

# Lead Score Case Study

---

SADANA  
RUPESH  
SAHIL

# Problem Statement

---

1. X Education gets a lot of leads, They acquire 100 leads in a day, only about 30 of them are converted.
2. To make this process more efficient, the company wishes to identify the most potential leads, also known as 'Hot Leads'.
3. After successfully identifying set of leads, the lead conversion rate should go up as the sales team will now focus more on communicating with the potential leads rather than making calls to all leads.

# Business Objective

---

1. X education wants to know most promising leads.
2. They want to build a Model which identifies the hot leads by giving score.
3. Model should be able to adjust for future changes.

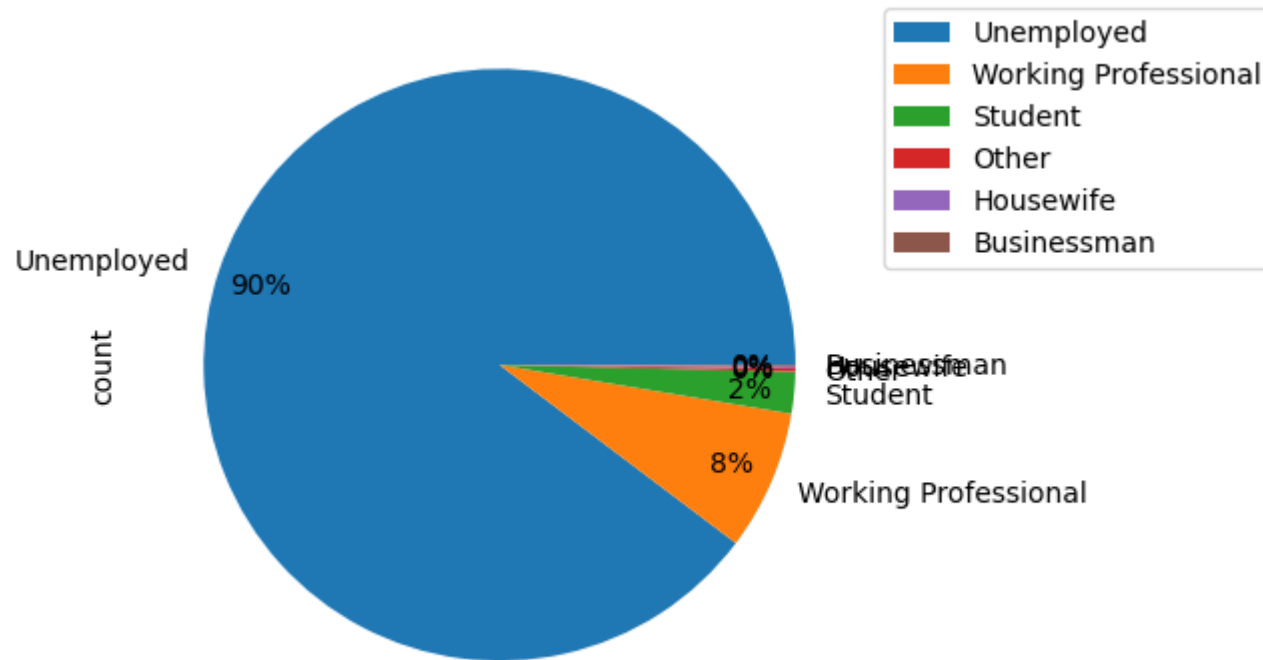
# Solution Steps

---

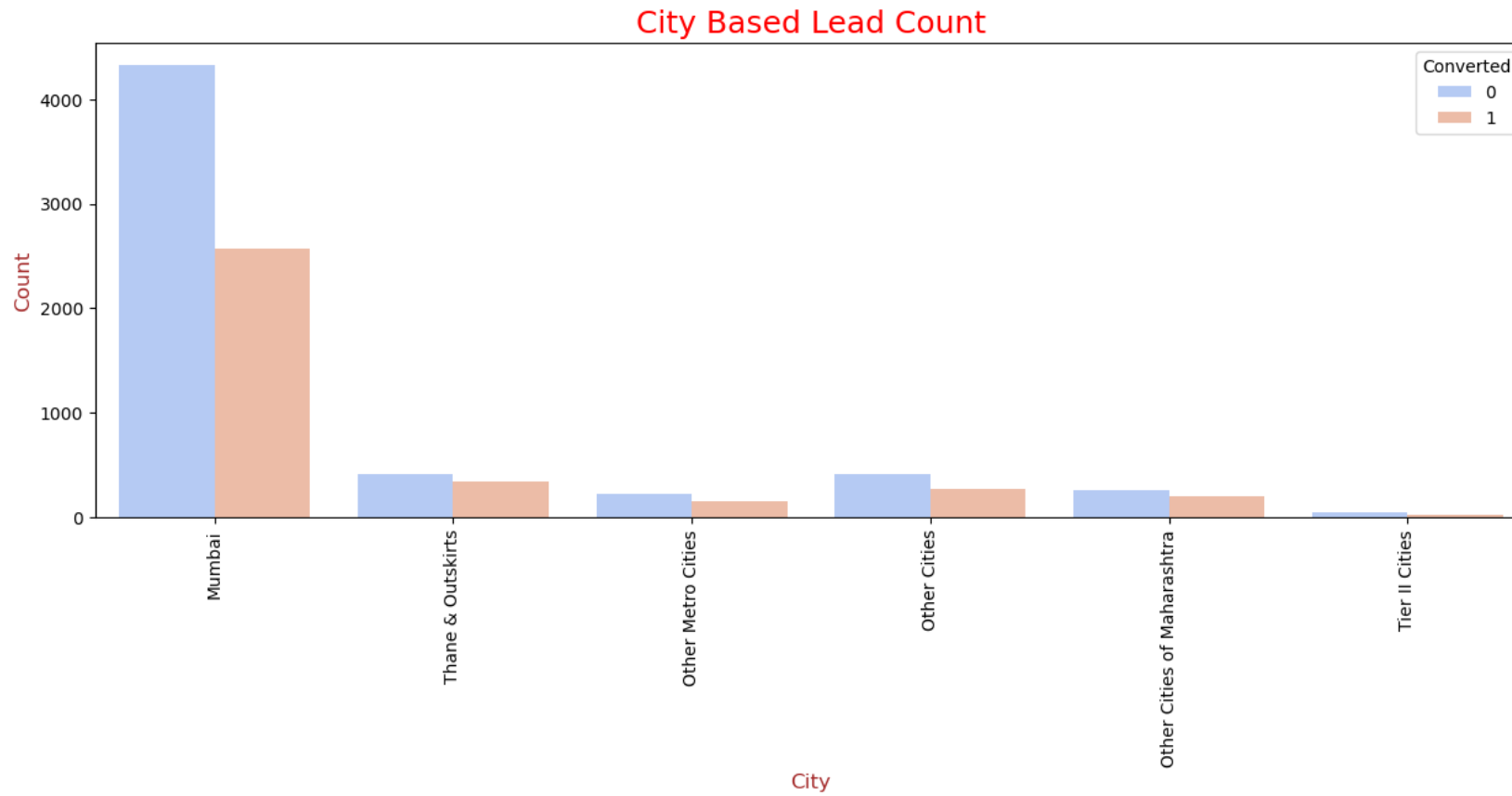
1. Reading Data
2. Cleaning Data
3. EDA
4. Creating Dummy variables
5. Splitting data into train and test set
6. Building Model using Logistic Regression
7. Making Predictions
8. Model Evaluation.

