LEAD SCORE CASE STUDY

Submitted By-

Edwin Novel

Sutanuka

Shivam Garg

Lead Score CaseStudy For X Education

- Problem Statement-
- Industry professionals can purchase online courses from X Education, a company that provides education. Many experts interested in the courses visit their website on any given day and search for courses.
- Upon arriving at the website, these visitors may browse the courses, submit a form for the course, or watch some videos. These persons are categorised as leads when they fill out a form with their phone number or email address. Additionally, the business receives leads from earlier recommendations.
- Once these leads are obtained, sales team members begin calling, sending emails, etc. Some leads are converted during this procedure, but most are not. At X Education, the normal lead conversion rate is roughly 30%.

Business Goal:

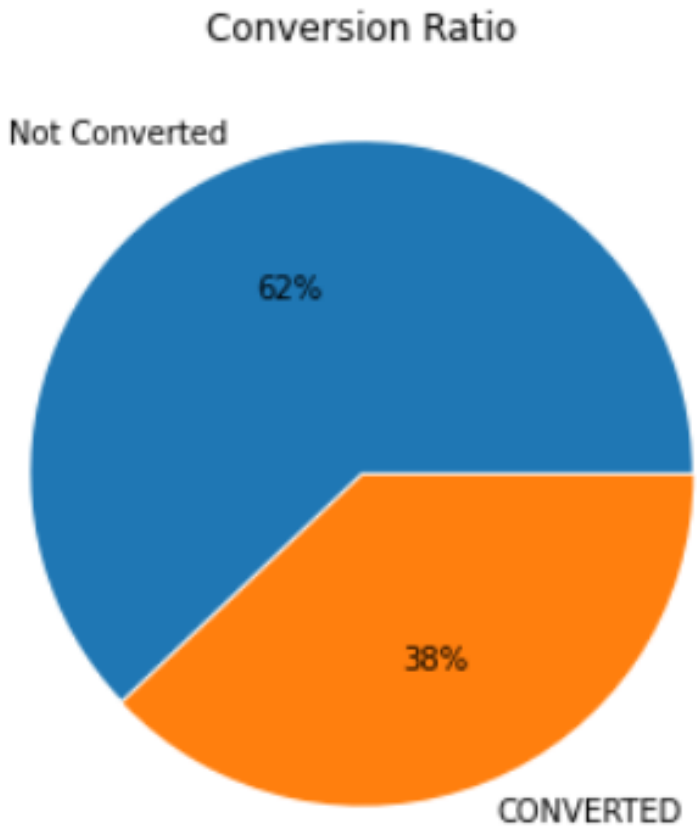
- In order to choose the leads that have the best chance of becoming paying clients, or the most promising prospects, X Education requires assistance.
- The business needs a model where each lead is given a lead score, and leads with higher lead scores have a better chance of converting, while leads with lower lead scores have a lower chance of converting.
- The desired lead conversion rate has been estimated by the CEO to be in the range of 80%.

Model building Process

- The data source for the analysis
- Exploratory Data Analysis after data preparation and cleaning of data
- Feature Scaling
- Dividing the dataset into a Train and Test dataset.
- Construction of a logistic regression model and calculation of Lead Score.
- Assessing the model using several metrics, such as precision and recall or specificity and sensitivity.
- Using the most appropriate model for the test data based on the sensitivity and specificity metrics.

Exploratory Data Analysis

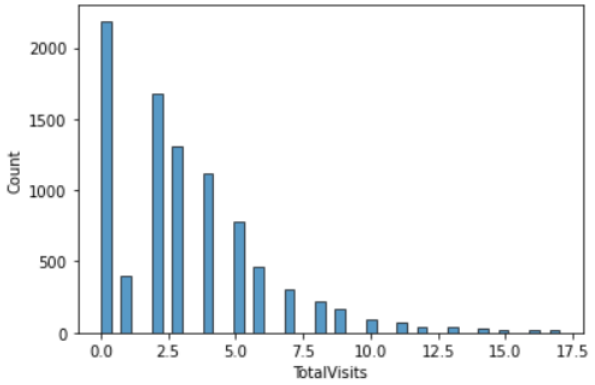
- Univariate Analysis-
  Conversion Ratio



