Lead Score Case Study

Logical Regrassion

Problem Statement:

- An education company named X Education sells online courses to industry professionals. The company markets its courses on several websites and search engines like Google.

- X education wants most promising leads that can be converted to paid customers.

- Though the company gets a lot of leads but around 30% of them gets converted. That is out of 100 leads only 30 leads get 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.

- We need to built a model where lead conversion rate is around 80%

Strategy

- Importing all necessary libraries
- Reading the data and inspecting them for basic data understanding
- Taking care of missing values and outliers
- Exploratory Data Analysis for finding out most helpful columns for building model
- Prepare the data for model building
- Split the data in train and test sets
- Scaling features
- Built a logistic regression model
- Model evaluation
- Test the model on test set
- Measure the accuracy and other metrices

Lead origin vs Converted: API and Landing Page Submission bring higher number of leads as well as conversion.Lead Add Form has a very high conversion rate but count of leads are not very high.
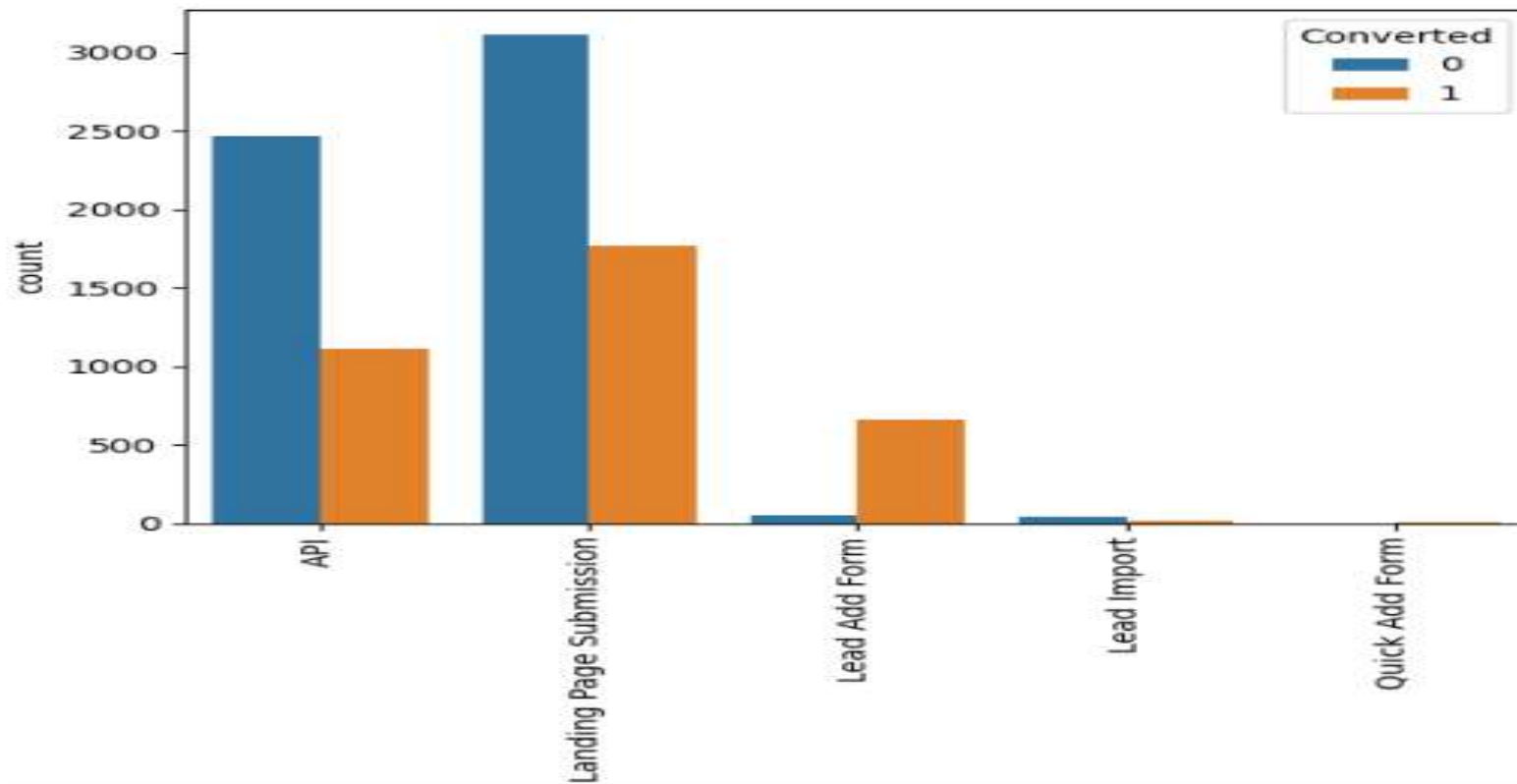