# Exploring Data

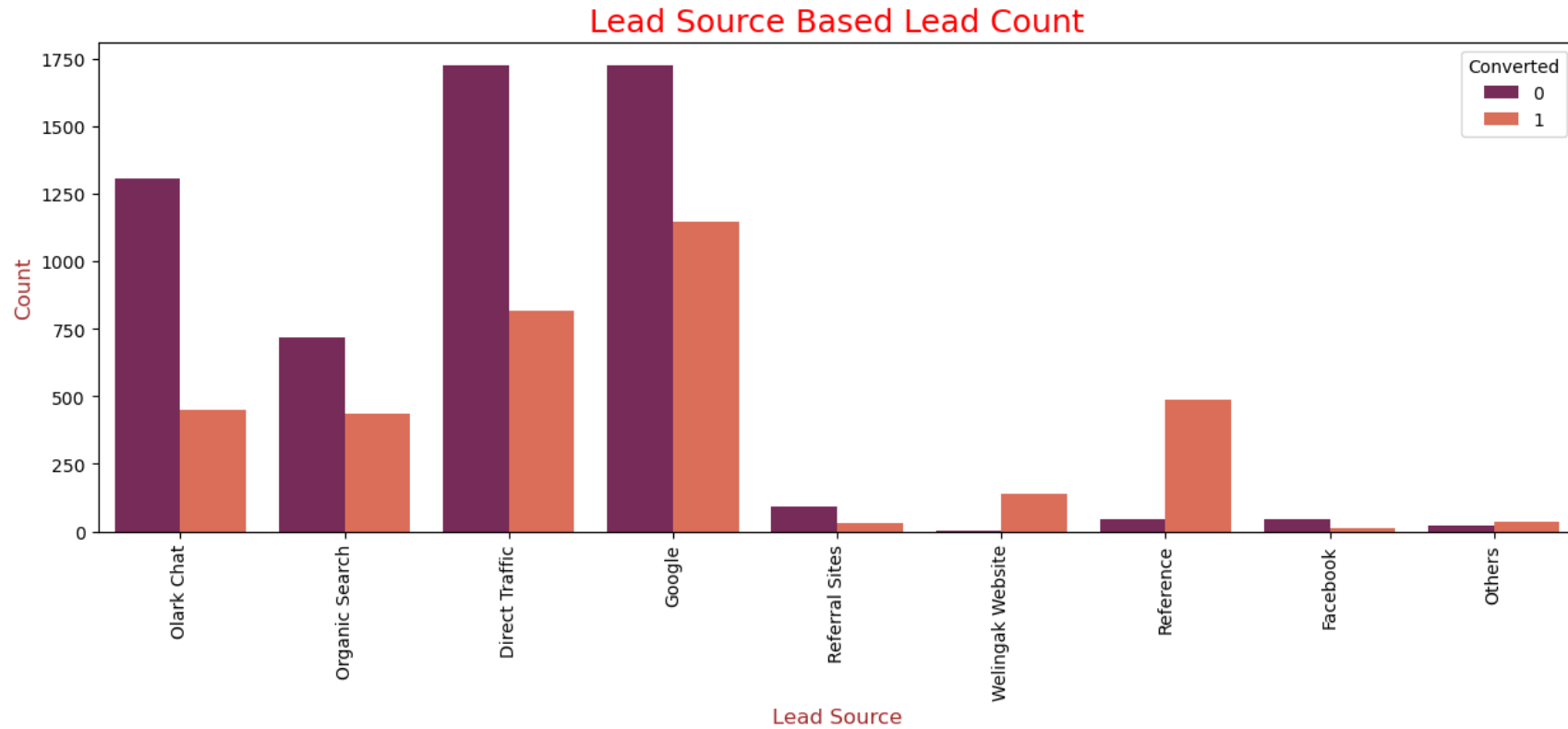
---



- Unemployed individuals seem to submit more inquiries.
- Working professionals have 2<sup>nd</sup> highest enquiries.
- Students have 3<sup>rd</sup> highest enquiries.

