LEAD SCORE
CASE STUDY

STUDY GROUP
ABHINAV SINGH
RITIKA ARYA

BUSINESS OBJEC TIVE

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. THE COMPANY GETS LEADS THROUGH PAST REFERRALS, ONLINE SOURCES AND MARKETING TEAM. IT IS FOUND THAT CONVERSION RATE AT X EDUCATION IS AROUND 30%.
THE COMPANY REQUIRES YOU TO BUILD A MODEL WHERE LEAD CONVERSION RATE COULD BE INCREASED TO AROUND 80%.

STEPS TAKEN

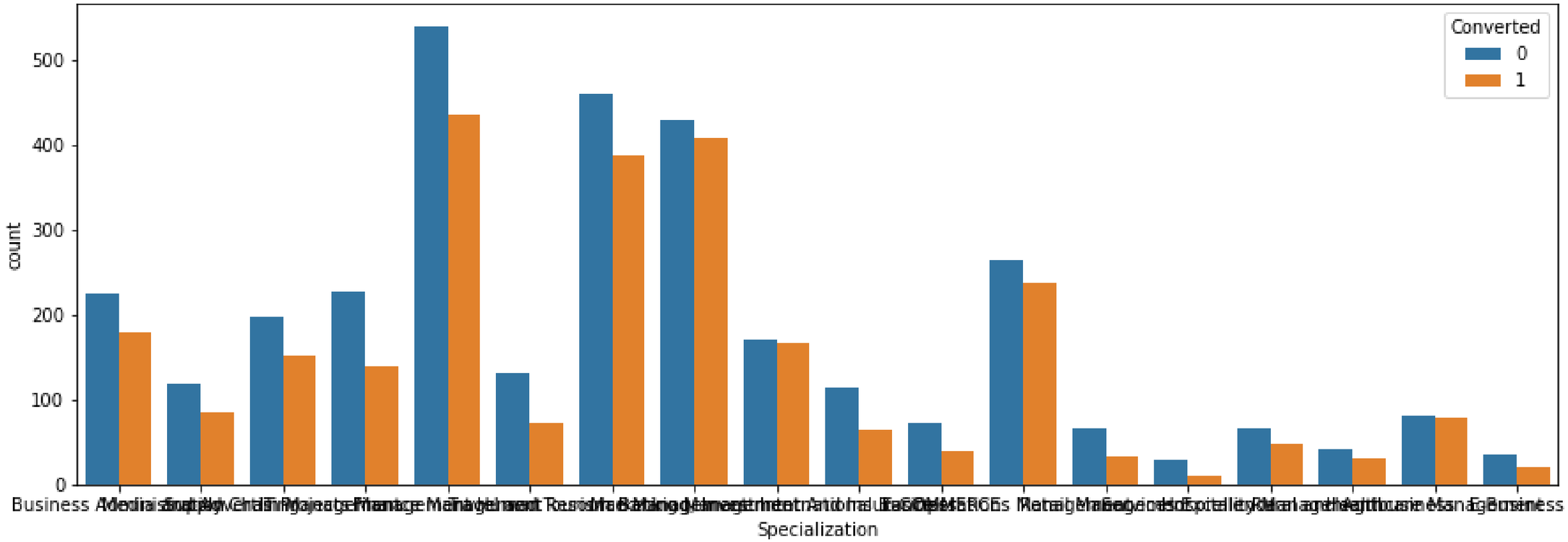
1. Data Understanding
2. Clean the data
3. Univariate Analysis
4. Bivariate Analysis
5. Prepare the data for Model Building
6. Model Building
7. Model Evaluation