Lead Source VS Converted: Top Three lead source are Google, Direct Traffic and Olark Chat followed by Organic search. Google has more converted leads. Reference has higher conversion ratio but leads are low.
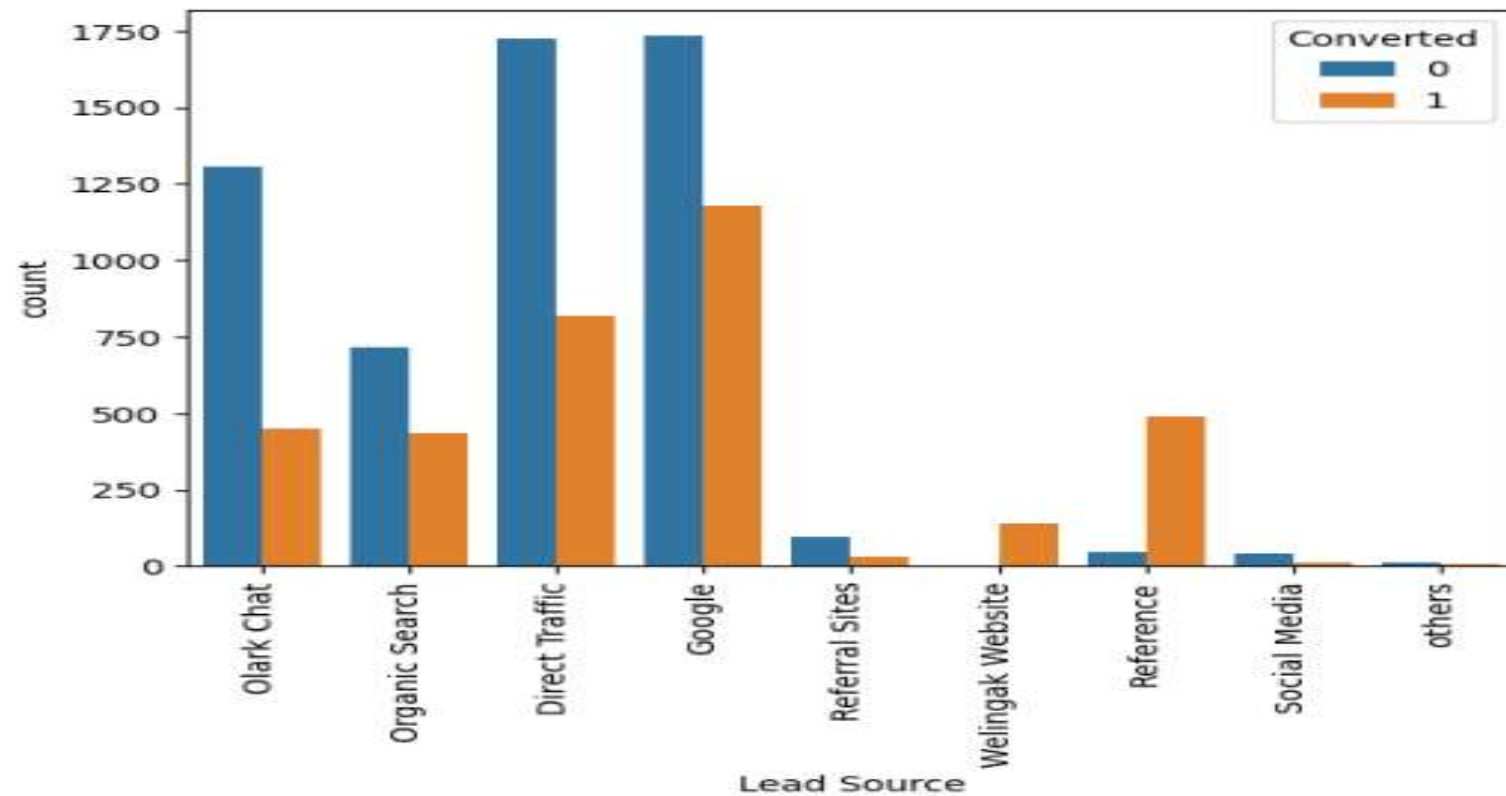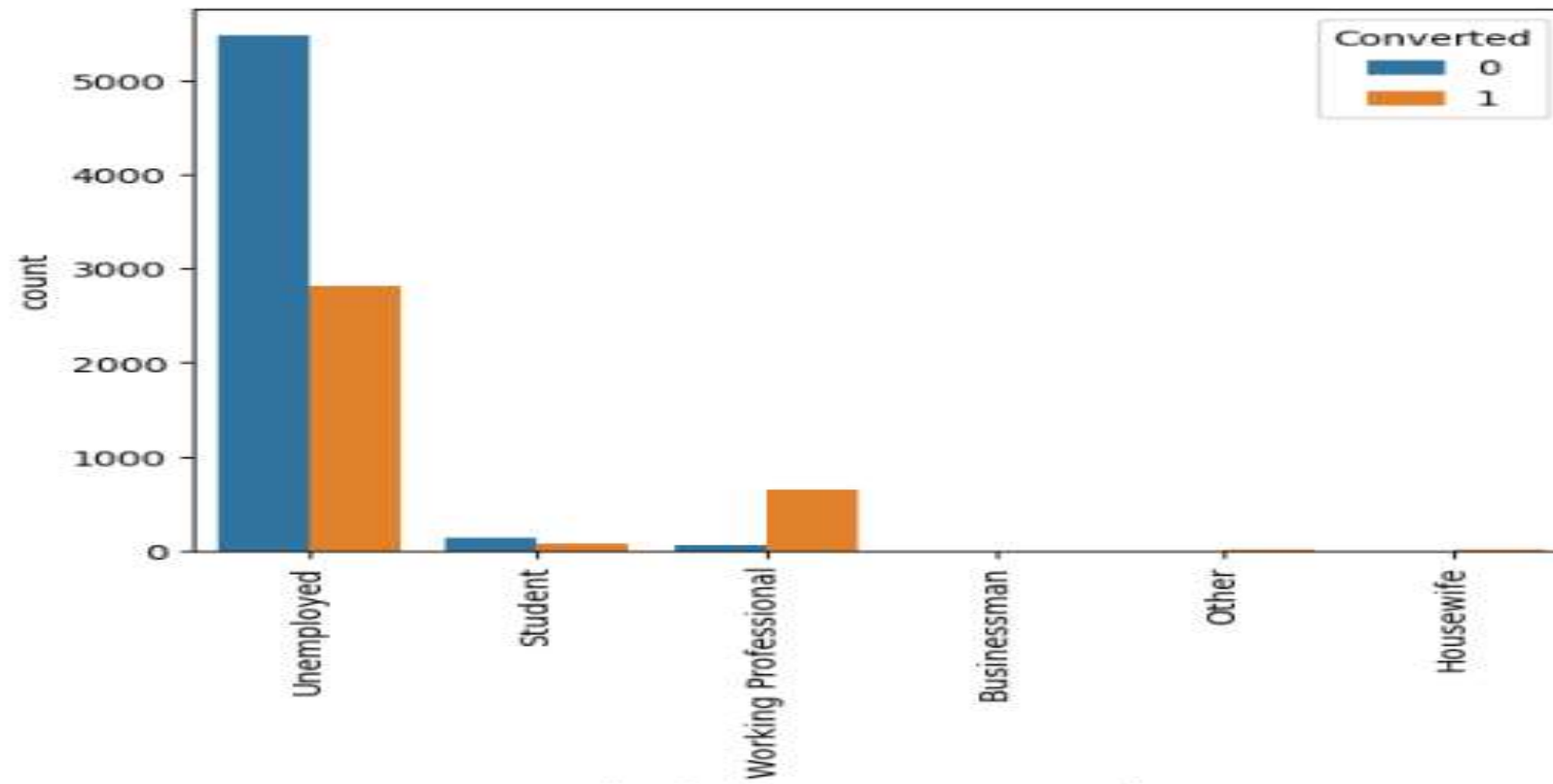
