



International
Institute of Information
Technology Bangalore



LEAD SCORING CASE STUDY

NIKUNJ PATEL

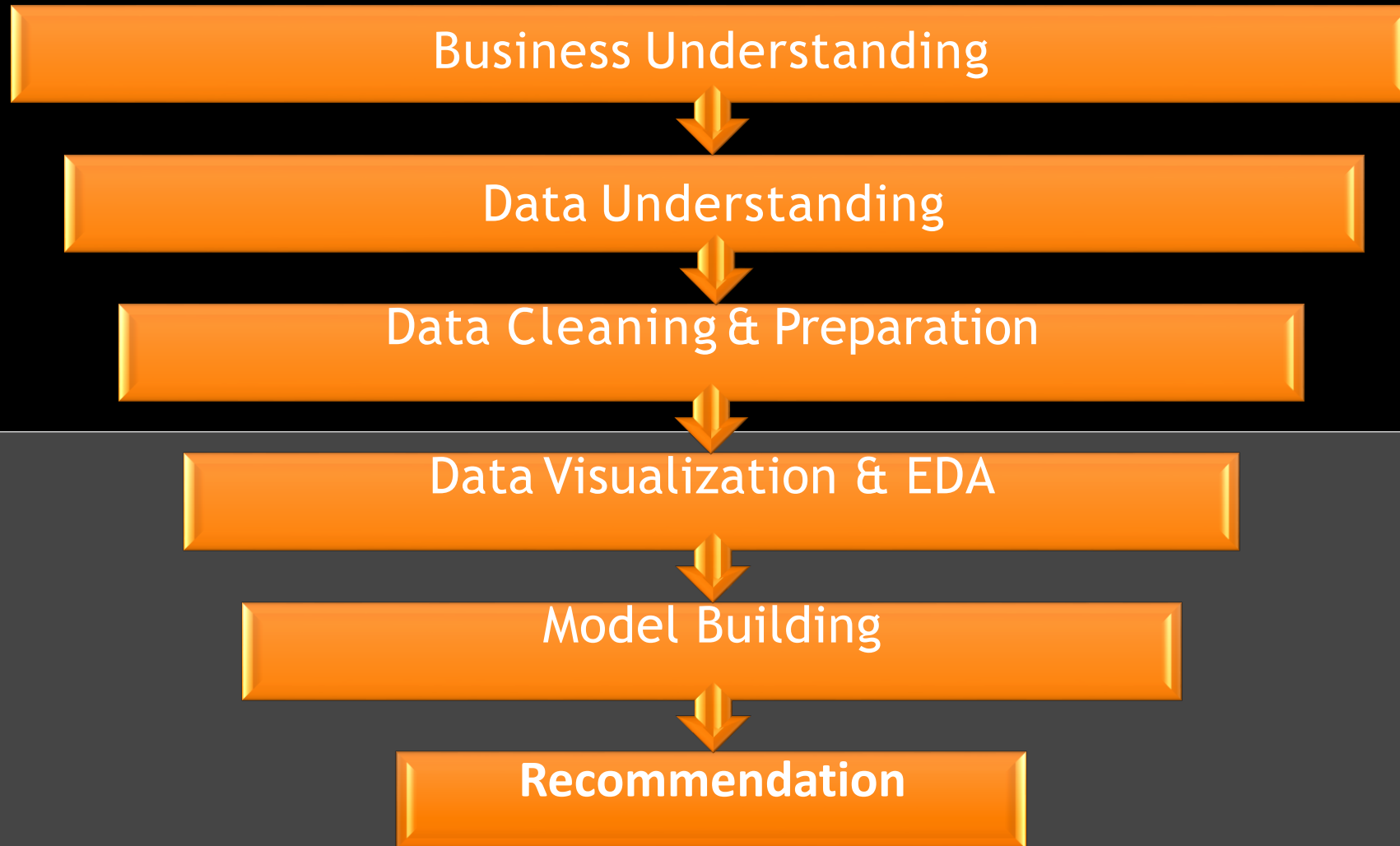
LAXMAN RAO

ABSTRACT

An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses.

Now, although 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.

PROBLEM SOLVING METHODOLOGY



BUSINESS UNDERSTANDING

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.