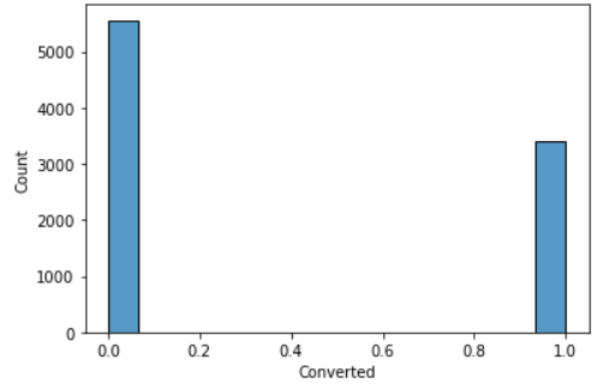
Univariate Analysis

Histogram Plots -

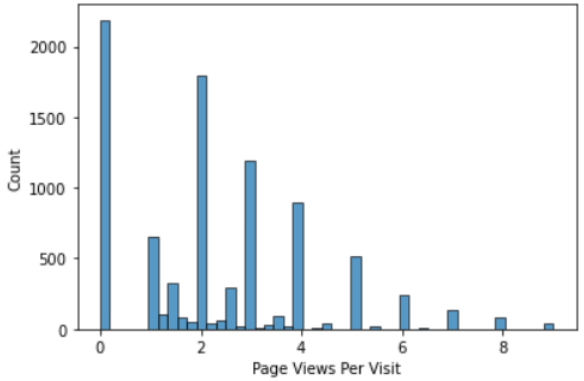
Histogram of TotalVisits :



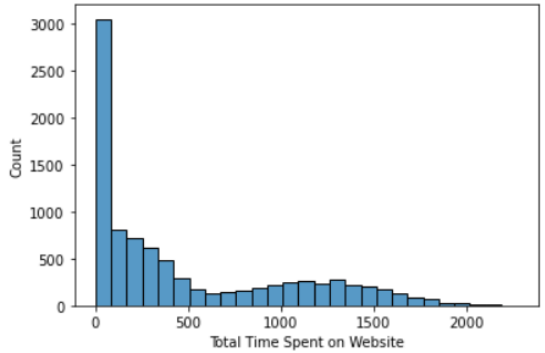
Histogram of Converted :



Histogram of Page Views Per Visit :

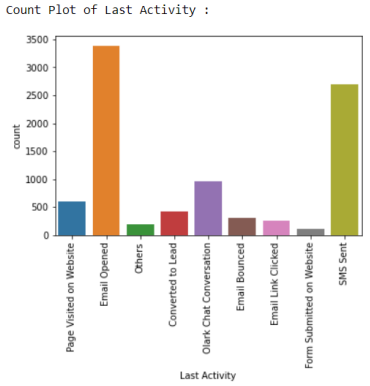


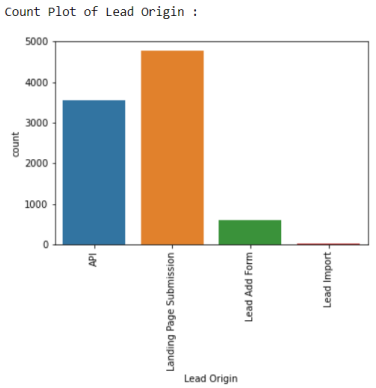
Histogram of Total Time Spent on Website :