Last notable activity vs Converted: Mostly a leads get converted if potential leads opened the mail or get SMS.
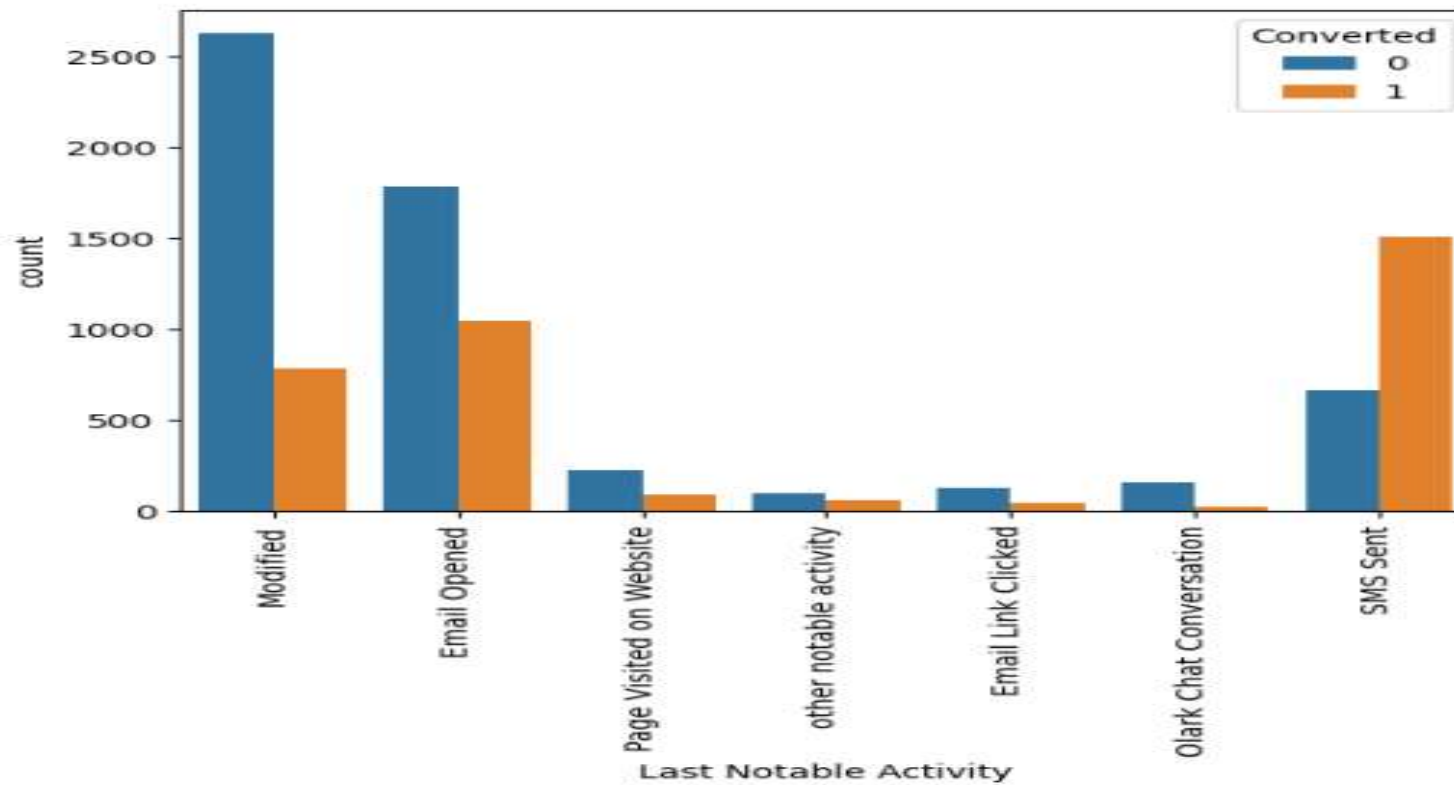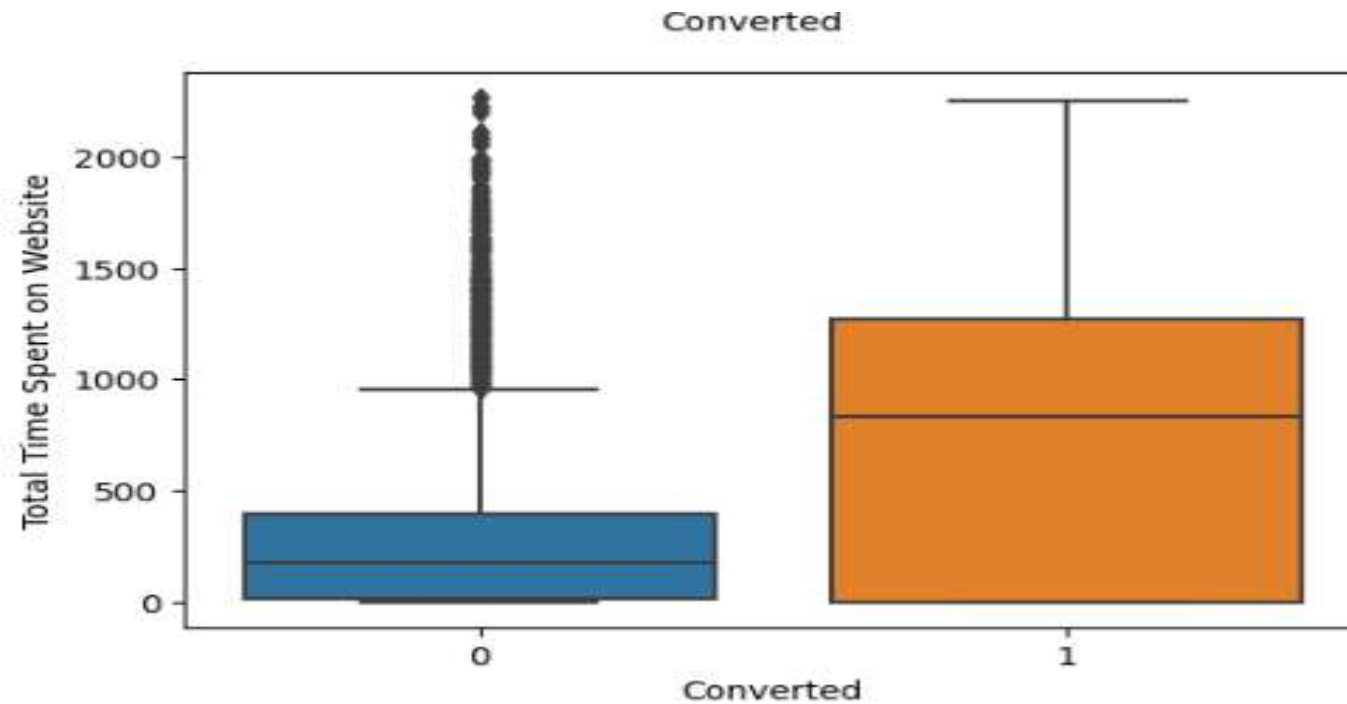
City VS Converted: Higher number of leads are from mumbai as well as conversion.
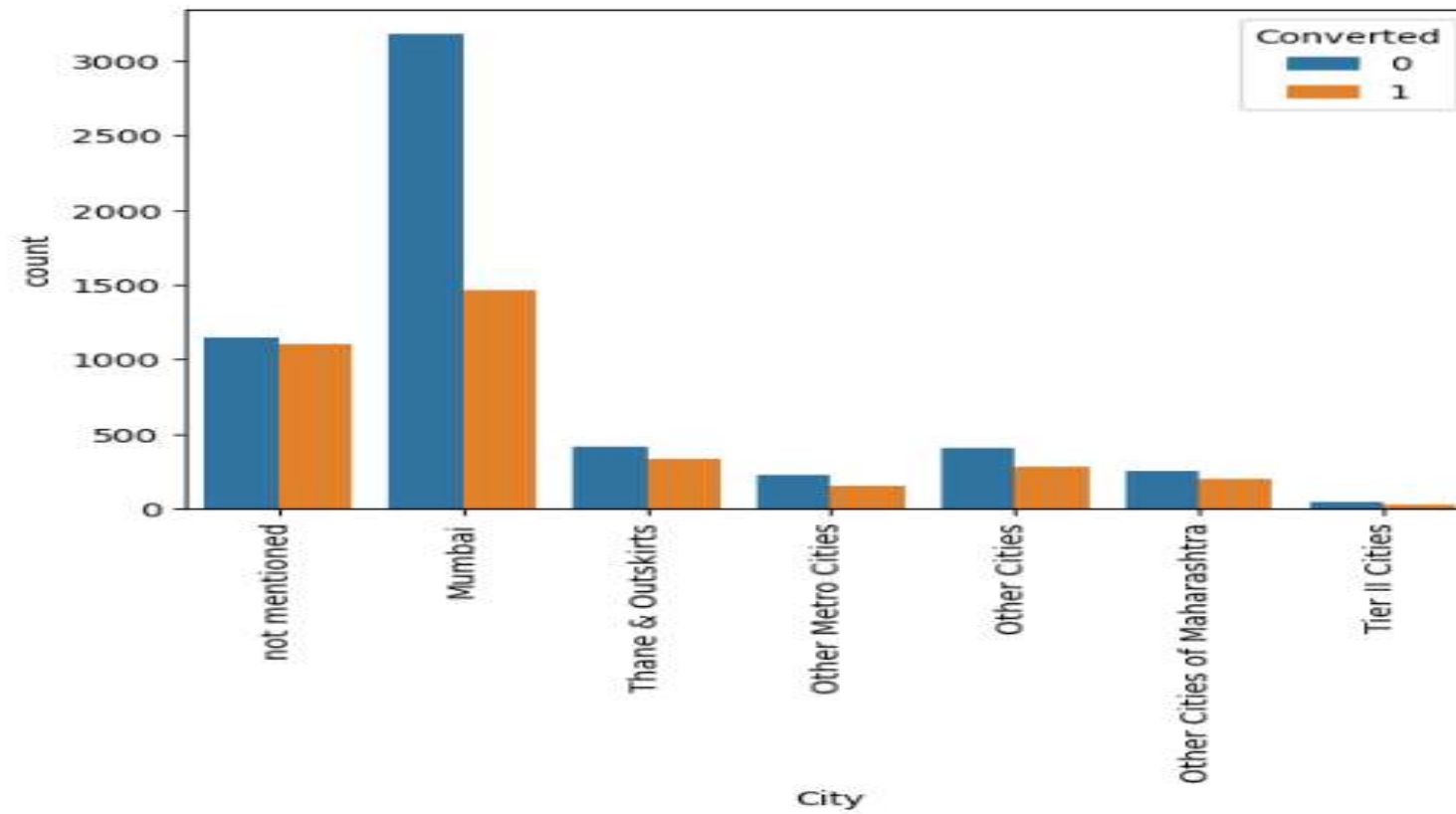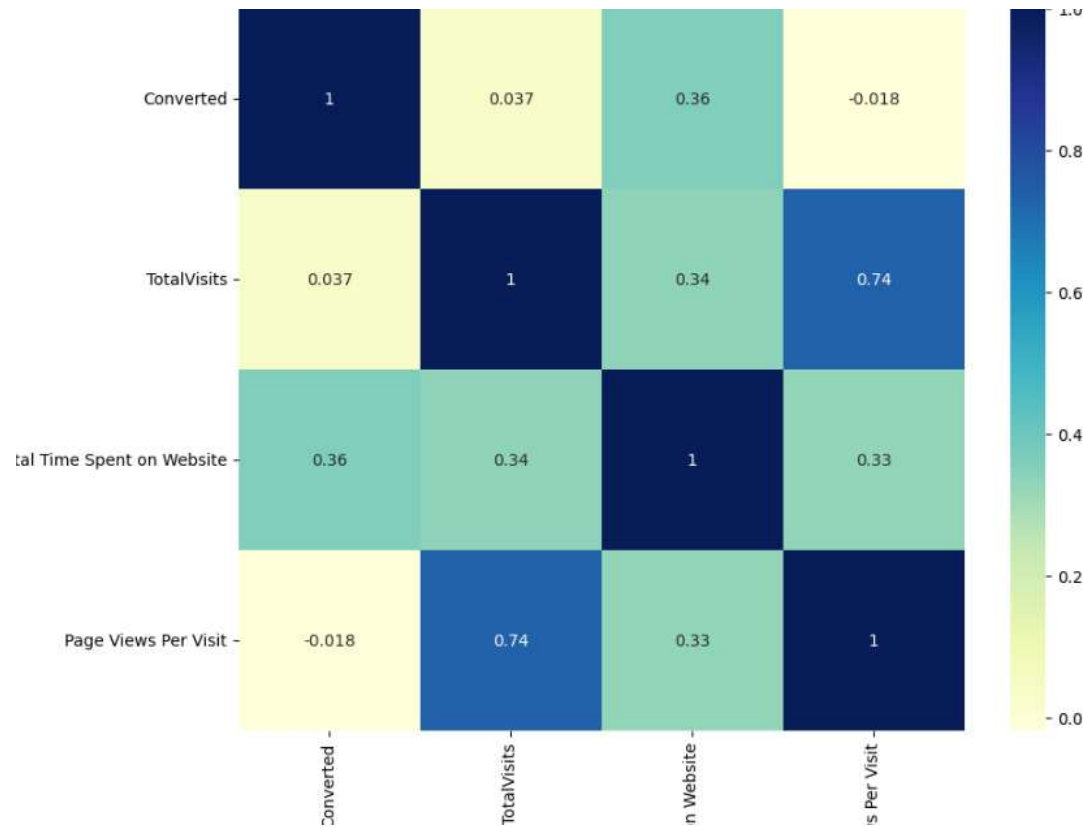