DATA MANIPULATION

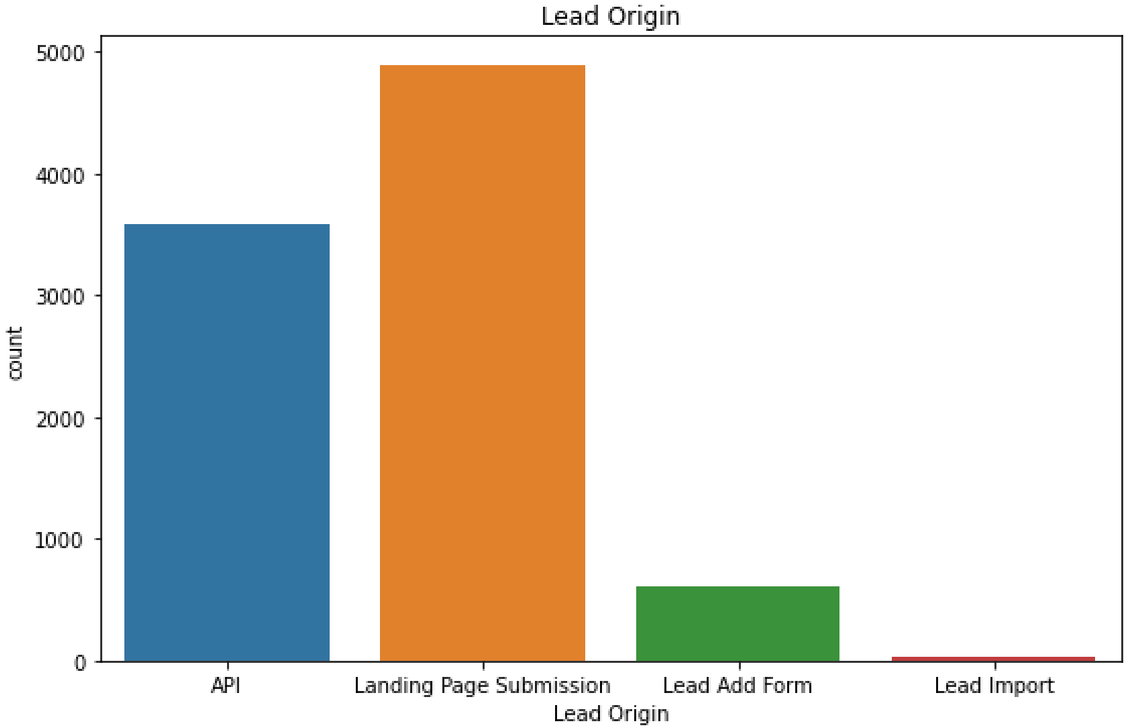
1. Total number od rows-9240 and columns-37
2. Removed the "prospect and lead number because it will not impact the model
3. Check for the null values and replacing the null values
4. Droping the Columns having more than or equal to 45% of missing values
5. EDA