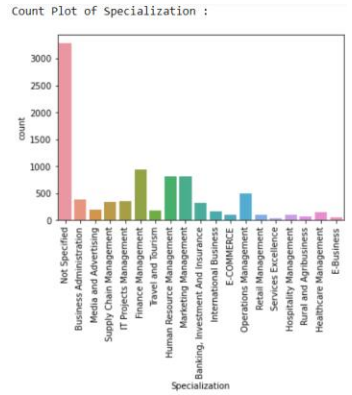


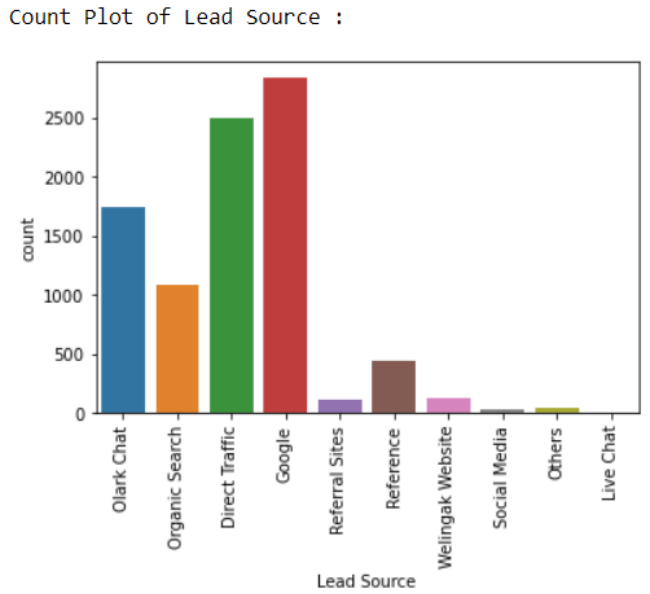
Univariate Analysis

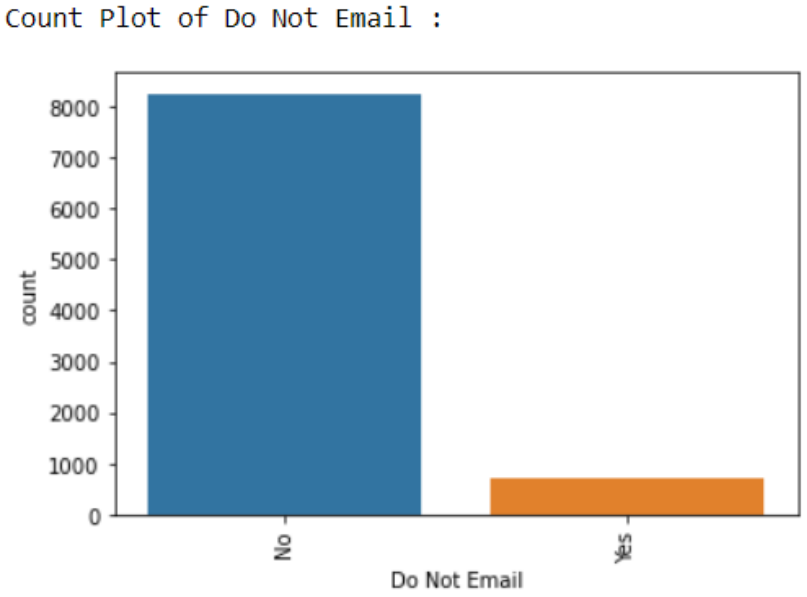
- Categorical Data-Most Leads are from Landing Submission Page, and API.
- Lead are coming Google, Direct Traffic and Olark Chat.
- Most of the leads want to get email about their course hence they have opted No.
- Lead counts are high for SMS Sent and Email Opened.





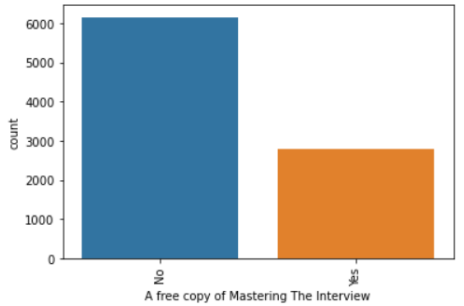




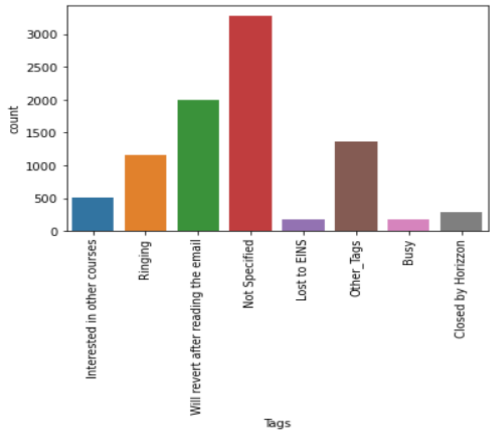


- We can see most of the leads are from Finance, HR and Marketing Management.
- Unemployed and Working Professional have high leads count.
- Most of the leads are from Mumbai.

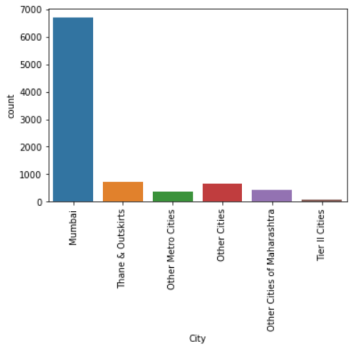
Count Plot of A free copy of Mastering The Interview



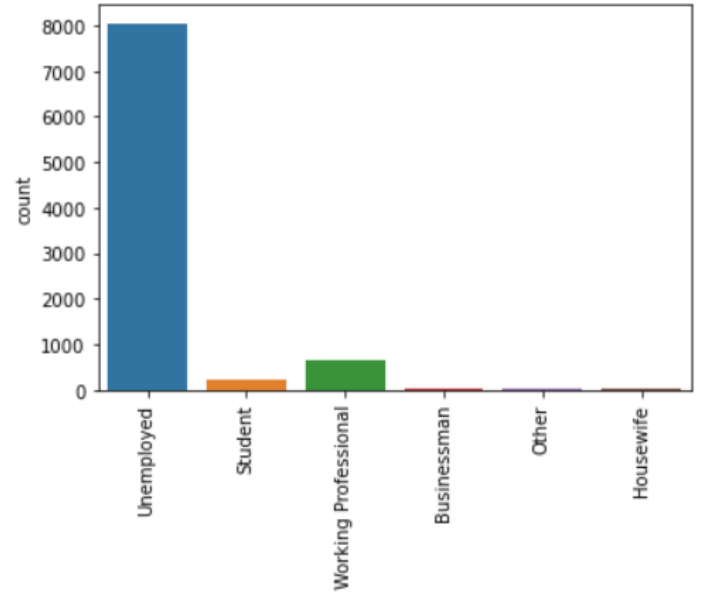
Count Plot of Tags :



