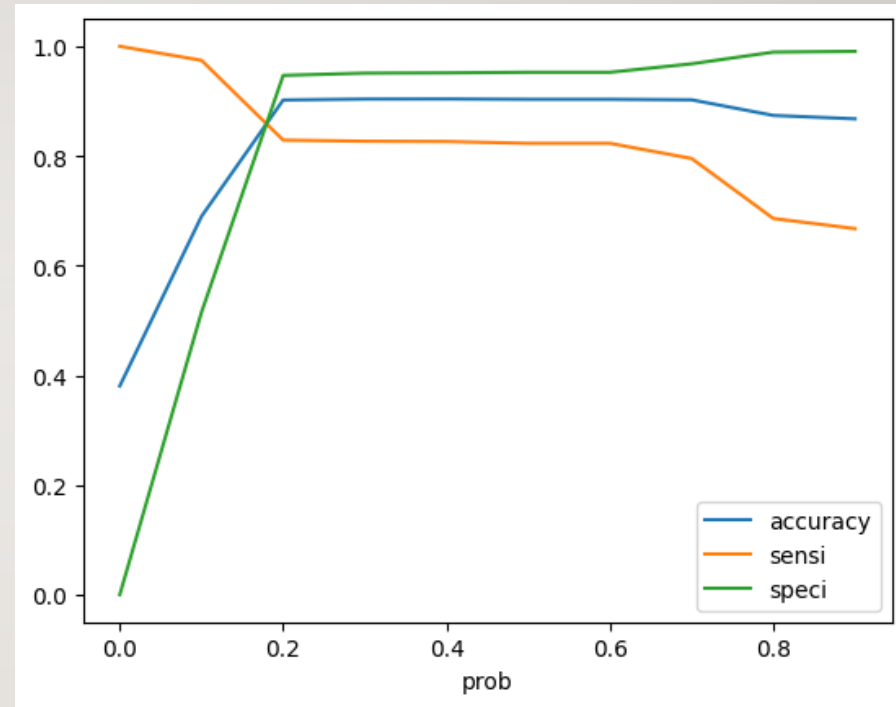
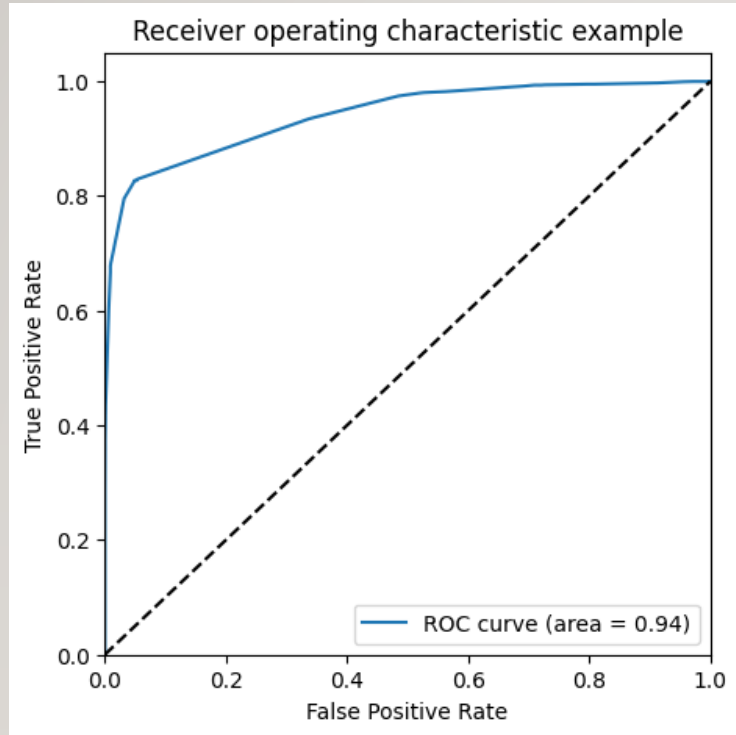


DATA CONVERSION

- Numerical Variables are normalized
- Dummy variable are created for object type variables
- Total number Rows for analysis: 6267
- Total numbers of Columns: 48

MODEL BUILDING

- Split the Data in to Training and Testing Dataset
- The first basic step for regression is performing a train-test split, we have chosen 70:30 ratio.
- Use RFE for Feature Selection
- Running RFE with 15 variables as output
- Building Model by removing the variable whose p-value is greater than 0.05 and VIF value is greater than 5
- Predictions on test data set
- Overall Accuracy 90%



WE HAVE GOOD ROC CURVEVALUE OF 0.94

WE HAVE BALANCED SENSITIVITY AND SPECIFICITY OPTIMAL CUT OFF AT 0.2

CONCLUSION

- It was found that the variables that mattered the most in the potential Lead conversion are (In descending order) :
 - The total time spend on the Website.
 - When the lead source was:
 - a. Google
 - b. Direct traffic
 - c. Organic search
 - d. Reference
 - When the last activity was:
 - a. SMS Sent
 - b. Email Opened
 - When the lead origin is Lead add format.
 - When their current occupation is as a working professional. Keeping these in mind the X Education can flourish as they have a very high chance to get almost all the potential buyers to change their mind and buy their courses.