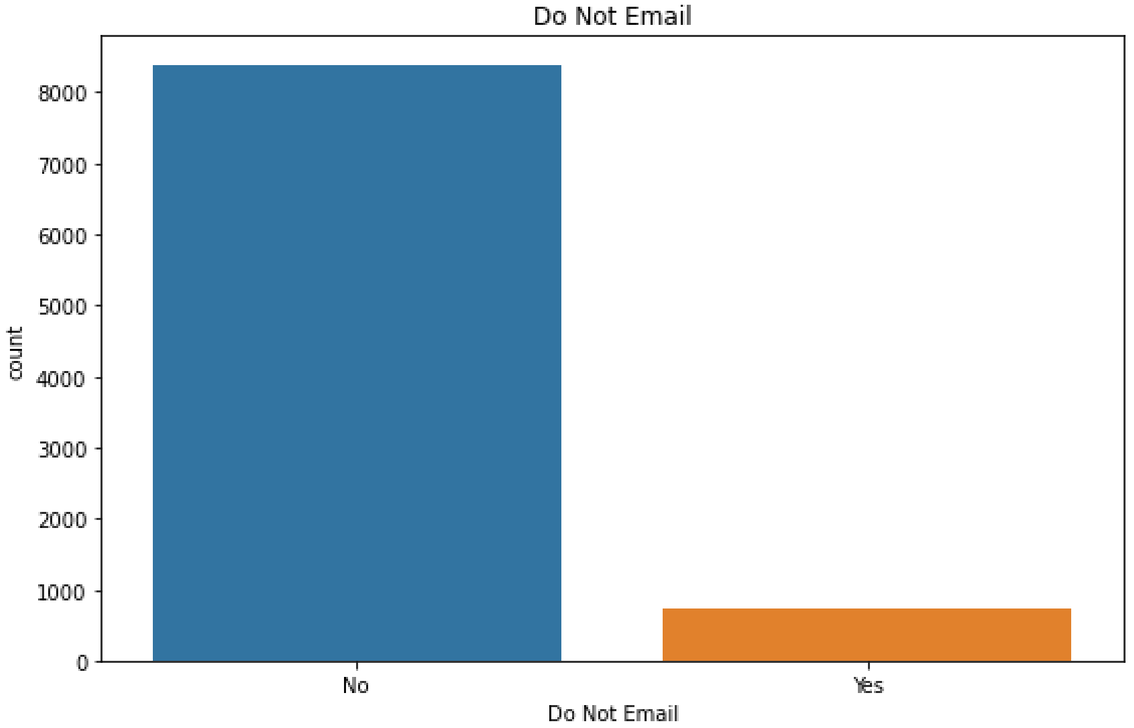
EDA

PLOT SPREAD OF SPECIALIZATION COLUMNN

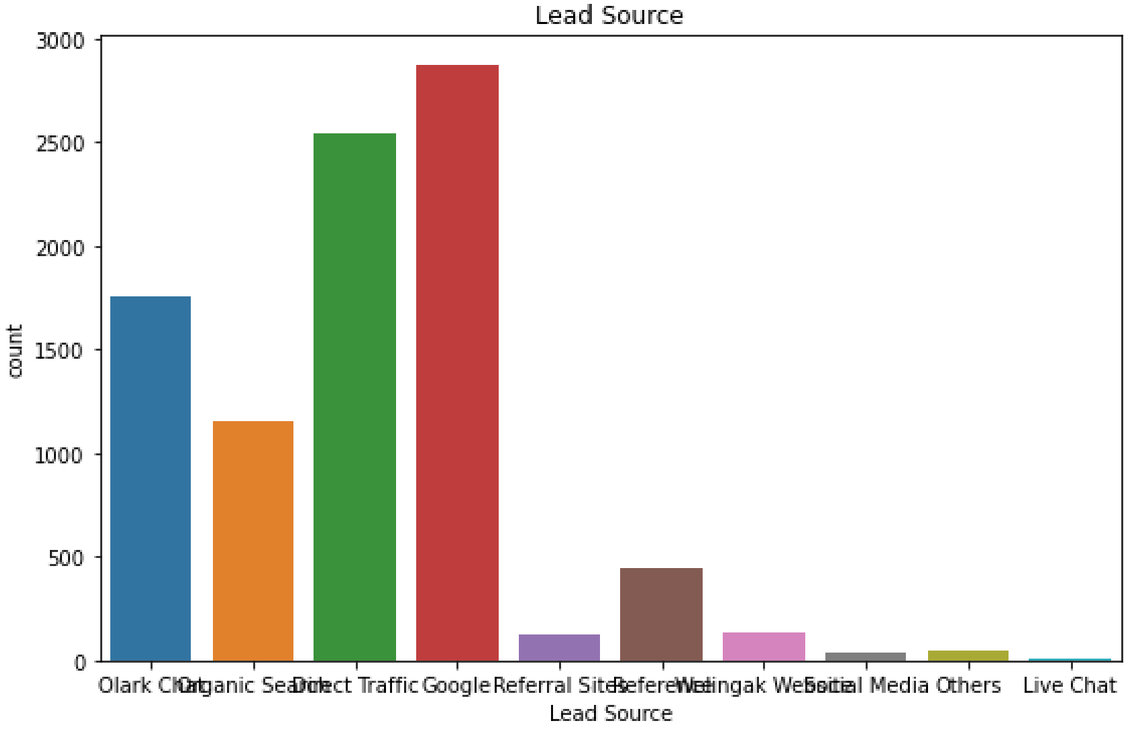


