LEAD SCORING CASE STUDY

By :

Snehal Kedar

Siva Sankare

Smith Desai

PROBLEM STATEMENT

X Education company sells online courses, they are getting lot of leads but the lead conversion rate is very poor. To make this process more efficient, the company wishes to identify the most potential leads so that they can focus more on communicating with the potential leads rather than making calls to everyone.

The company requires to build a model, wherein you need to assign a lead score to each of the leads such that the customers with a higher lead score have a higher conversion chance and the customers with a lower lead score have a lower conversion chance. Target lead conversion rate to be around 80%.

ANALYSIS APPROACH

➢ Understanding The Business Objective/Problem Statement

➢ Data Analysis (understanding data)

➢ Data Preparation – Handling Missing Values, Outliers, Dropping/Imputing columns

➢ EDA/Visualization

➢ Dummy Variable Creation For Categorical Columns

➢ Feature Scaling (Manual & RFE to select 15 features in output)

➢ Run The Logistics Regression Model On Training And Testing data (70:30 Split, P-value   should less than 0.05 and VIF less than 5) & Prediction.

➢ Model Evaluation

➢ Conclusion And Recommendation

FEATURE SELECTION

- In data there are many columns has missing values. This incomplete data can lead to bias/wrong result hence it is necessary to check and treat missing values. Below are the details.

In Scope For Model Building
Lead Origin
Lead Source
Do Not Email
Last Activity
Specialization
Last Notable Activity
What is your current occupation
TotalVisits
Total Time Spent on Website
Page Views Per Visit
Converted

Dropped Columns Due To Missing Values =>40%.
How did you hear about X Education
Lead Quality
City
Asymmetrique Activity Index
Asymmetrique Profile Index
Asymmetrique Activity Score
Asymmetrique Profile Score
Tags
Lead Profile

| Dropped Columns As Not Giving Much Information |
| --- |
| Country |
| What matters most to you in choosing a course |
| Search |
| Magazine |
| Newspaper Article |
| X Education Forums |
| Newspaper |
| Digital Advertisement |
| Through Recommendations |
| Receive More Updates About Our Courses |
| Update me on Supply Chain Content |
| Get updates on DM Content |
| I agree to pay the amount through cheque |
| A free copy of Mastering The Interview |
| Do Not Call |
| Lead Number |
| Prospect ID |