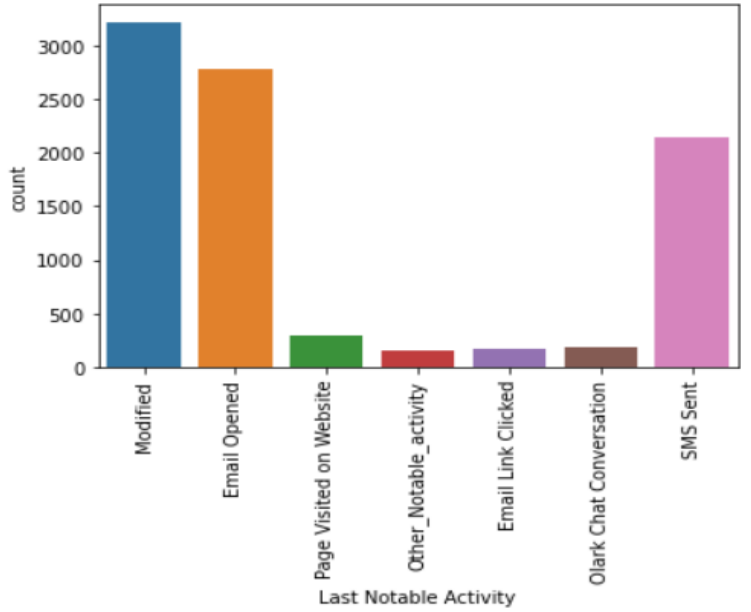
Count Plot of City :



Count Plot of What is your current occupation :

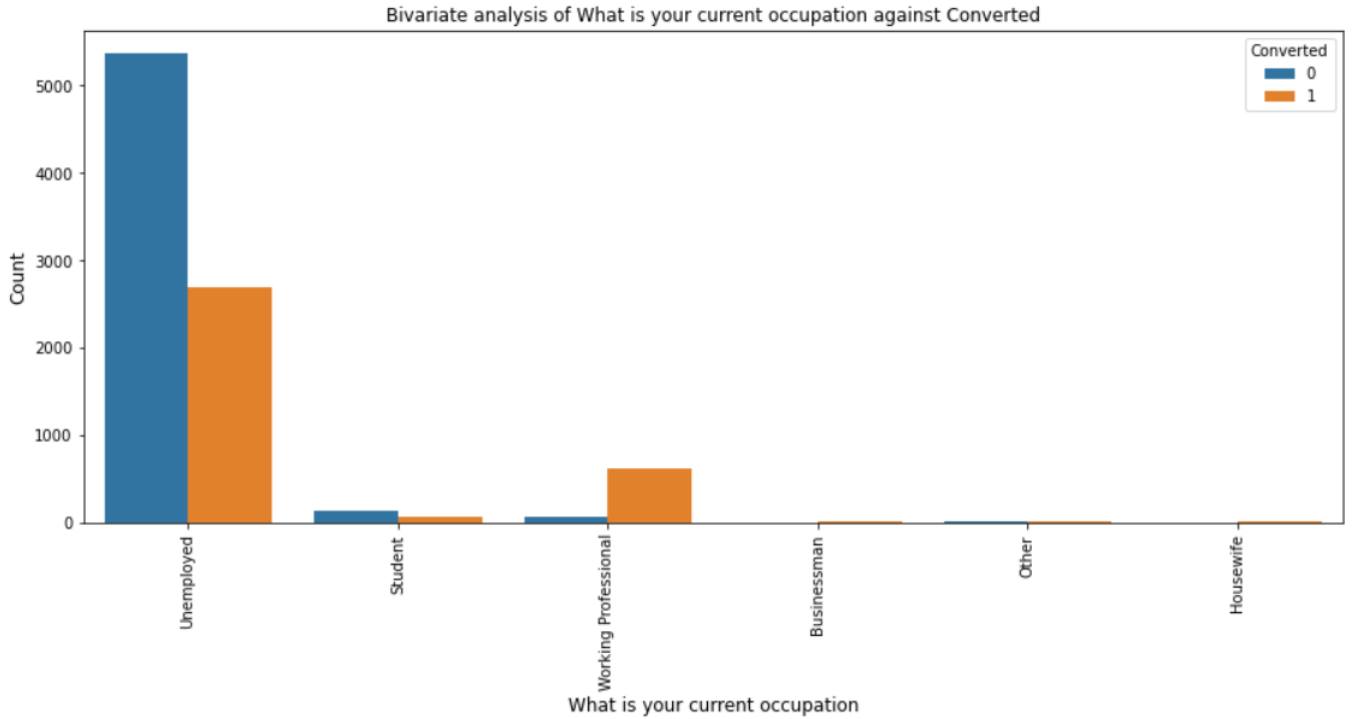


Count Plot of Last Notable Activity :