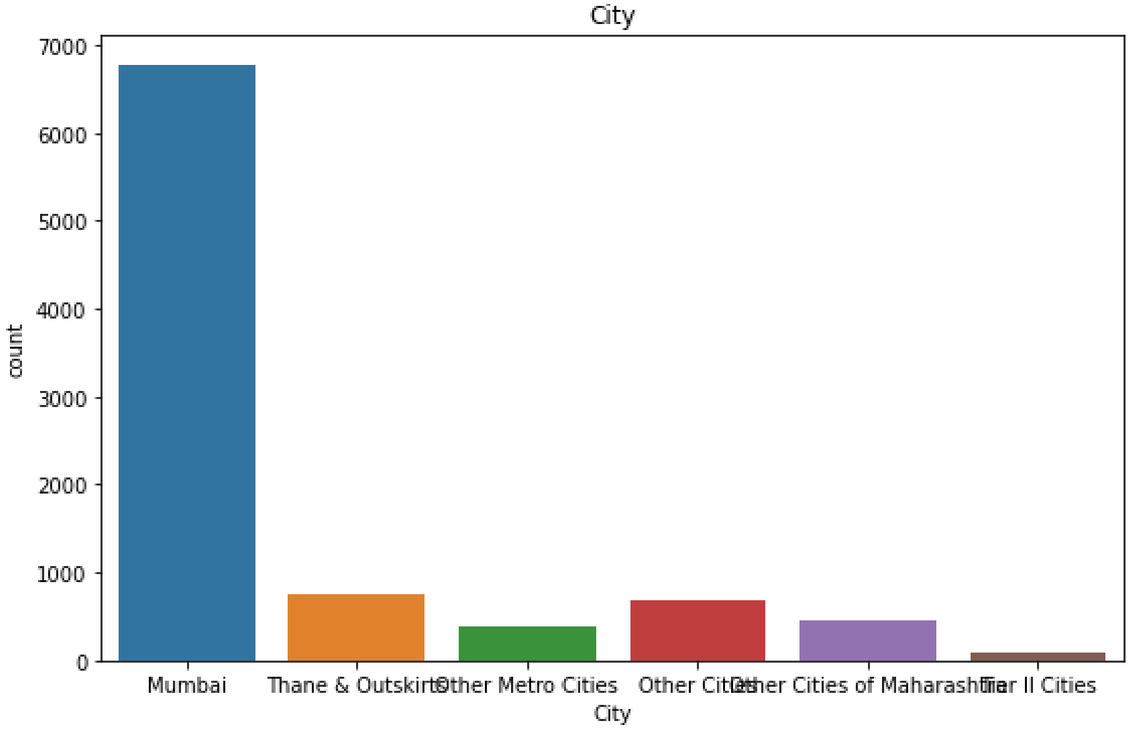
EDA


Lead Origin
5000
4000
3000
2000
1000
0
count
API
Landing Page Submission
Lead Add Form
Lead Import
Lead Origin