Notes:
Rows with some misc. missing values are dropped.
~ 70 % lines are retained and used for model building.
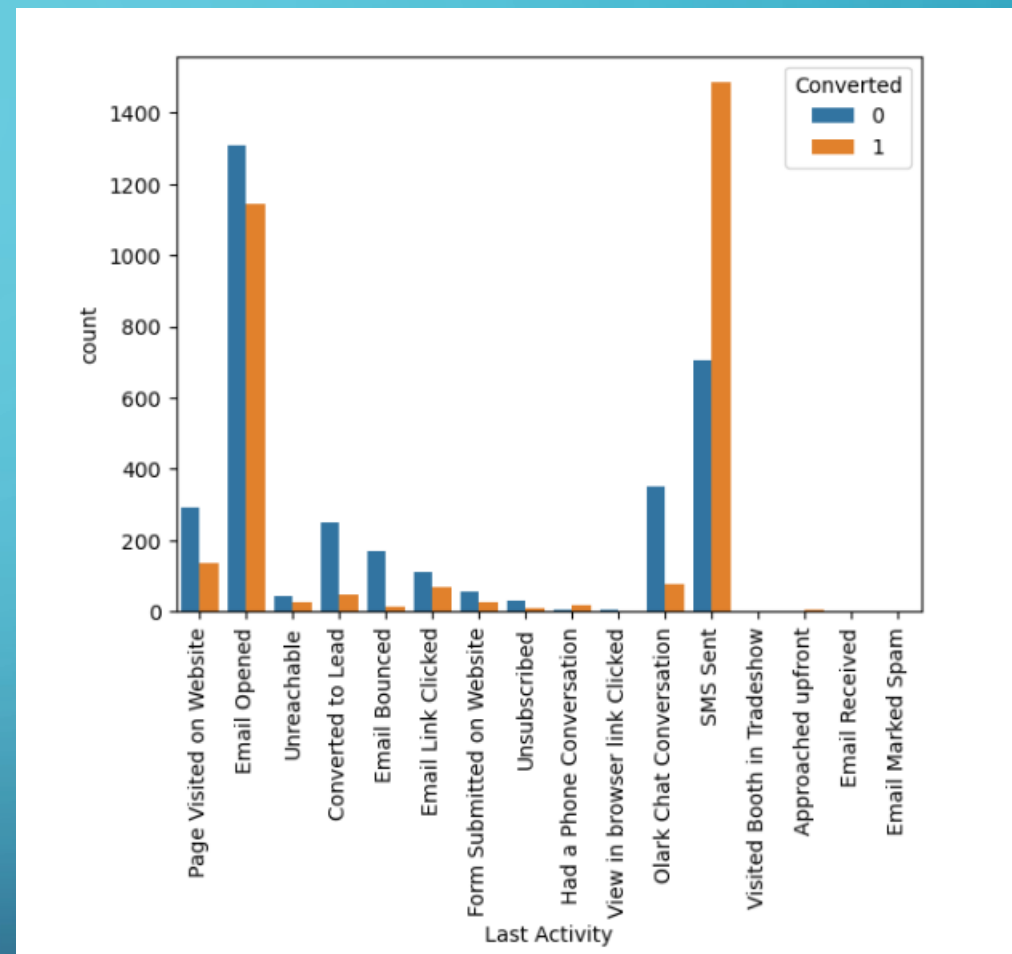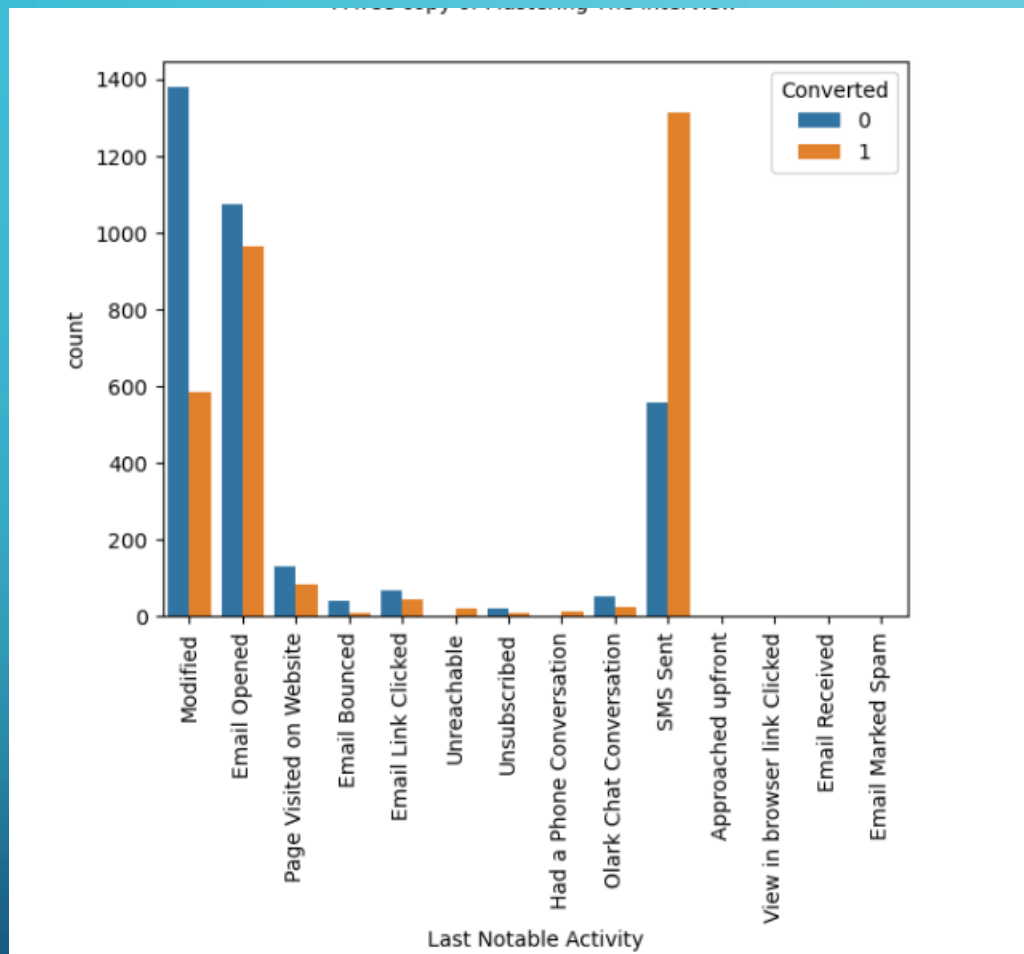