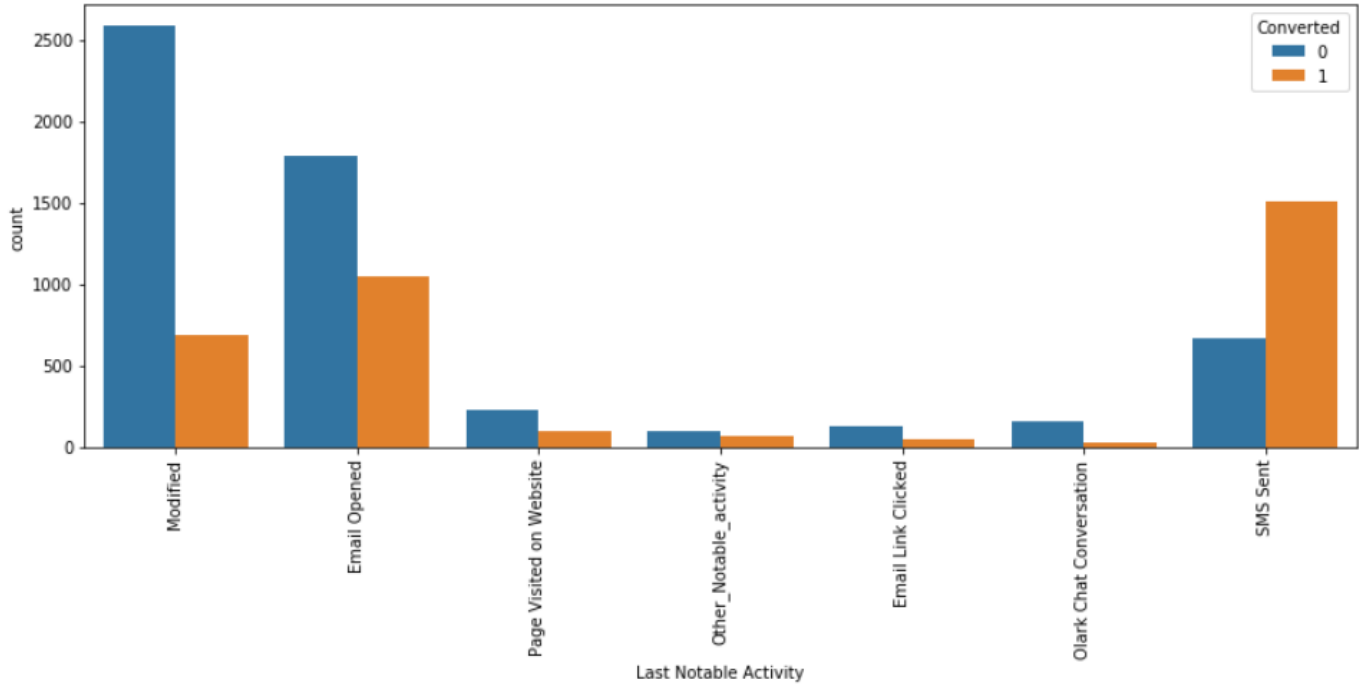


Bivariate Analysis

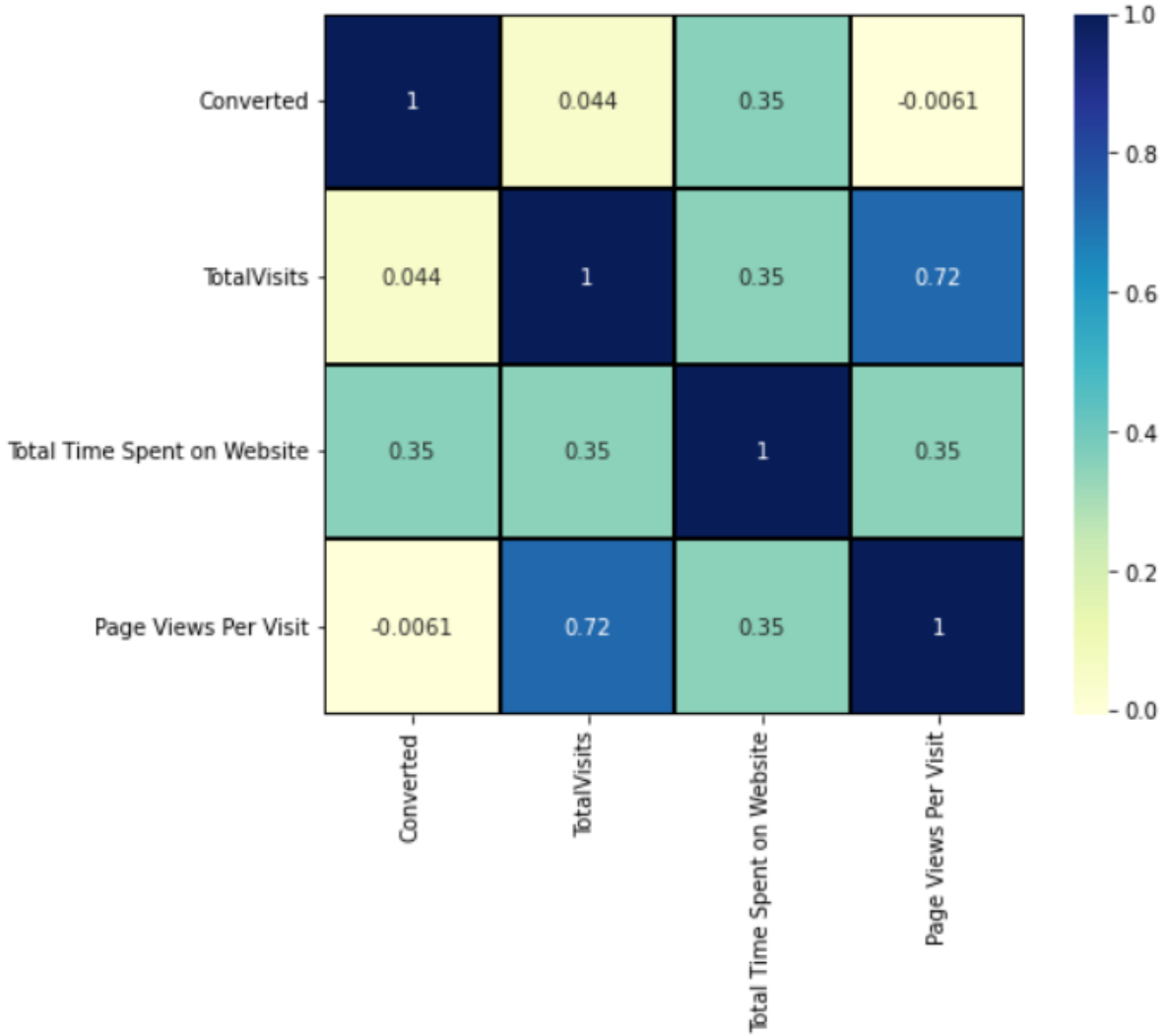
- More conversions occurred with unemployed people.