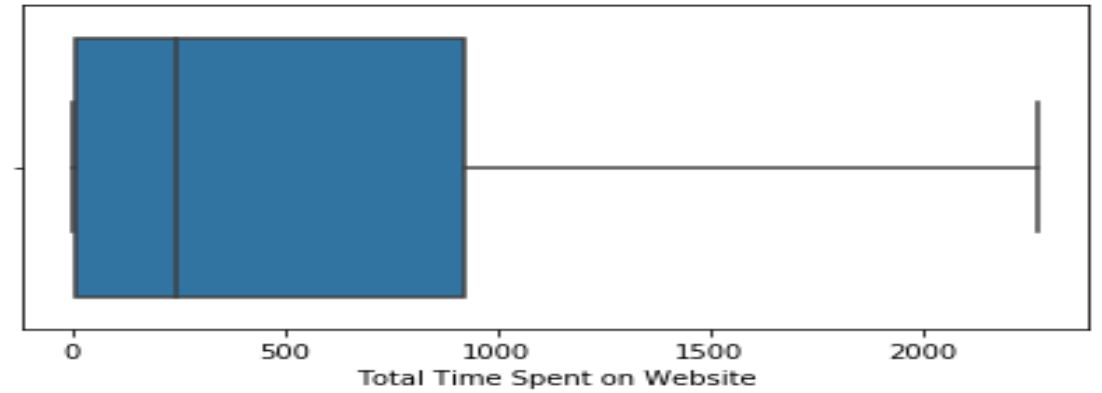
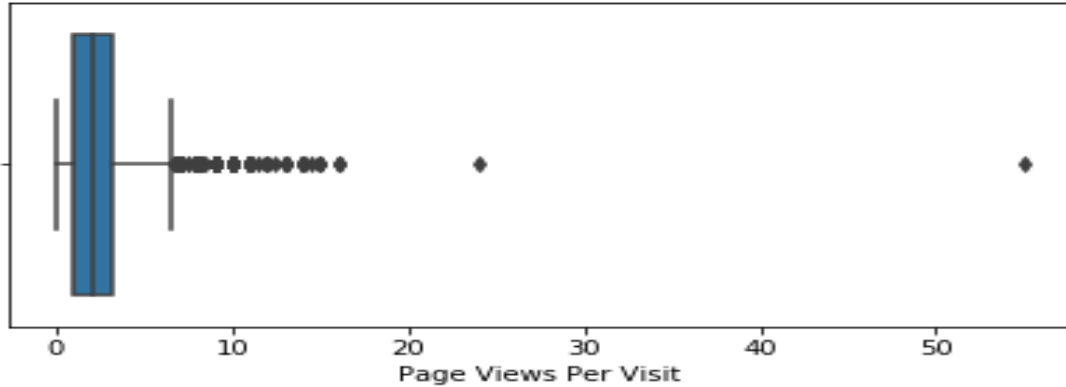
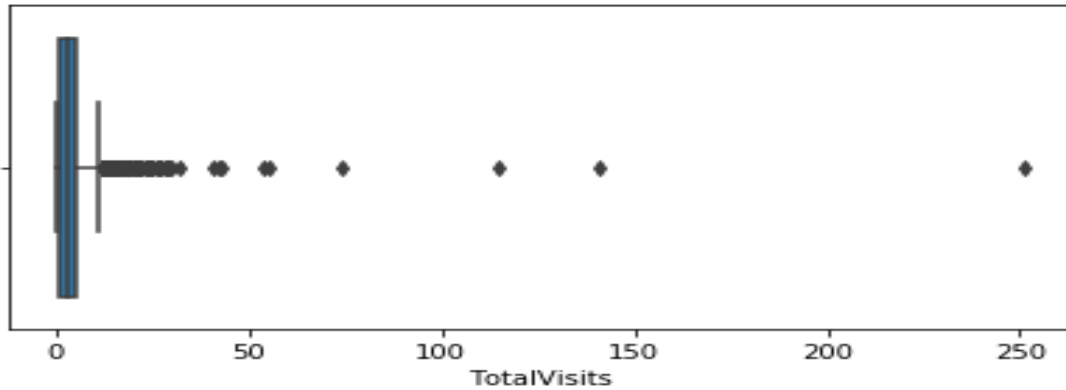
The company markets its courses on several websites and search engines like Google. Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos. When these people fill up a form providing their email address or phone number, they are classified to be a lead. Moreover, the company also gets leads through past referrals. Once these leads are acquired, employees from the sales team start making calls, writing emails, etc. Through this process, some of the leads get converted while most do not.

There are a lot of leads generated in the initial stage (top) but only a few of them come out as paying customers from the bottom. In the middle stage, you need to nurture the potential leads well in order to get a higher lead conversion.Help them to find hot leads and give education company to strategy to convert more leads and focus more some variable more.

DATA CLEANING & PREPARATION

- Identification of missing values and duplicate in data set.
- Treatment of missing values by dropping columns having missing values more than 30% from Dataset
- Replace some categorical column with Top Frequency. There is a “Select” value in some column so we Replace them with “Unknow” value and considered as we don’t know anything about that value
- Later Perform, EDA with Target variable “Converted” with all the columns one by one and
- Perform outlier Analysis and Treat outlier in “ Total Visits “ & “ Page views per visit “
- Dropping off few irrelevant columns for analysis like(Newspaper,magazine,search)
- Create Dummy Variable Based on final EDA columns

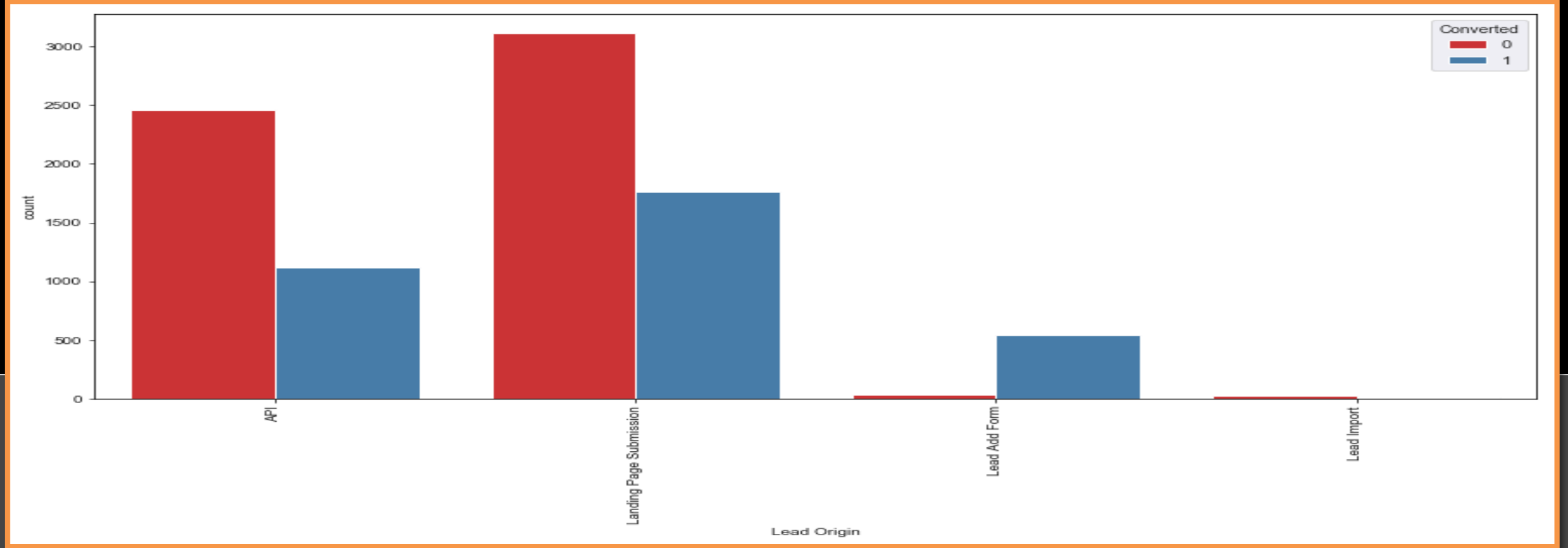
OUTLIER ANALYSIS



It can be seen that “ Total Visits ” and “ Page Views per visit ” has the outlier so we treat them using the quantile range 0.05 to 0.95