EDA / VISUALIZATION -1
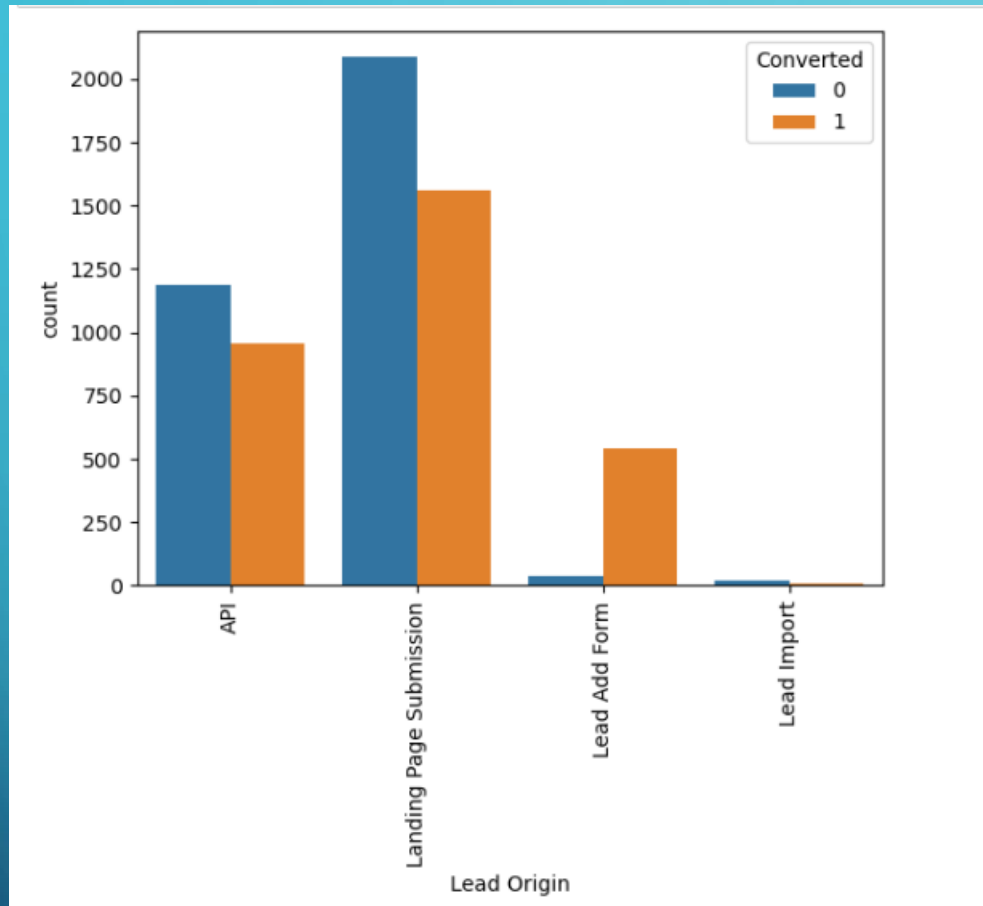
Count for Landing Page Submission is more but Conversion is higher for Lead Add Form