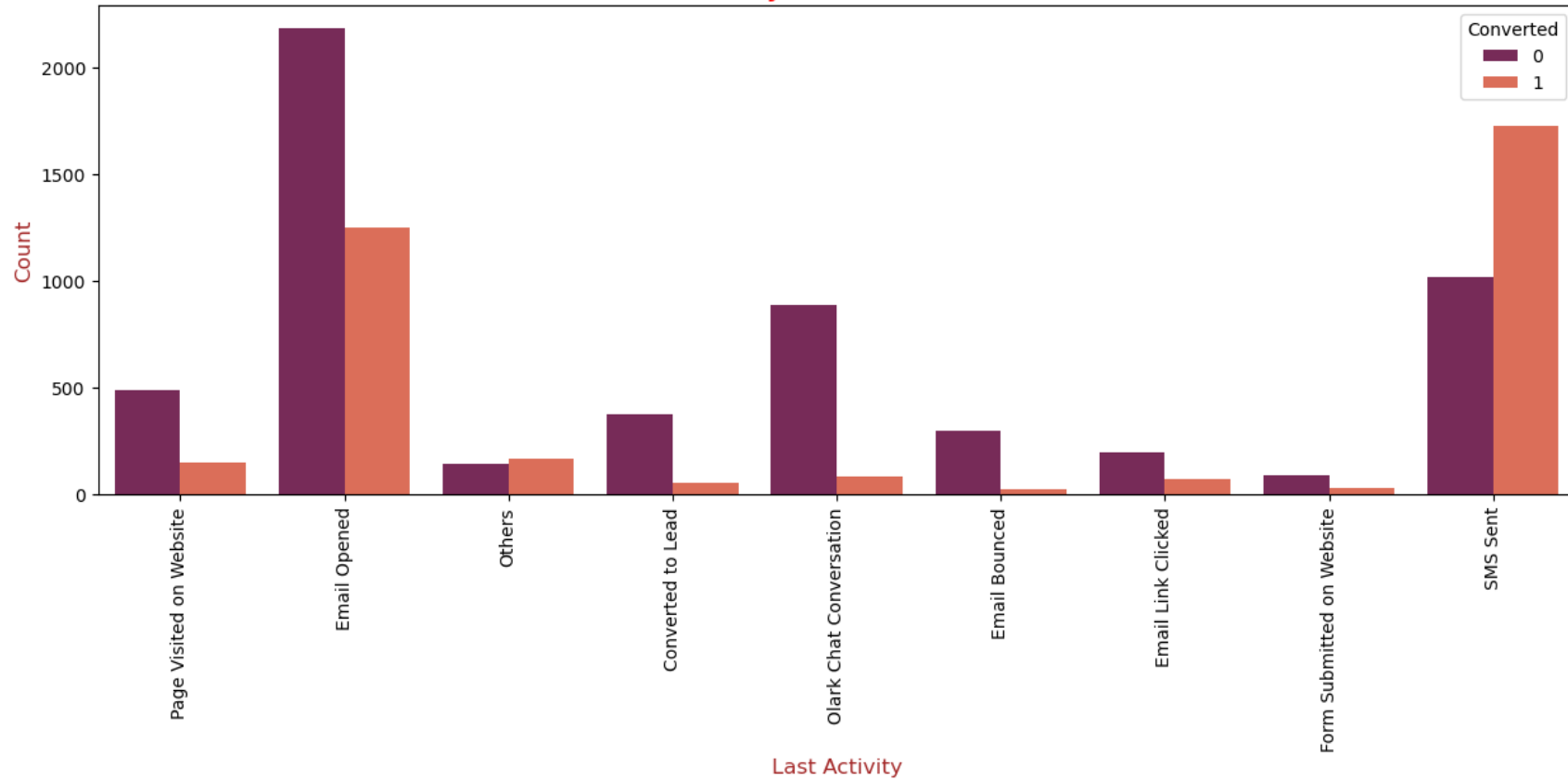


- Mumbai has the highest lead count
- Thane has the second highest lead count



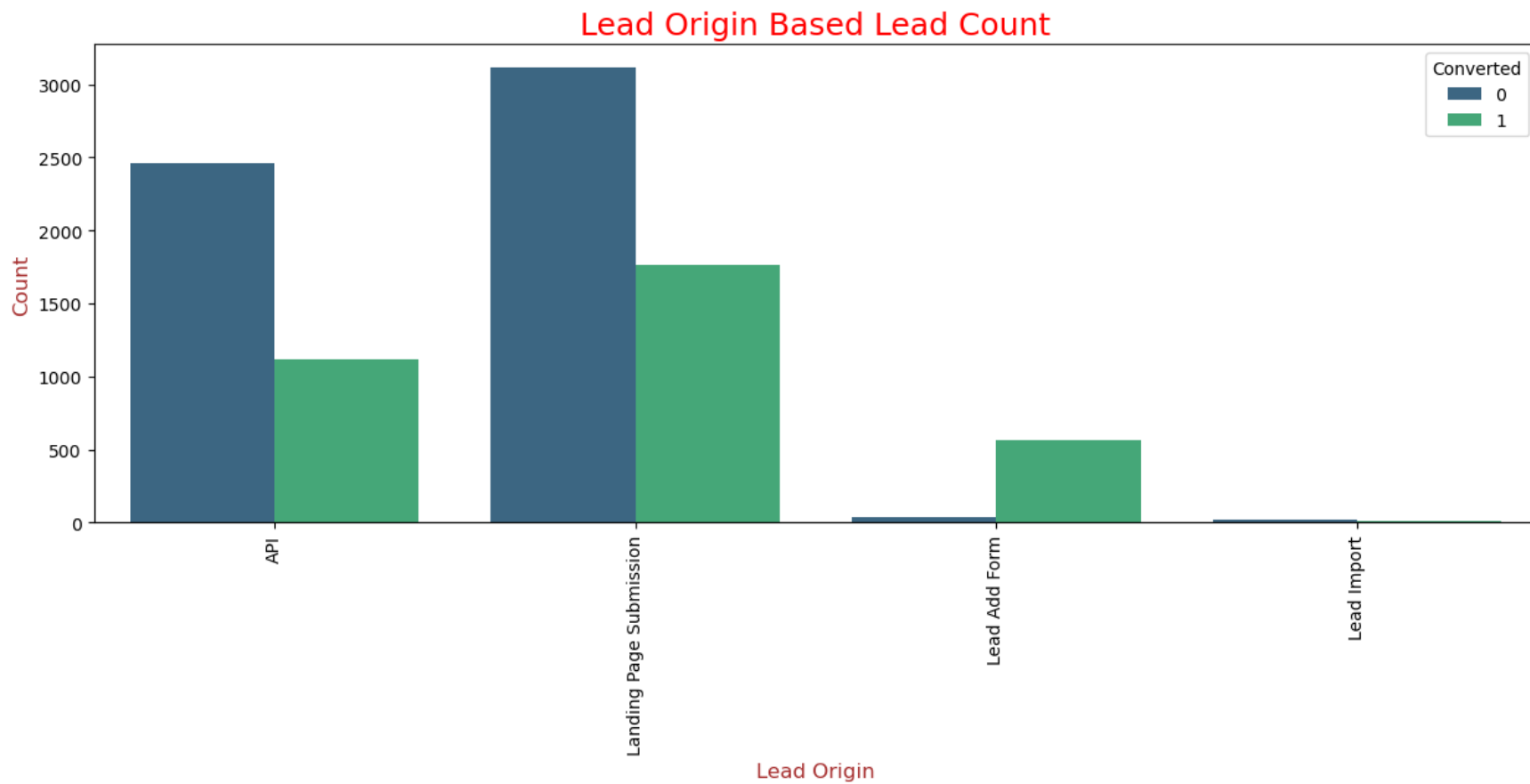
- Direct website traffic and Google generate highest lead count
- Olark Chat has the second highest lead count