Occupation Vs Converted: From unemployed category we get higher no of leads as well as conversion. Working professional has higher conversion ate.
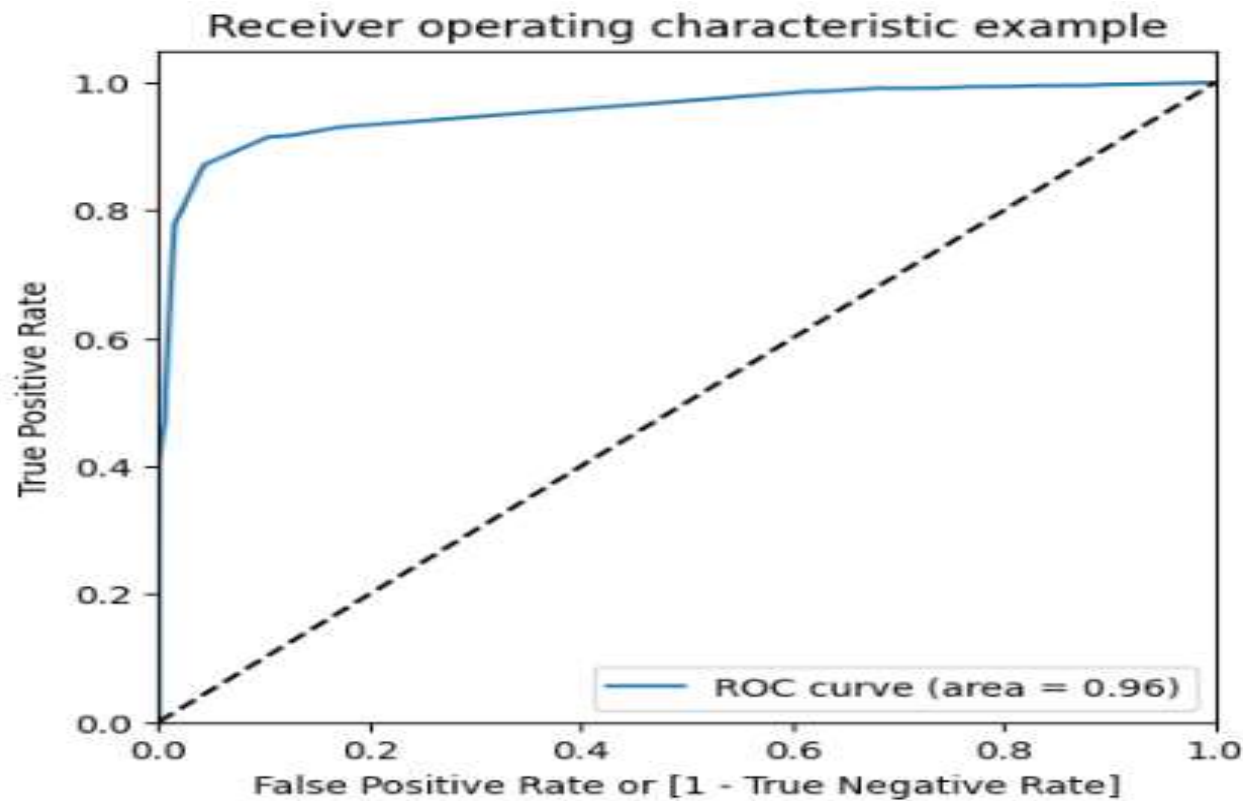
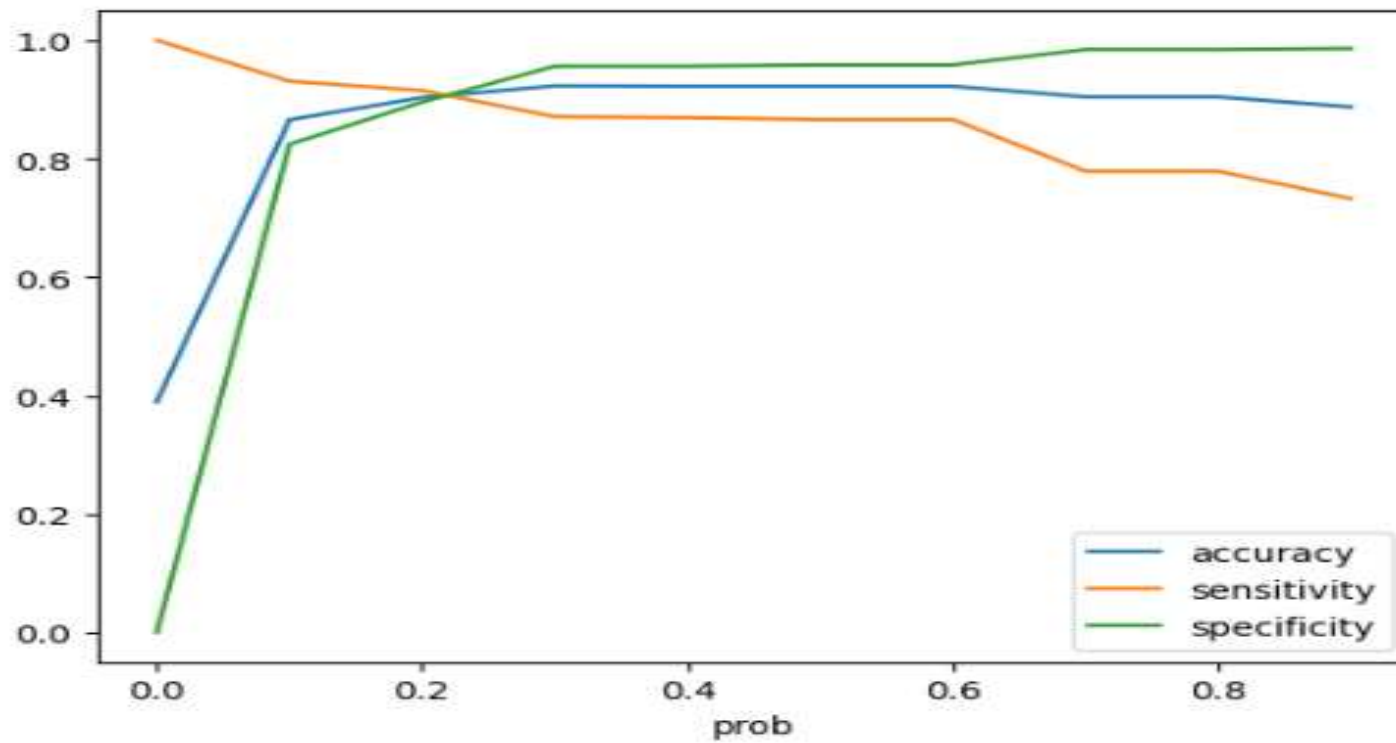# Total time Spent on Website vs Converted: More time spent on website more leads get converted.