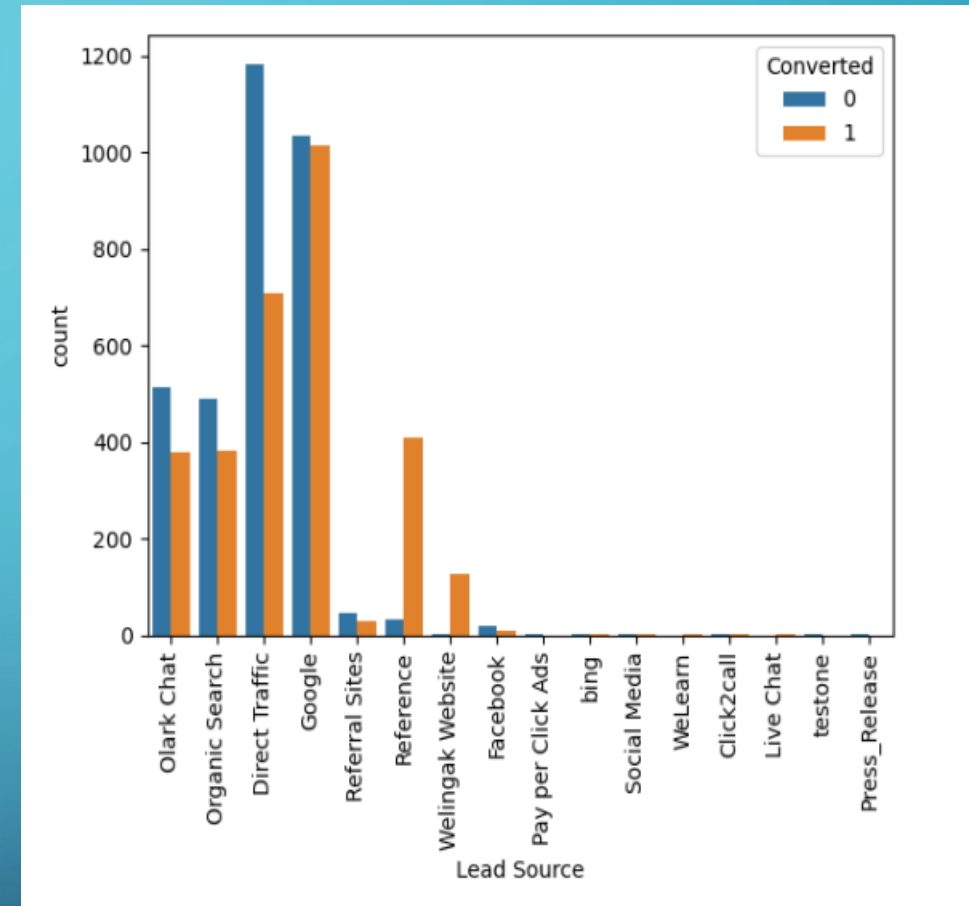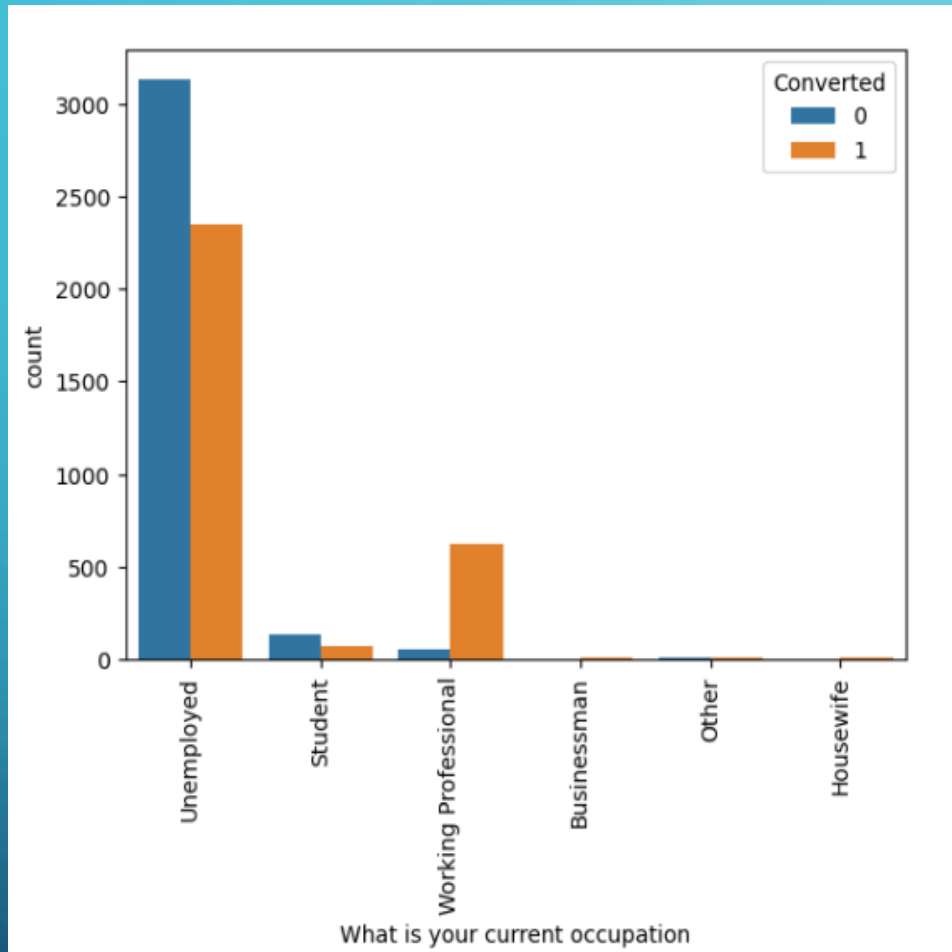
Lead Source is one of the important column for insights. There are many categories