EDA (Exploratory Data Analysis)

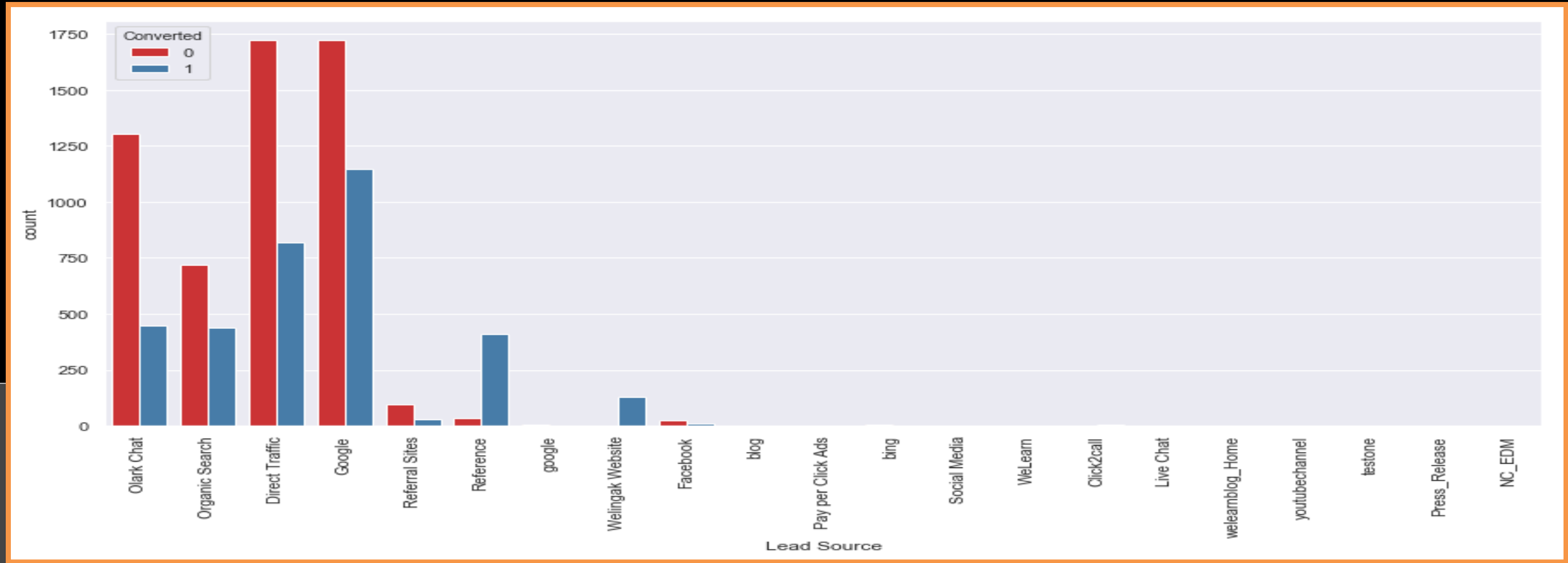
(1) Lead Origin



- As per above plot we can see most of the lead is converted from API and Landing page submission
- Also Lead Add From has the very good rate of Lead conversion but the count of lead is very less

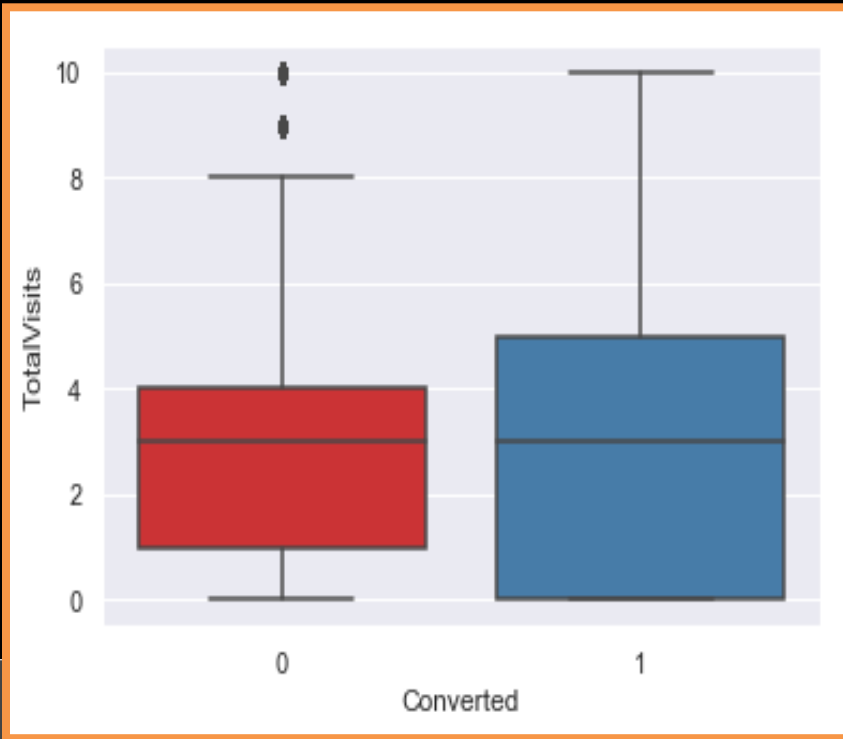
EDA (Exploratory Data Analysis)