- Organic Search generates third largest lead count
- Reference and Welingak has higher conversion rate. We can maximize leads from these sources
- Campaigning using the Chat system, Google and direct site may increase the number of leads

Last Activity Based Lead Count



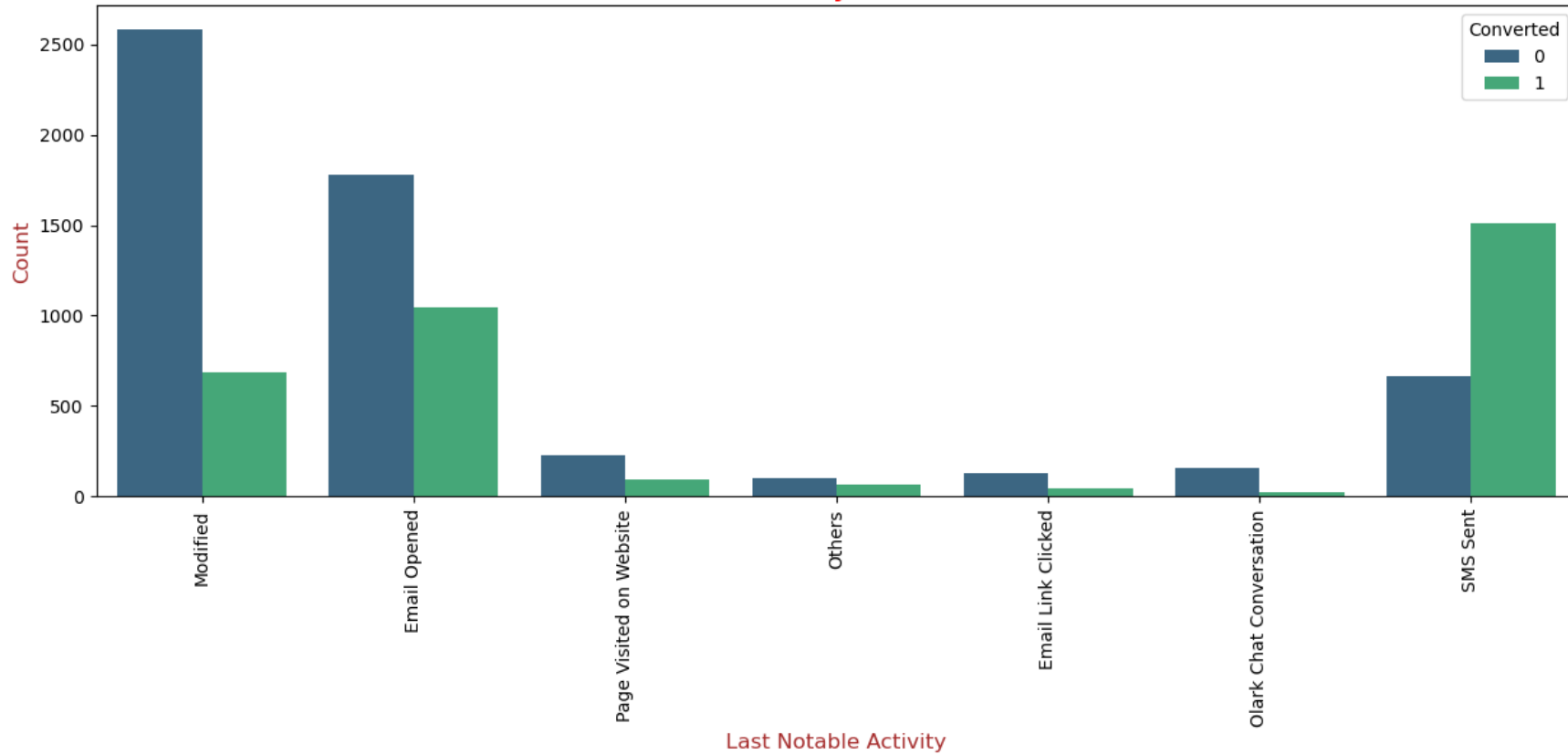
- Olark Chat and Emails generate more leads
- SMS has more lead conversion
- We can concentrate on converting Olark Chat and Email leads
- We can concentrate on generating more leads via SMS