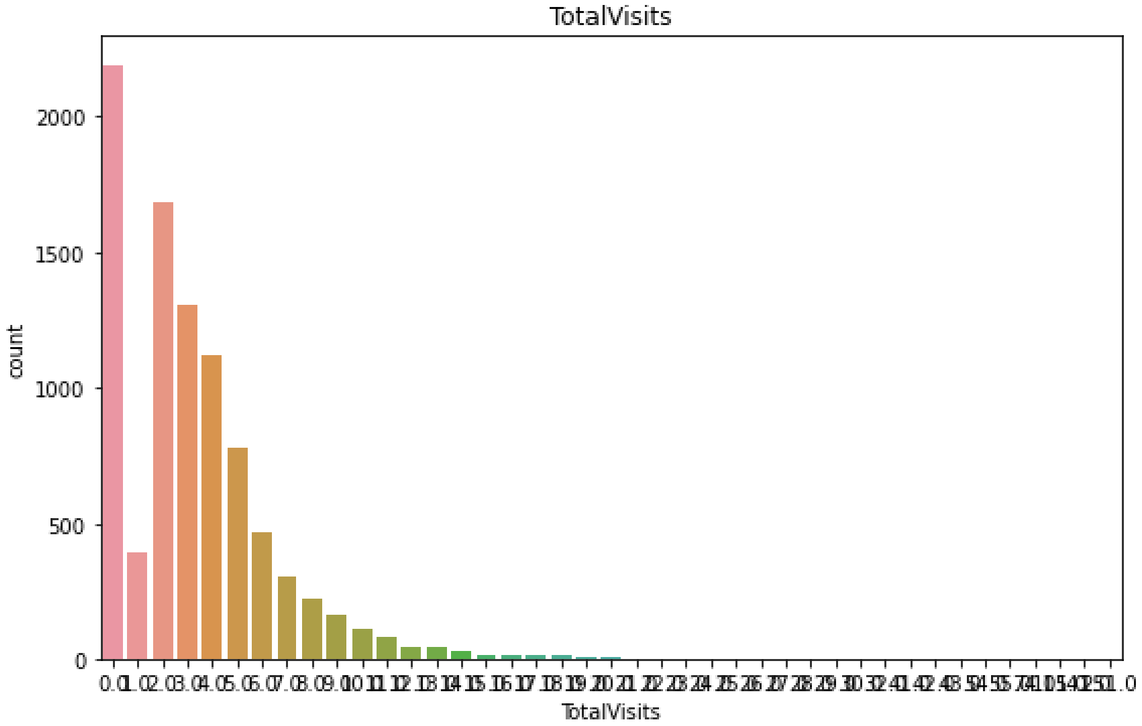

Do Not Email
8000
7000
6000
5000
4000
3000
2000
1000
0
count
No
Yes
Do Not Email

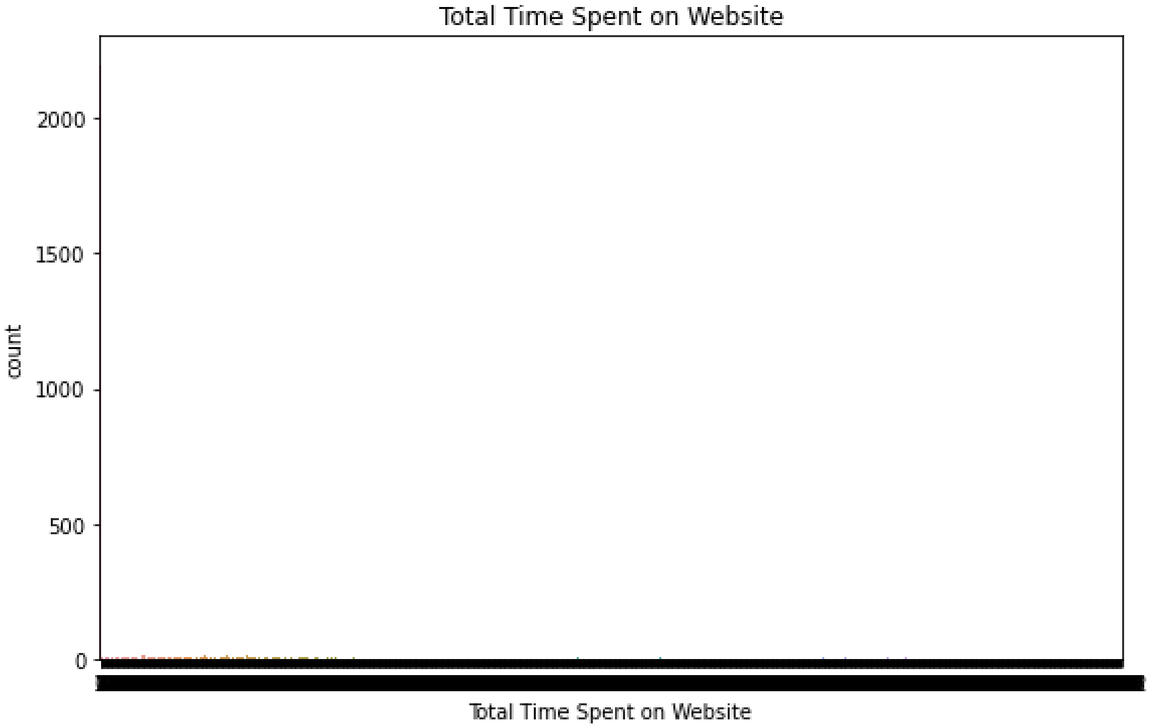
EDA


Lead Source
count
3000
2500
2000
1500
1000
500
0
Olark Chat
Organic Search
Direct Traffic
Google
Referral Sites
Reference
Welingak Website
Social Media
Others
Live Chat
Lead Source