(2) Lead Source



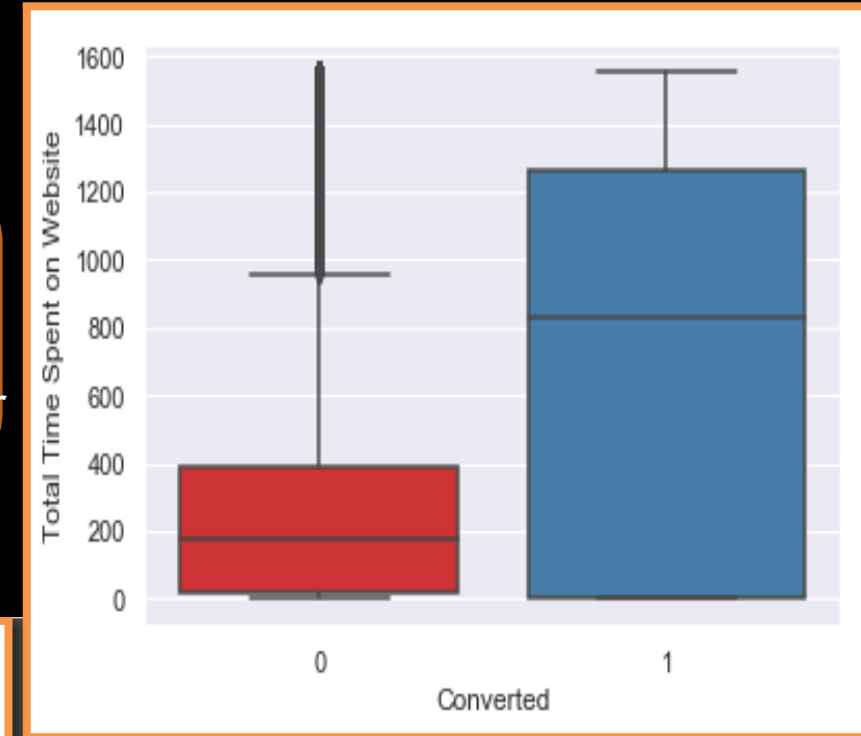
- As we see Google & Direct Traffic Genrate the number of leads
- on the other side , Reference & Welingak has the high number of chance to convert the leads

(3) Total Visit



- Median for converted and not converted leads are the same.
- Nothing conclusive can be said on the basis of Total Visits.

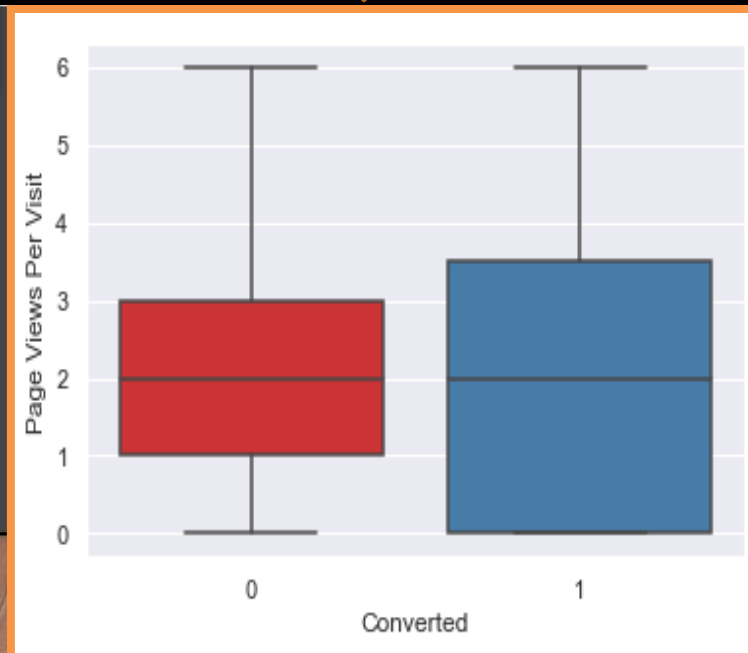
(5) Total Time spend on website



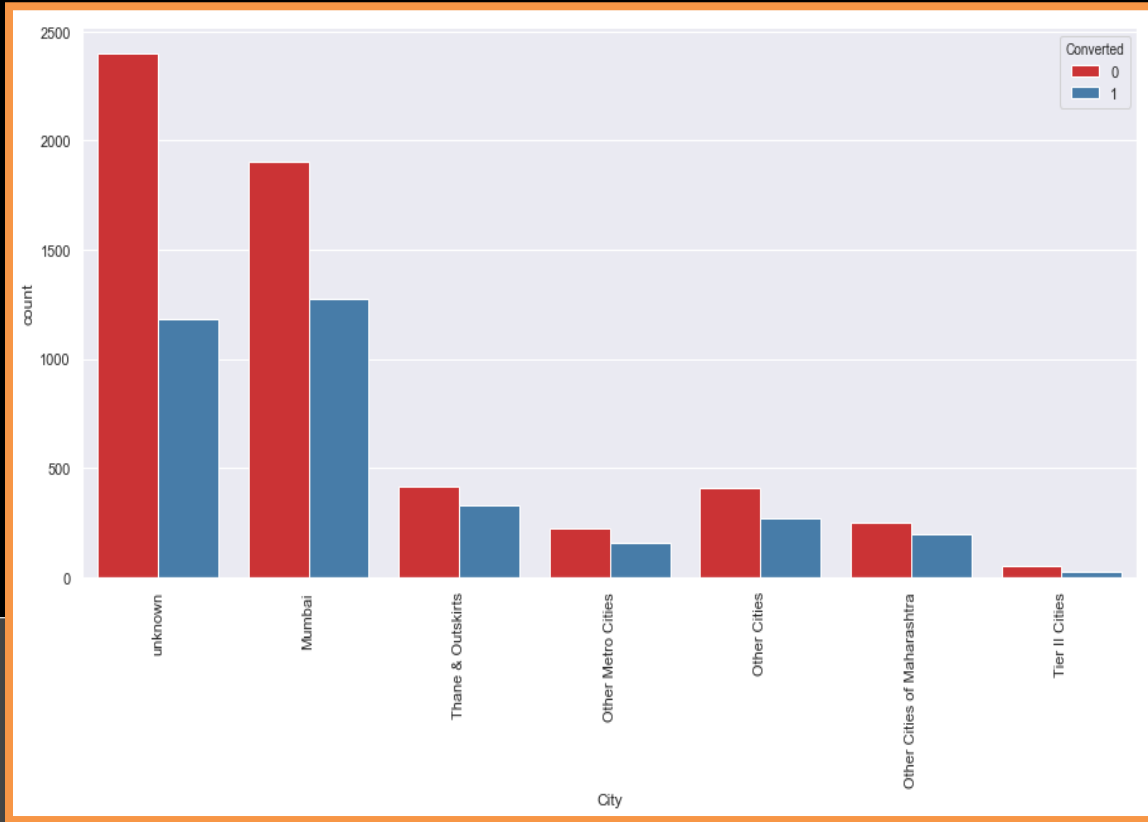
- We can say that people who spending more time on website lead will be converted