- SMS Sent's Last Activity value had a higher conversion rate.

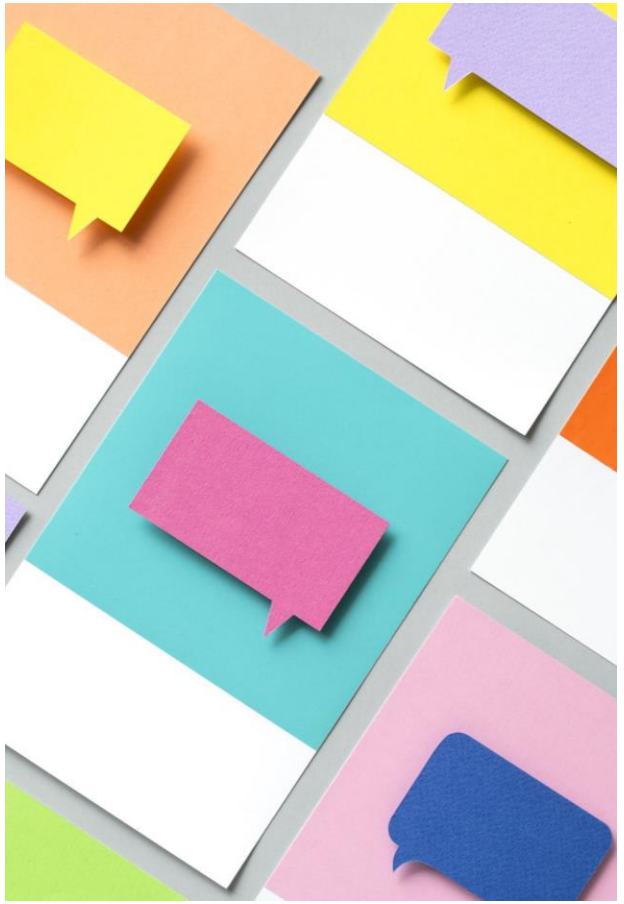


Correlation Matrix



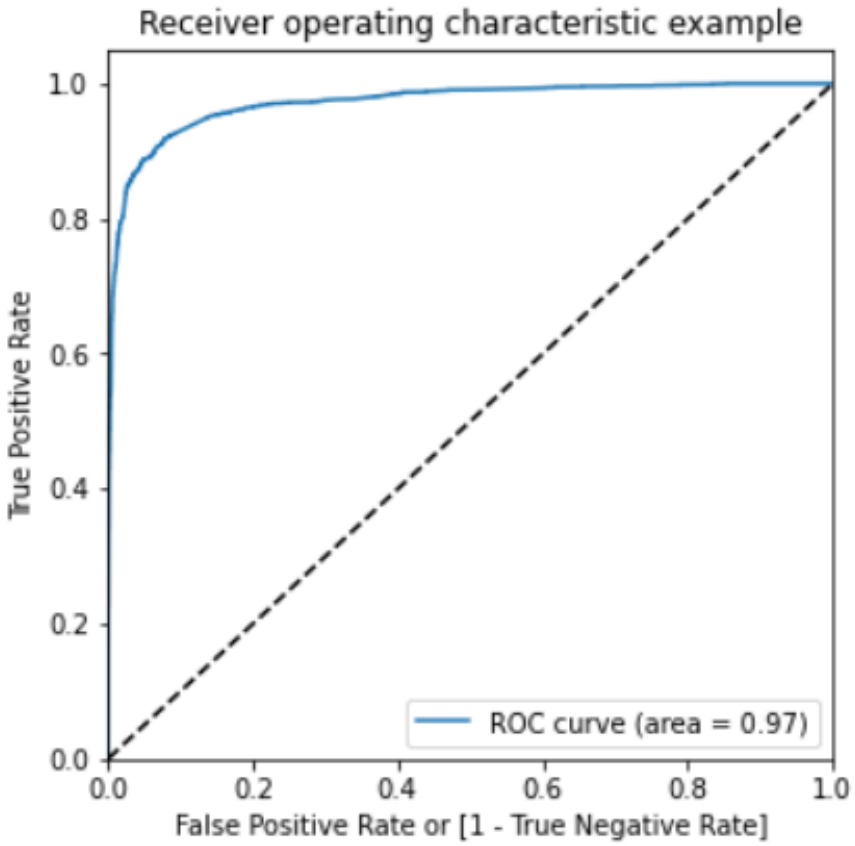
Features Impacting Converting Rate

- Last Notable Activity_Modified
- Lead Source_Google
- Lead Source_Direct Traffic
- Last Activity_SMS Sent
- Tags_Will revert after reading the email
- Tags_Other_Tags
- Tags_Ringing
- Last Activity_Olark Chat Conversation