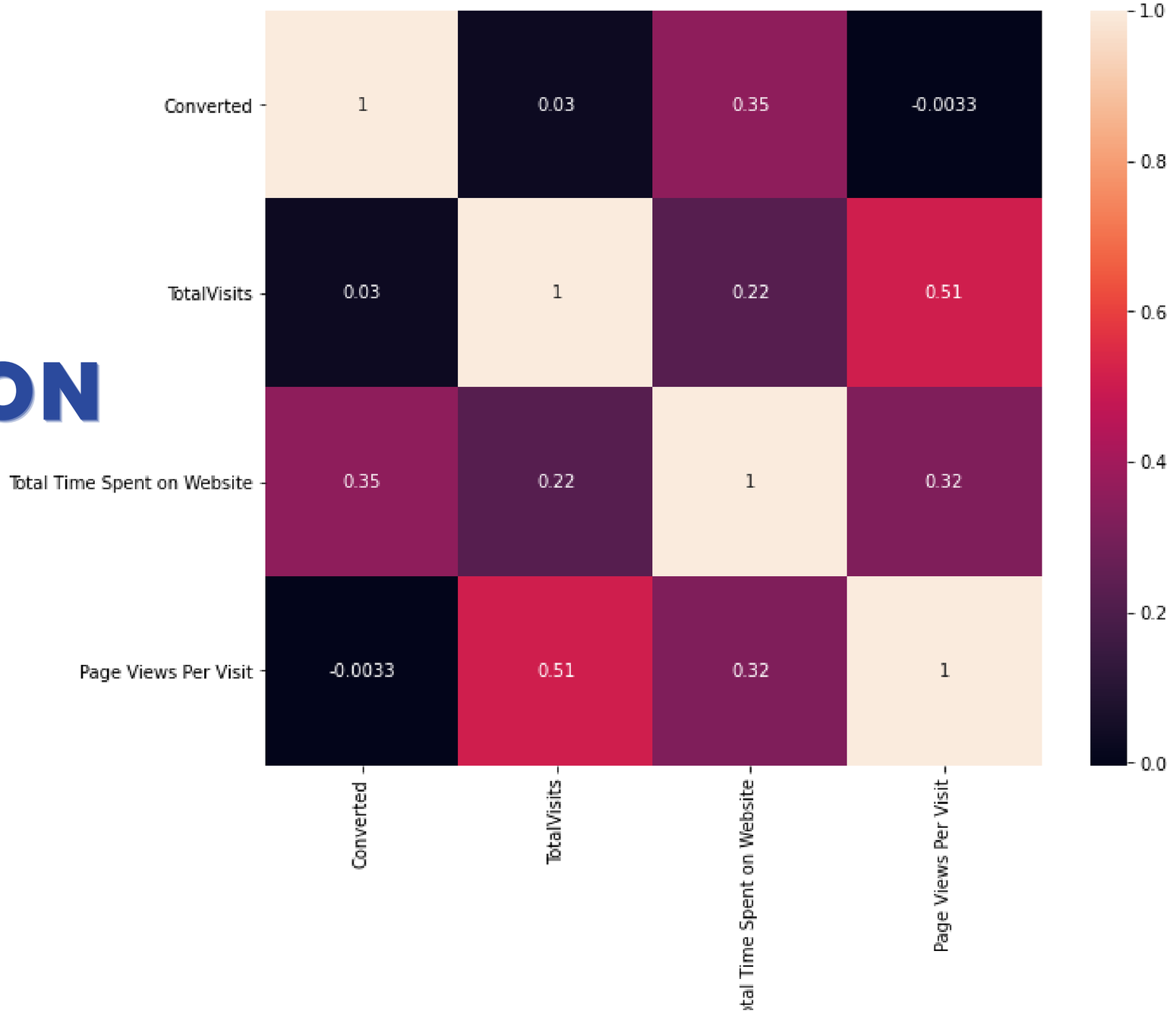

City
count
7000
6000
5000
4000
3000
2000
1000
0
Mumbai
Thane & Outskirts
Other Metro Cities
Other Cities
Other Cities of Maharashtra
Tier II Cities
City

EDA


TotalVisits
count
2000
1500
1000
500
0
TotalVisits


Total Time Spent on Website