Heat Map: Correlation among numerical variavles.

ROC CURVE: We are getting area under the curve at 0.96 which is a fair value.

Optimum Point at 0.3

Accuracy ,Sensitivity and Specificity on train and test set:

- **Train:**
- Accuracy: 92.30%
- Sencitivity: 86.61%
- Specificity: 95.82%
- **Test:**
- Accuracy: 96.69%
- Sencitivity: 86.61%
- Specificity: 95.82%

Findings:

- 1.API and Landing Page Submission bring higher number of leads as well as conversion. Lead Add Form has a very high conversion rate but count of leads are not very high, so we need to bring more lead through this.Lead Import and Quick Add Form get very few leads.

- 2.Top Three lead source are Google, Direct Traffic and Olark Chat followed by Organic search. Google has more converted leads. We should work on converting leads from Ozark chat and direct traffic. Reference has higher conversion ratio but leads are low, so try to bring leads through reference.

- 3.From Last activity and last notable activity it is cleared that, mostly a leads get converted if potential leads opened the mail or get SMS.

- 4.Word of mouth and online search has higher converted least.

- 5.From unemployed category we get higher no of leads as well as conversion. We need to focus on working professional as there is higher chance of conversion.

- 6.Higher number of leads are from Mumbai as well as conversion. We should try to bring leads from thane and other cities as it has higher conversion rate.

- 7.More time leads spent on website more chances to get it converted.

Thank You