(4) Page view per visit

- Median for converted and unconverted leads is the same.
- Nothing can be said specifically for lead conversion from Page Views Per Visit



(6) City



- As we seen Most of the lead and conversion comes from mumbai

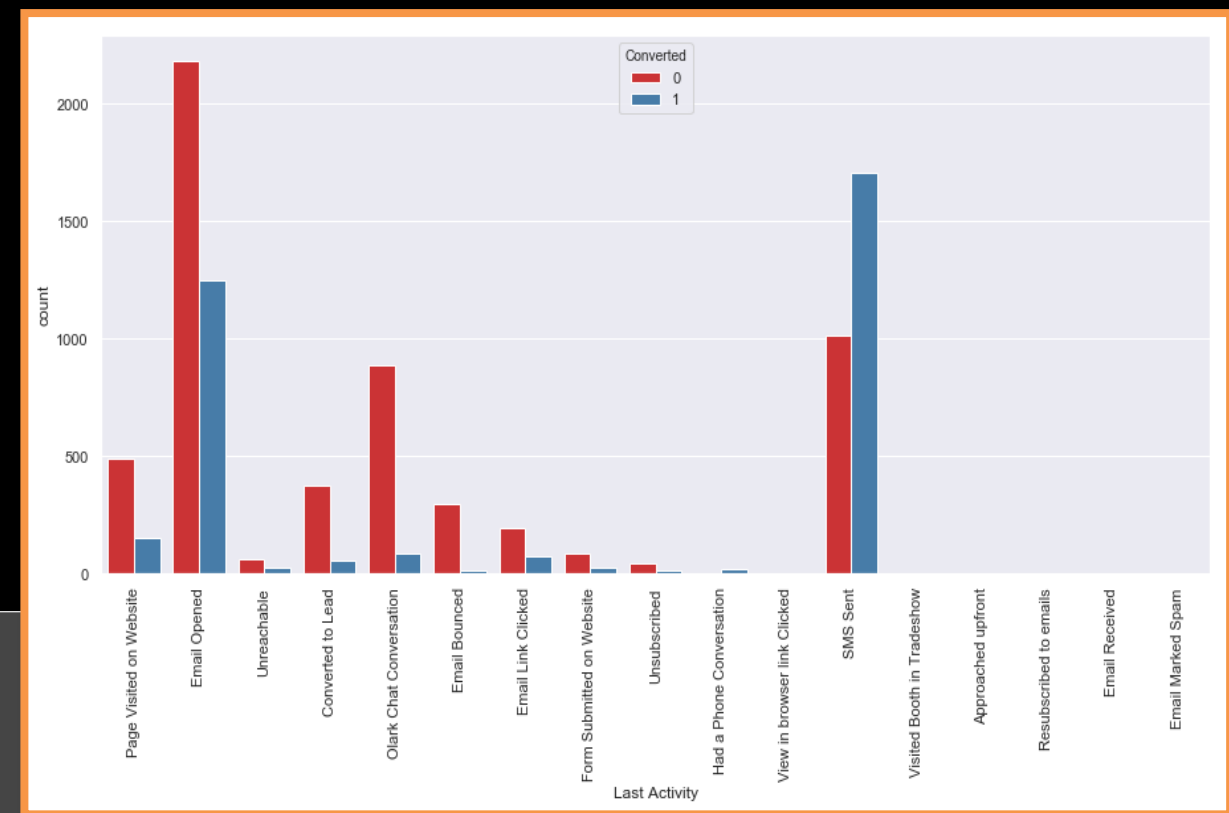
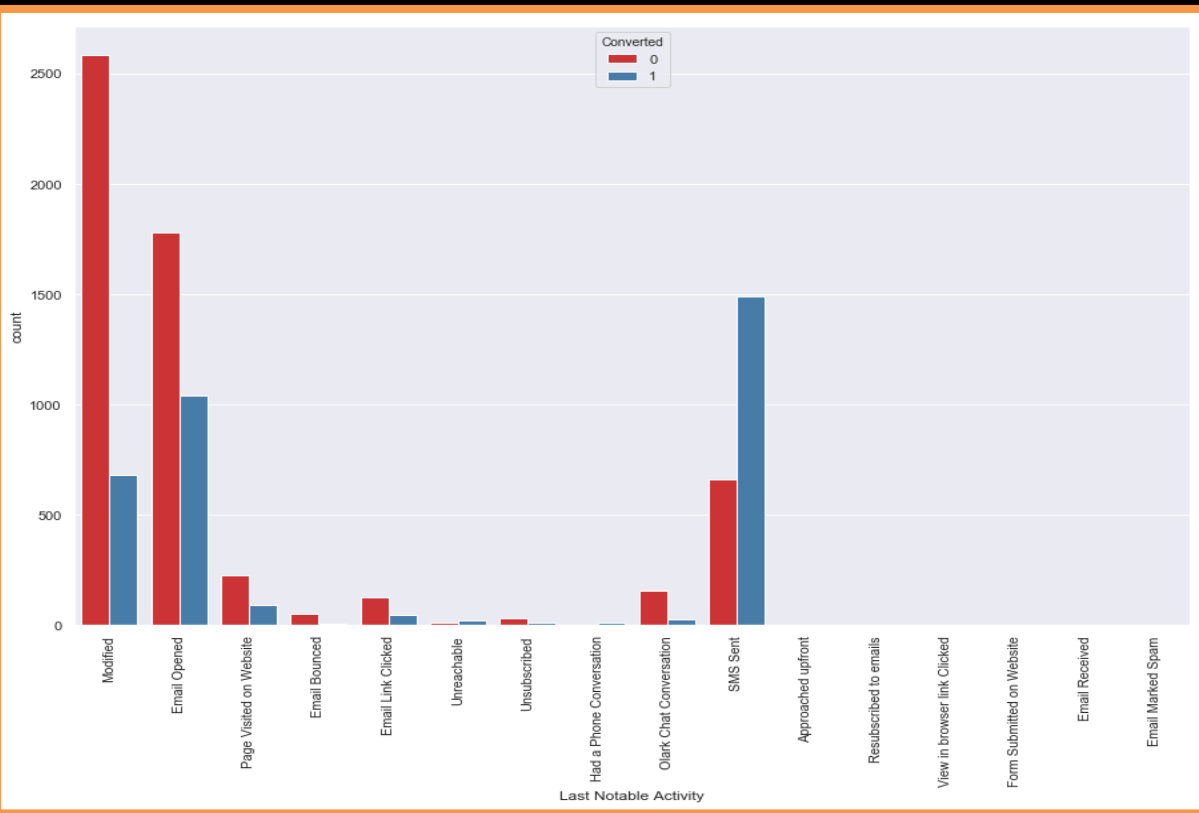
(7) Country