count
2000
1500
1000
500
0
Total Time Spent on Website

CORRELATION



CREATION OF DUMMY VARIABLES

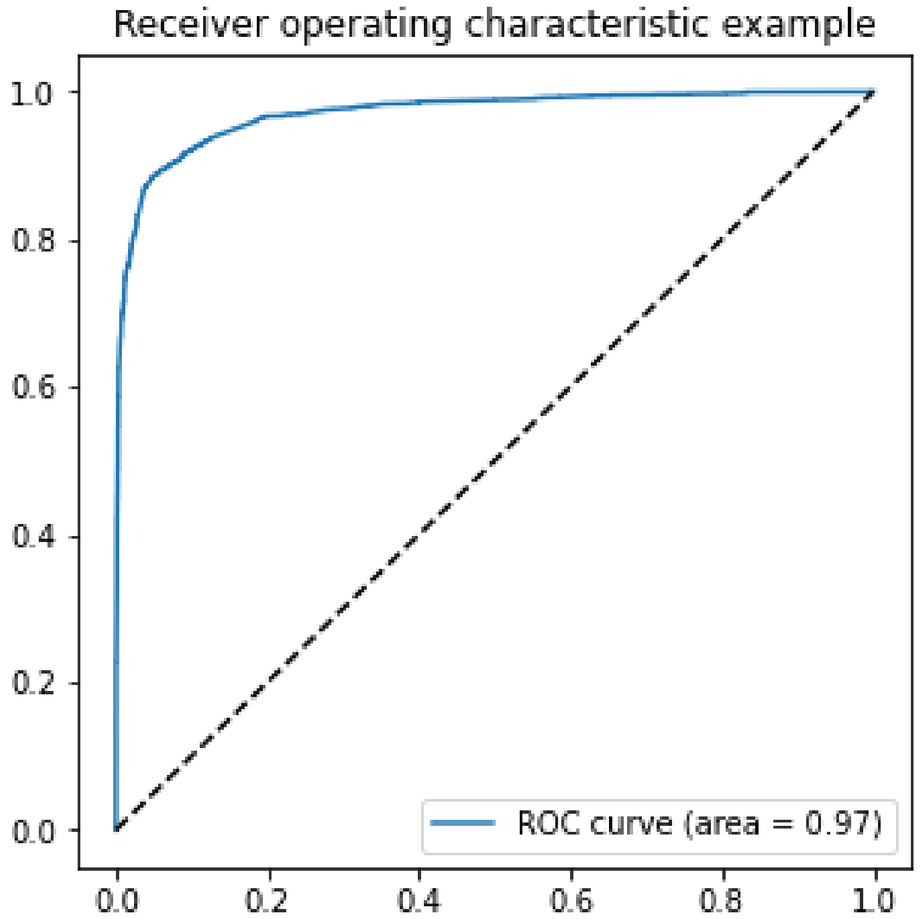
1. Total number of rows-9240 and columns-51
2. After Creating the dummy variables drop the original columns
3. following are the dummy variables