like Welingak Website, Olark Chat can be useful.
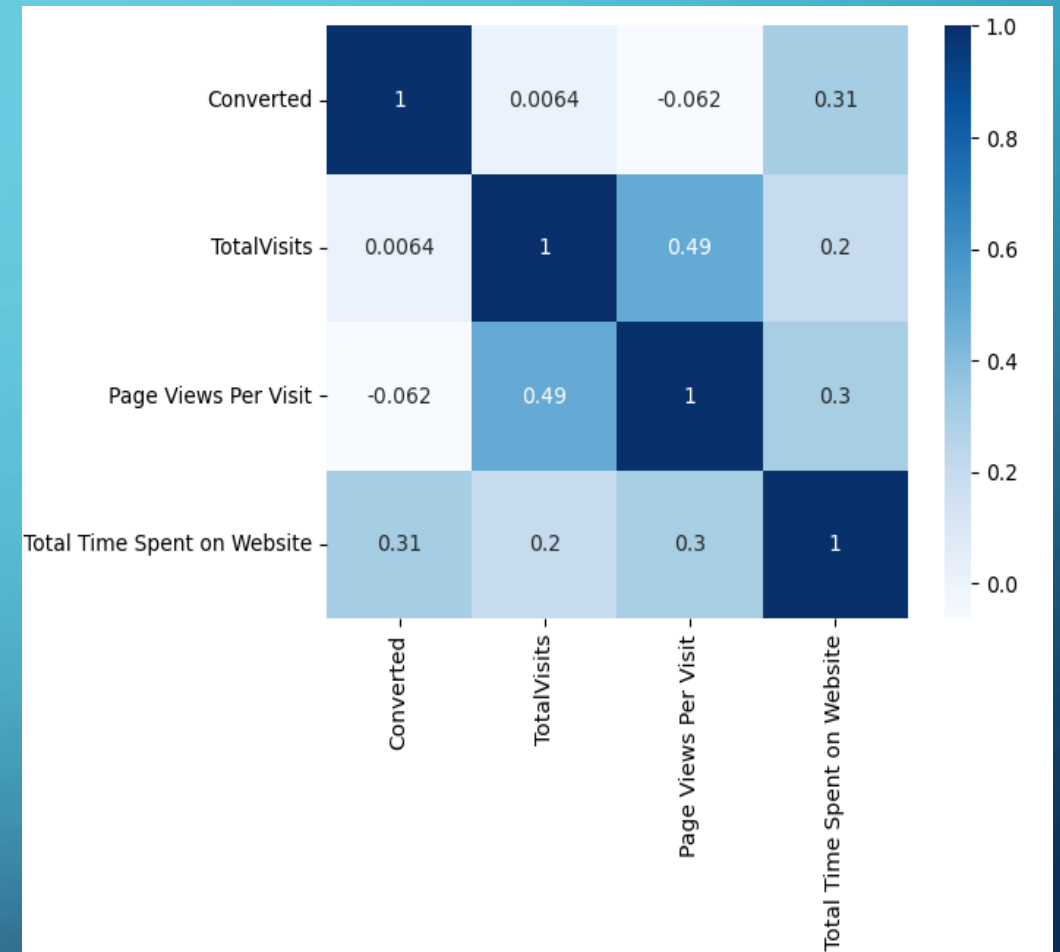
Last Activity & Last Notable Activity are useful features for Conversion.