- As we clearly seen india has the only country to target for lead generated.

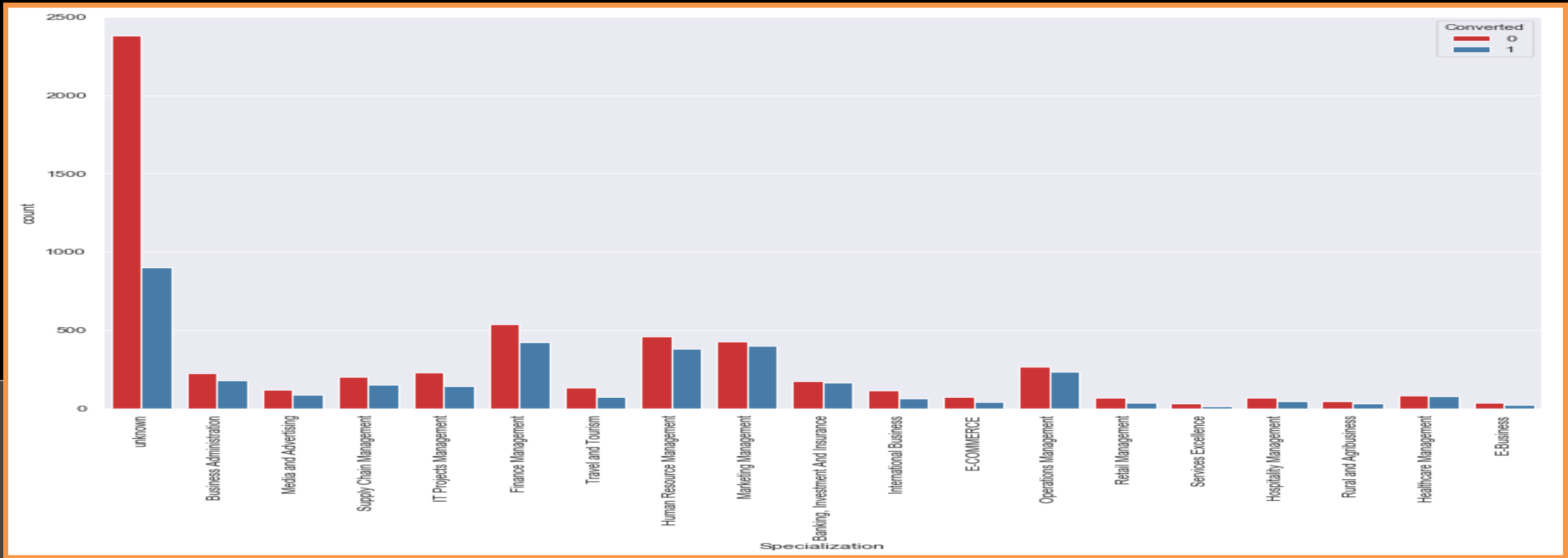
(8) Last Notable Activity

(9) Last Activity



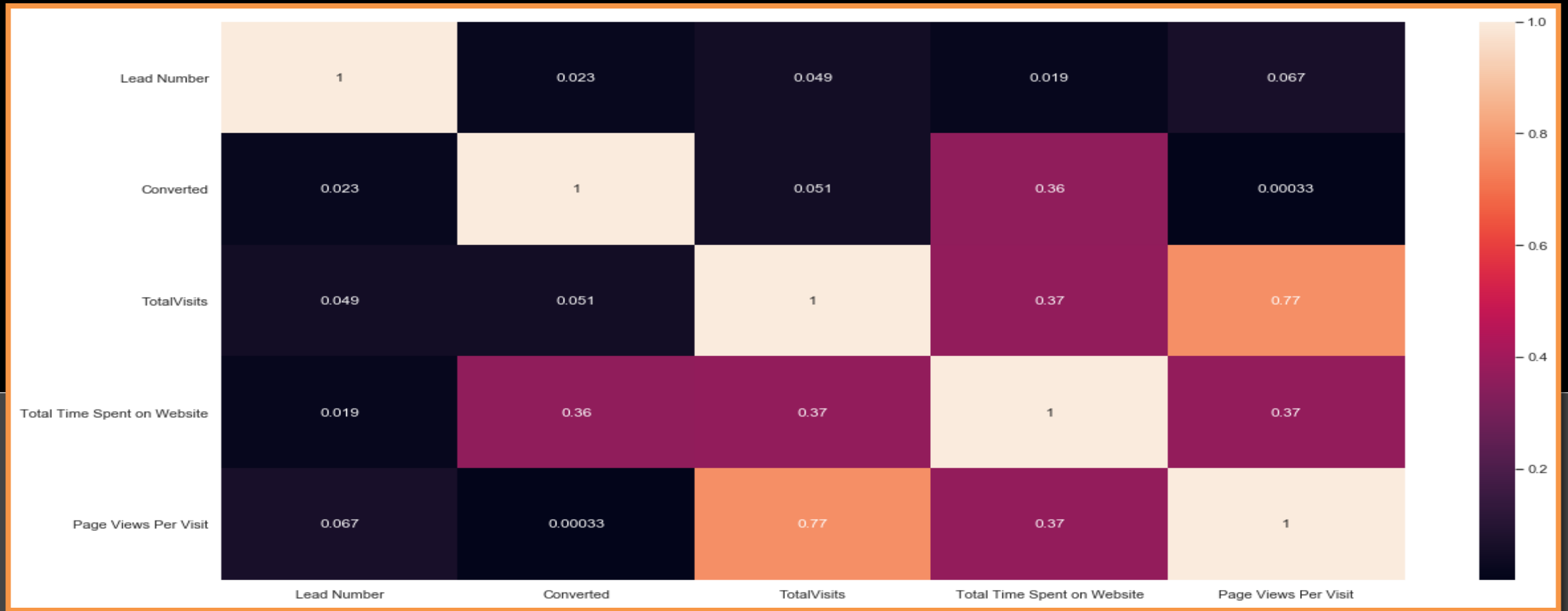
- As we seen in Both Picture Most of the lead and conversion comes from SMS sent & Email open