- Lead Source_Organic Search
- Total Time Spent on Website
- Tags_Interested in other courses
- Tags_Closed by Horizzon
- Tags_Lost to EINS
- Lead Source_Referral Sites
- Lead Source_Welingak Website
- Last Notable Activity_Email Link Clicked



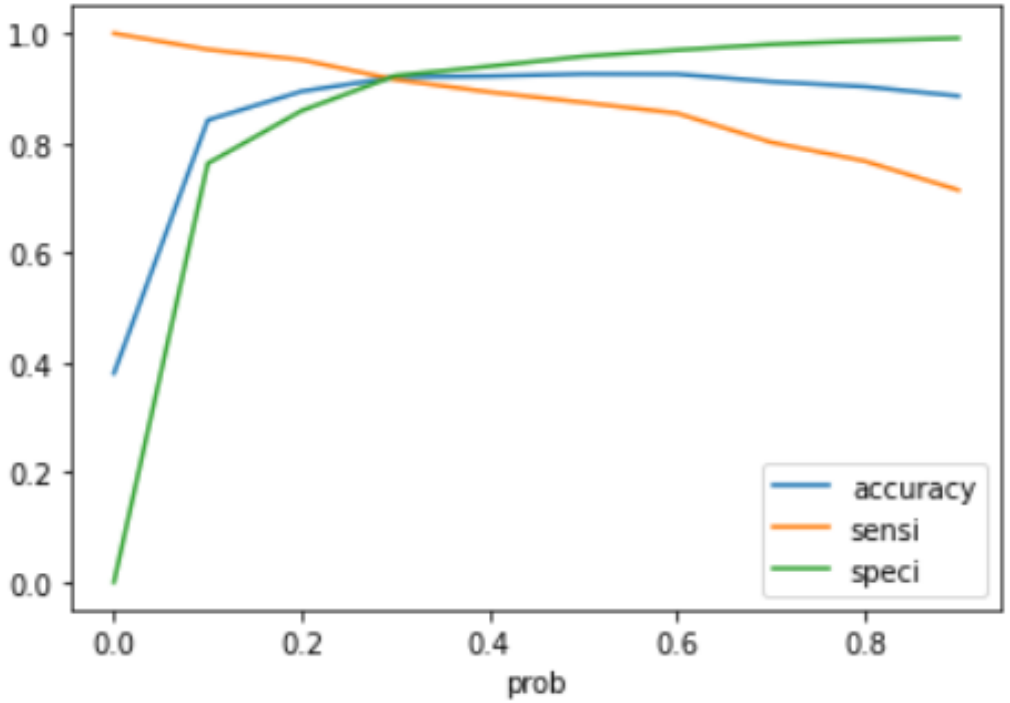
Model Evaluation - Sensitivity and Specificity on Train Data Set

- **ROC Curve**
- **The ROC curve we are seeing (area = 0.97) indicates a strong predictive model.**