EDA / VISUALIZATION - 3

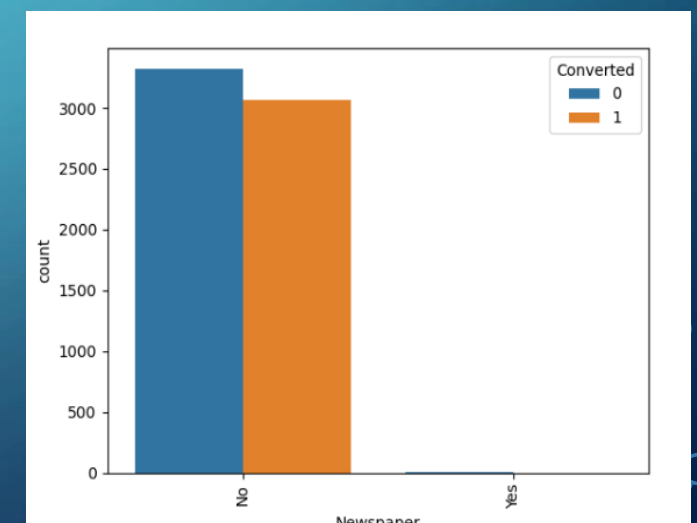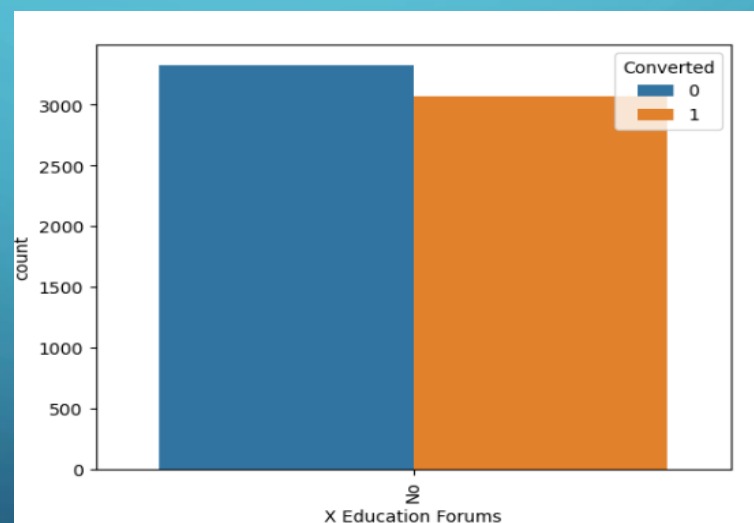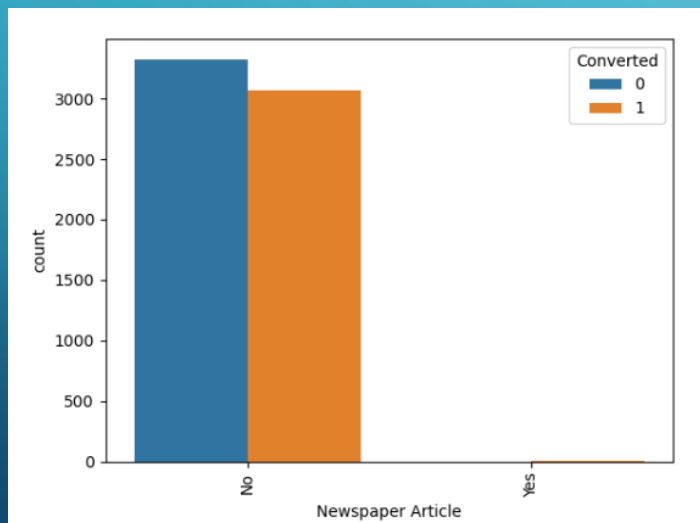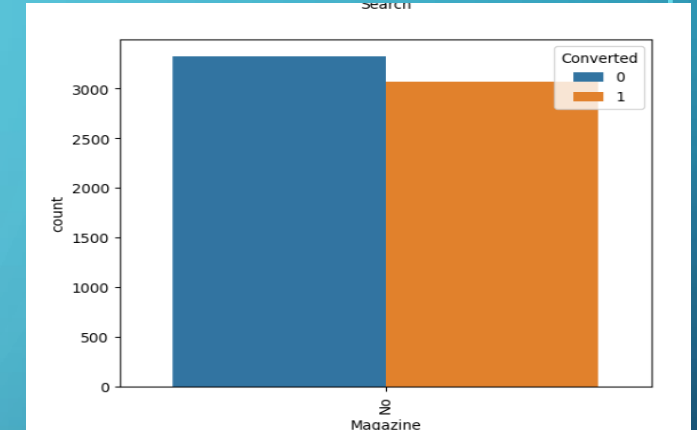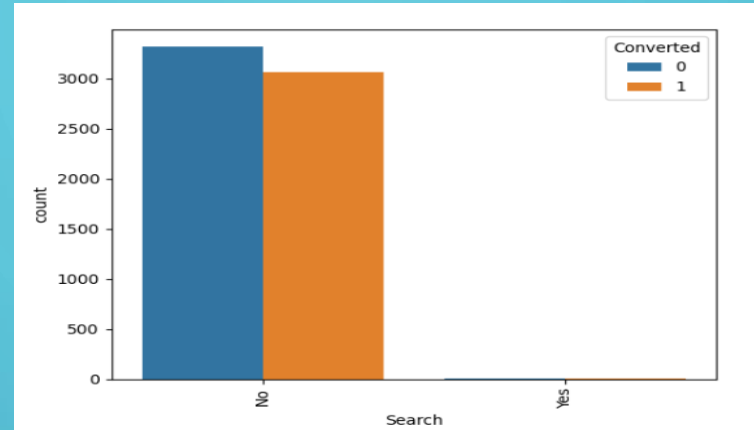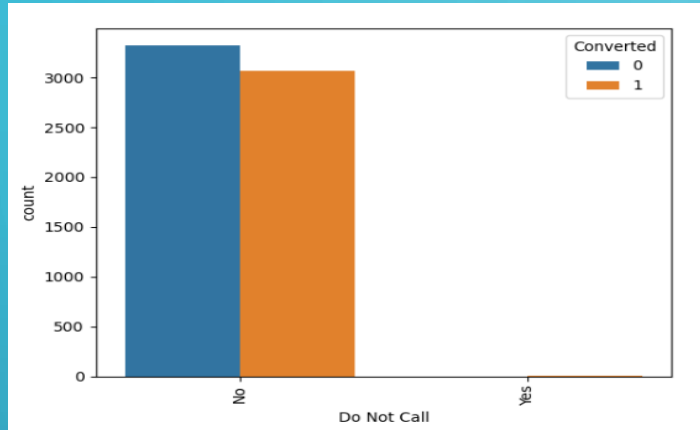Conversion rate for Working Professional is more