- API and Landing page Submission generates a greater number of leads and converts a good number of leads.
- Lead add form generates less leads but has higher conversion rate
- We can concentrate on generating leads through Lead Add Form

## Last Notable Activity Based Lead Count

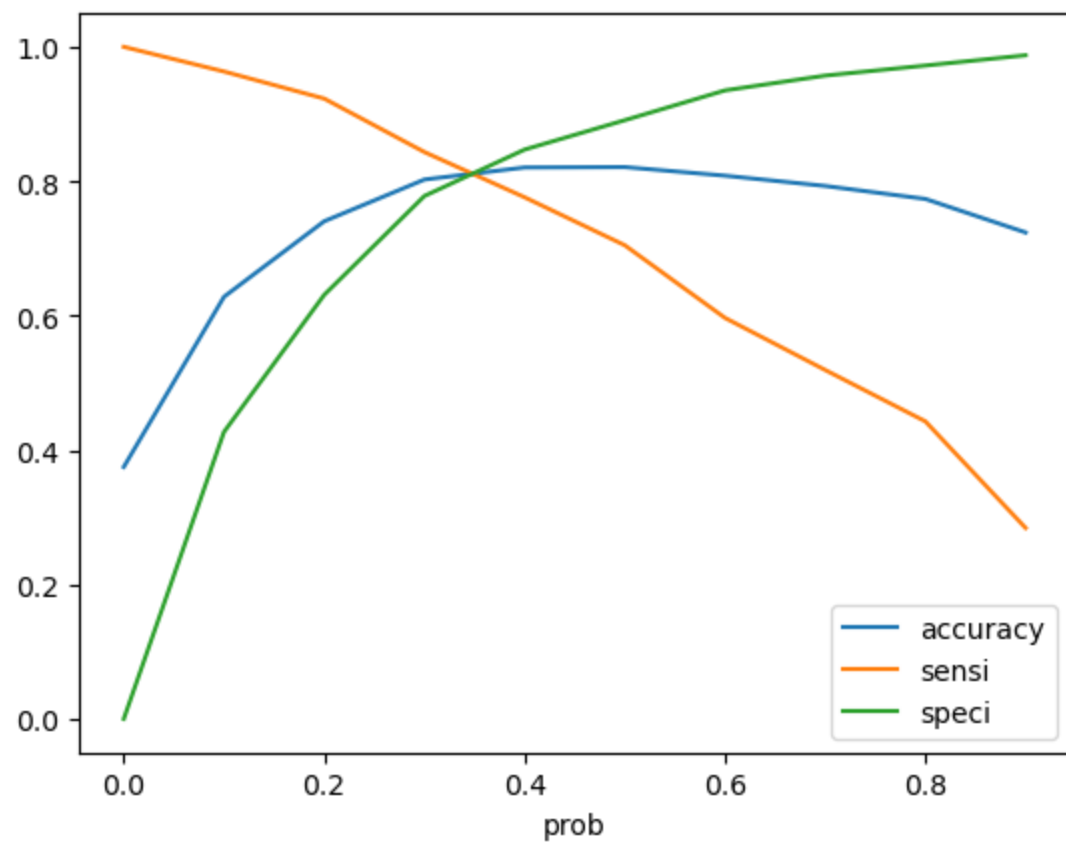


- Conversion is higher when SMS is sent
- Modified method seems to generate more leads
- We need to concentrate on generating more leads through SMS and E-mail method

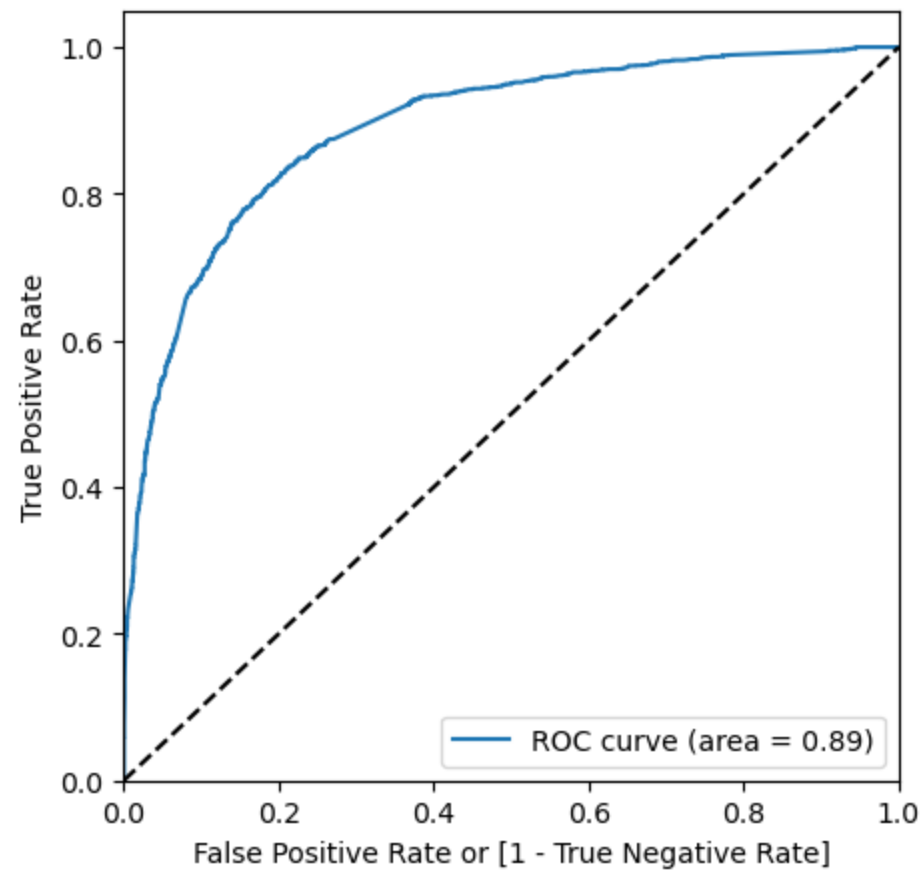
# Model Building

---

1. Splitting the Data into Training and Testing Sets
2. The train-test split is chosen at 70:30 ratio.
3. Use RFE for Feature Selection with 15 variables
4. Building Model by removing the variable using p- value and VIF
5. Predictions on test data set
6. Overall accuracy of model 80%



Optimal Cutoff point used 0.3



ROC curve area around 0.89

# Conclusion

---

It was found that the variables that mattered the most in the potential buyers are

1. The total time spend on the Website.
2. Total number of visits.
3. When the lead source was Google, Direct traffic, Organic search, Welingak website
4. When the last activity was SMS, Olark chat conversation,
5. When the lead origin is Lead add format.
6. When their current occupation is as a working professional.

Keeping these in mind the X Education can increase their buyers by focusing on the above to convert potential leads successfully.