- Lead Origin','What is your current occupation
- Specialization
- Lead Source
- Last Activity
- Tags

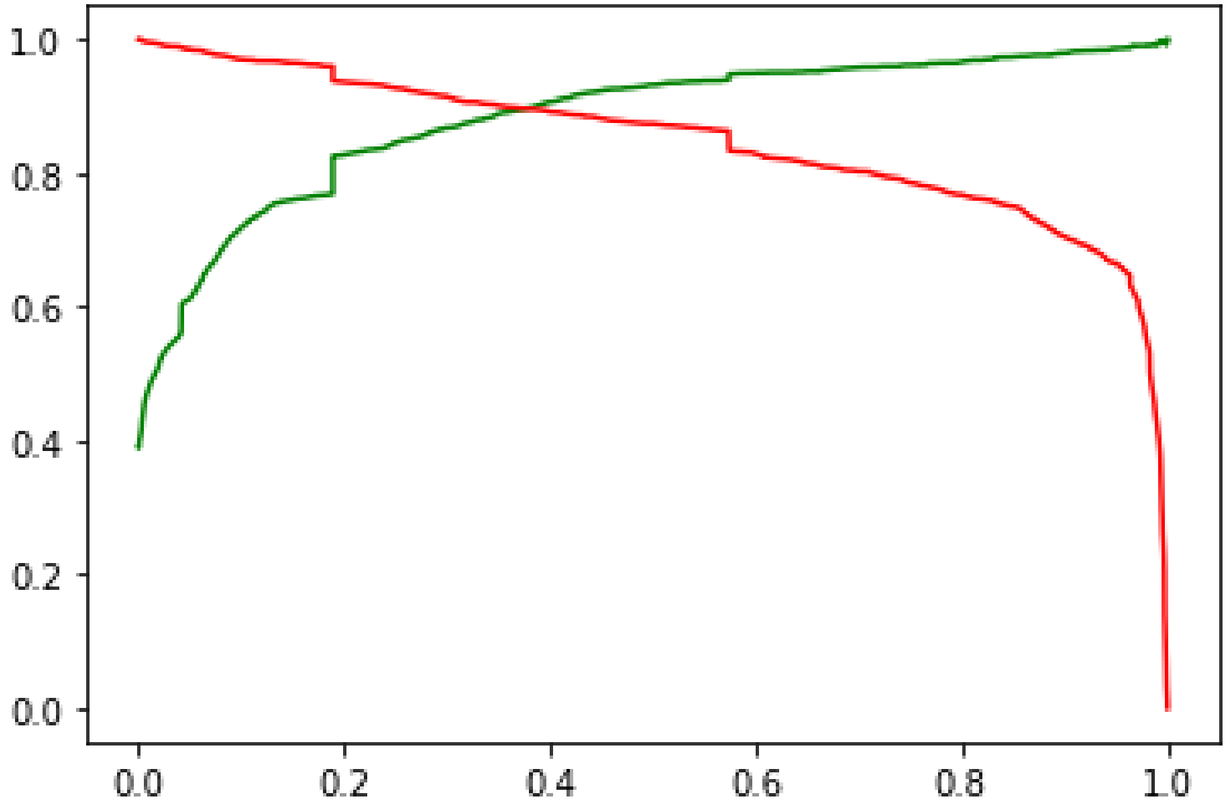
MODEL BUILDING

1. Import the required library
2. Split the dataset into 70% and 30% for train and test respectively
3. scaling
4. Use rfe for feature Selection
5. Put all the columns selected by RFE in the variable 'col'
6. Building Model According to the p-value
7. Accuracy 92.3%

ROC CURVE


Receiver operating characteristic example
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
ROC curve (area = 0.97)

ROC CURVE



CONCLUSION

Factors that are impact full

- **The total time spend on the Website.**
- **Total number of visits.**
- **Lead Sources with element Google**

Maximum number of leads are generated by Google and Direct traffic.

Conversion Rate of reference leads and leads through welingak website is high.

SOLUTIONS

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.