Chances of conversion is more for Leads with higher number of visit and leads spending more time on website

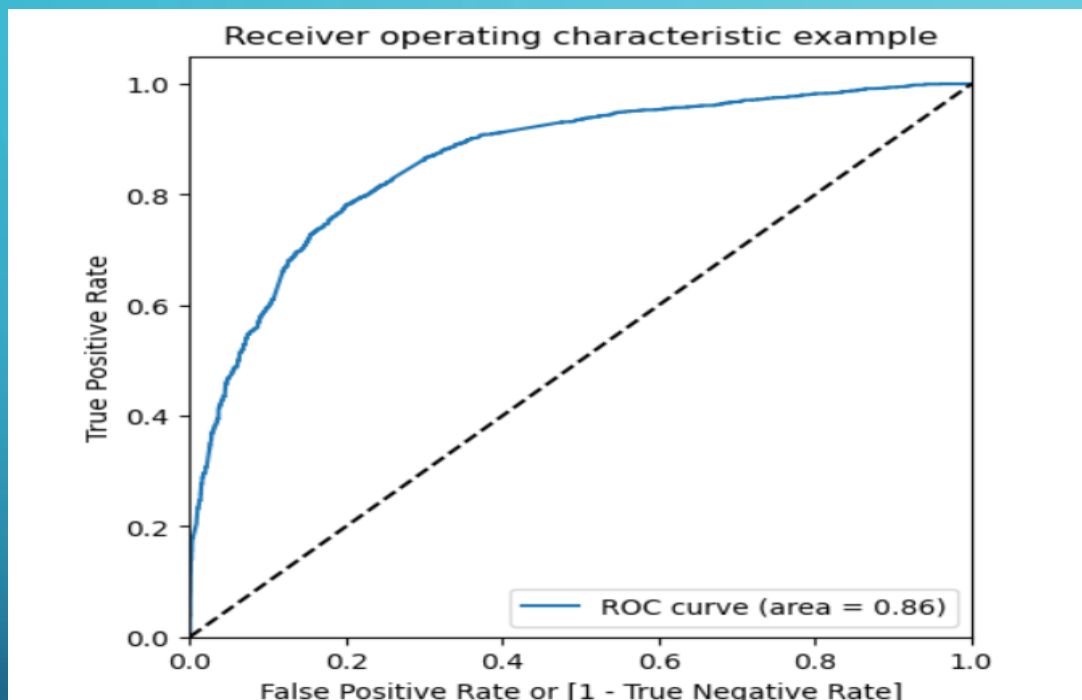# EDA / VISUALIZATION (COLUMNS DROPPED AS THOSE ARE NOT PROVING MUCH INFORMATION)