(10) Specialization



- As we clearly seen Finance Management , Human Resource management and Marketing management has the high **Specialization** to convert lead and conversion

Correlation between various attributes.



Correlation matrix of Various Attributes :

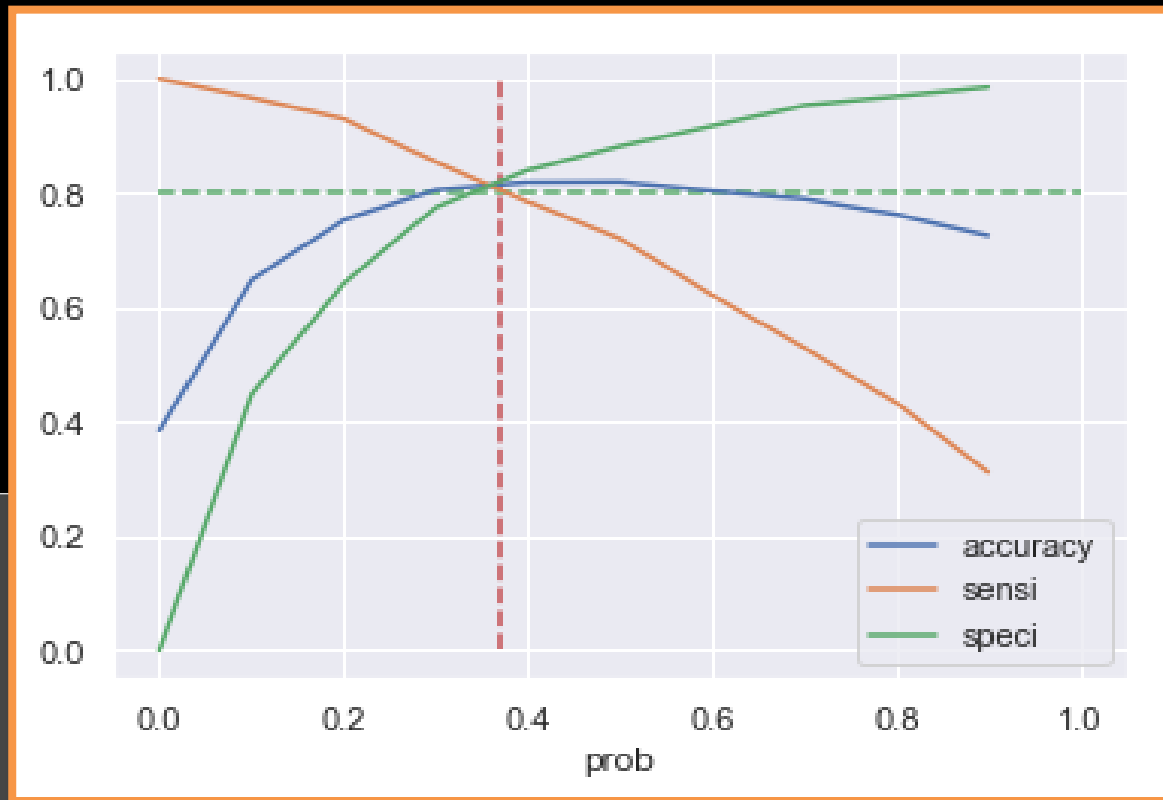
1) Total Visits 2) Total Time spent on Website 3) Page Views Per visit.

As we seen in plot there is no any attribute which is Highly Correlated with each other. So we say that Correlation Matrix is Good

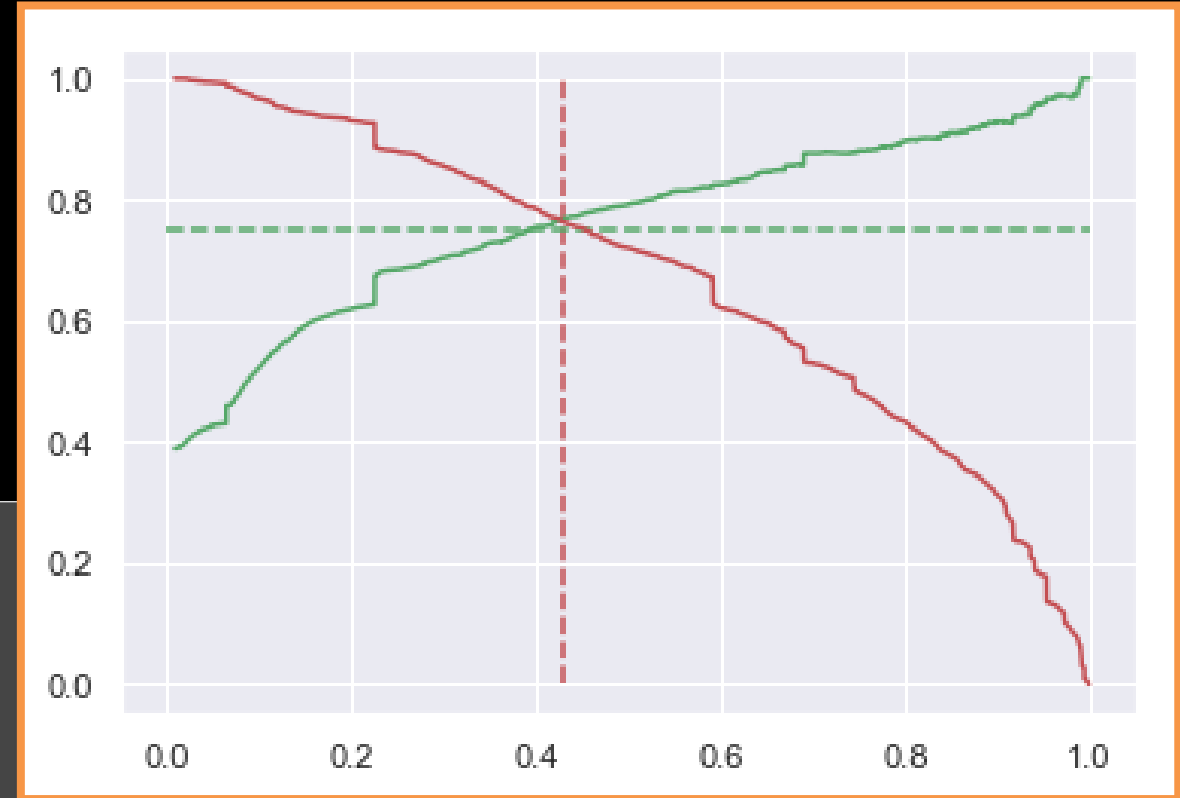
Model Building using Logistic Regression