Note : There are more such graphs are plotted in Jupyter notebook, few of them shared in this slide.
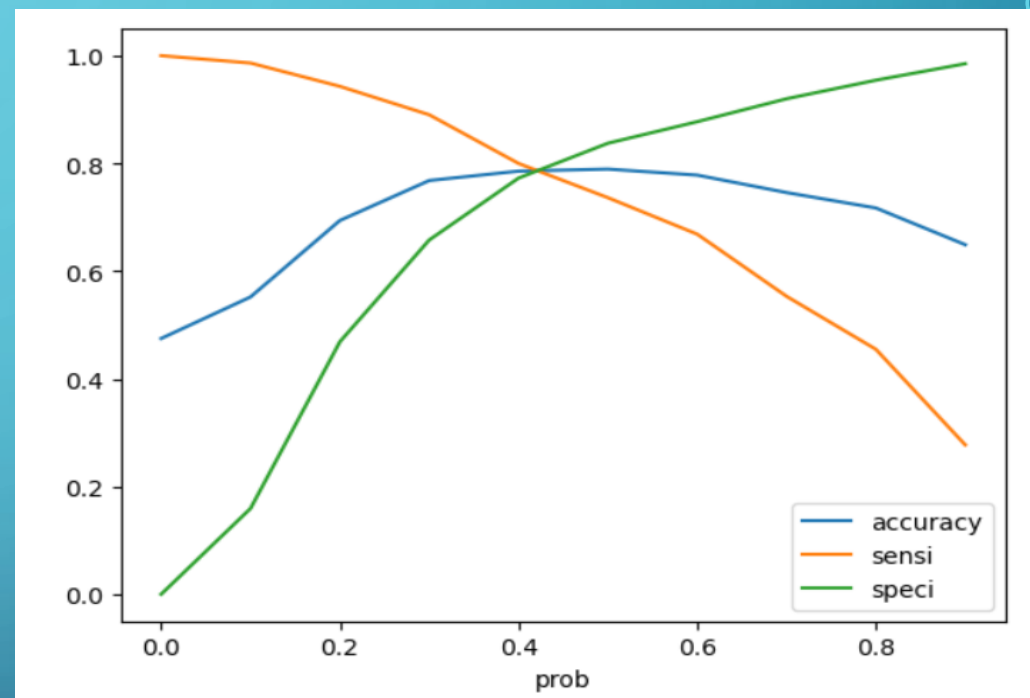
MODEL EVALUATION - ROC CURVE

ROC Curve (are=0.86)

Cutoff probability is 0.41

Accuracy of training data
Accuracy   : 78.81%
Sensitivity : 79.24%
Specificity : 78.42%

Accuracy of testing data
Accuracy   : 79.04%
Sensitivity : 79.28%
Specificity : 78.81%

CONCLUSION & RECOMMENDATIONS

Below are important/Top 3 Features
1. TotalVisits
2. Total Time Spent on Website
3. Lead Origin to be more specific - Lead Origin_Lead Add Form

Recommendations
Company should make calls to the leads following in below category as chances of they are getting converted to pai customer is more.

1. Leads with more total visits.
2. Leads who are spending more time on website.
3. Leads whose origin is Lead add form.
4. Leads whose current occupation is Working Professional.
5. Leads whose source is Welingak Website.
6. Leads whose last activity is 'Had a Phone Conversation'.
7. Leads whose source is Olark Chat.