- Step which I perform on Logistic Regression :
 - After Creating the Dummy Variable.
 - Next Step is to Split the Data in to Train and Test.
 - Model Analysis on Train DataSet
 - Feature Scaling on the Numerical Columns (total visit , Total time spend on website , Page view per visit)
 - Feature Selection Using RFE technique (It will give you the top 15 variable as output).Using that we build the model
- If the model has the high P-value then Drop that Column and Run model again until all the variable p-value is < 0.05 obtain
- Check the VIF if the vif has > 0.5 Drop that column one by one
- plot the ROC curve
- Choose the optimal Cutoff - point for accuracy sensitivity and specificity for various probabilities.
- Check Precision and Recall
- Model analysis on Test DataSet

Finding Optimal Cutoff point for sensitivity, specificity



Precision and recall tradeoff



- As we seen Sensitivity and Specificity Cutoff point is 0.37
- Where as Precision and recall Tradeoff point is 0.42

Assigning the Lead Score

- After perform the logistic Regression we assign the lead score means that the higher number of score is “ Hot Lead ” and with Low number it indicates the “ Cold Lead ”
- We seen in image there is some lead number which has high number and which has low number lead score

Assigning Lead Score

```
y_train_pred_final['Lead_Score'] = y_train_pred_final.Converted_prob.map( lambda x: round(x*100))
```

```
y_train_pred_final.head()
```

	Converted	Converted_prob	Lead Number	predicted	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	final_predicted	Lead_Score
0	0	0.149639	630949	0	1	1	0	0	0	0	0	0	0	0	0	15
1	0	0.018732	649355	0	1	0	0	0	0	0	0	0	0	0	0	2
2	0	0.225031	579735	0	1	1	1	0	0	0	0	0	0	0	0	23
3	1	0.882743	614238	1	1	1	1	1	1	1	1	1	1	0	1	88
4	1	0.690436	588625	1	1	1	1	1	1	1	1	0	0	0	1	69

As per our Understanding and Knowledge, Based on insights drawn out of the data , The following Result can be made.

1) As per model building the overall Accuracy of model is 81%

2) Top three variables in model which contribute most towards the probability of a lead getting converted?

Ans :

- 1) Lead Origin Lead Add Form
- 2) Last Activity had a Phone Conversation
- 3) Lead Source_Welingak Website

3) Top 3 categorical/dummy variables in the model which should be focused the most on in order to increase the probability of lead conversion?

Ans :

- 1) Lead Add Form
- 2) Had a Phone Conversation
- 3) Welingak Website

These following strategy might be followed by the company to have more number of leads converted Based on Model Predicted By 1 :

- 1) API & landing page submission and also generate more lead from the Lead add form
- 2) Olark chat, organic search, direct traffic, and google and generate more leads from reference and welingak website.
- 3) Website should be made more engaging to make leads spend more time. Leads spending more time on the website are more likely to be converted.
- 4) Focus on Mumbai but we also focus on other popular Cities to get more traffic on website and getting more leads.
- 5) Most of the lead have their Email opened as their last activity Conversion rate for leads.
last activity as SMS Sent is almost 60%. So we have to focus more on SMS Alert and Email open

If the Company Reach its Trget before time and company don't want to waste their time on phone calls so provide the some useful Strategy :

- 1) We can say that who have replied the SMS and E-mail Are more liketly to interested to join the Eduction Program. So in Order to Reduce the Phone calls team should be focus on Sending mail and Sms by Auto Respond. Using Auto Respond E-mail and SMS will be sent on the specific time and data so it will Save the Team time.
- 2) Also We have to Use Validation and Avoid the Dirty Data because Some people write down the wrong Email address and Phone number So we have to Verfify and Clean the Data.
- 3) We Also Use the Social Media platform to Engage more client and using Referal program should be advertising through converted leads.

Recommendation For Improving Lead Conversion Rate

Important Variables

- API & Landing page Submission
- Lead Add Form
- Google , Olark chat, Direct Traffic
- Total Time spend on Website
- City
- E-mail Opend
- SMS Alert
- Lead Source Welingak Website
- Had a Phone Conversion

- ✓ As we seen in Important Variable Which Should we have to focus on Converting the more leads and Traffic on Website
- ✓ So to improve the overall performance we have to focus on API & landing page submission and also genrate more lead from the Lead add form
- ✓ To improve overall lead conversion rate, we should Also focus on olark chat, organic search, direct traffic, and google and generate more leads from reference and welingak website.
- ✓ We can say that people who spending more time on website lead will be converted
- ✓ We also focus on Not only on Mumbai City but also focus on some other poplular city