Accuracy, sensitivity and specificity for various probabilities

From the curve 0.3 is the optimal point of Cutoff - Probability



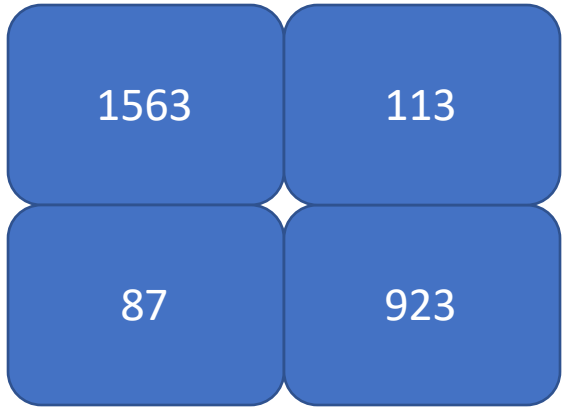
Confusion matrix

3583	299
201	2184

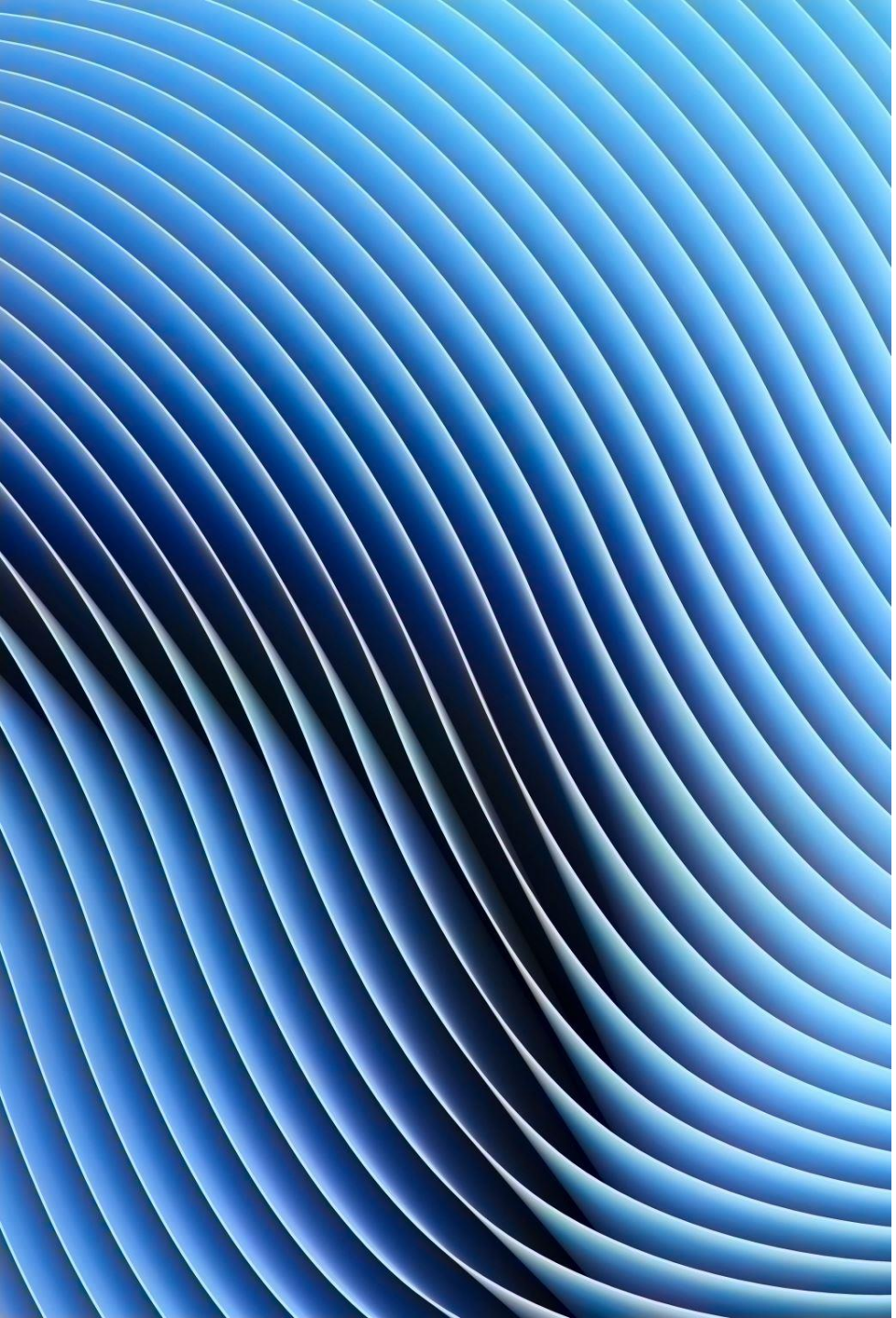
- **Accuracy** : 92.02%
- **Sensitivity** : 91.57%
- **Specificity** : 92.29%
- **.Precision** : 87.96%
- **Recall** : 91.57%

Prediction on test dataset

- Confusion Matrix



- **Accuracy** : 92.55%
- **Sensitivity** : 91.39%
- **Specificity** : 93.26%
- **Precision** : 89.09%
- **Recall** : 91.37%



Conclusion

1. Because the lead conversion rate from lead sources (such Google and direct traffic) is so high in comparison to other sources, they can help a firm generate a lot of business.
2. The likelihood of a lead converting increases if they spend more time on the website and were last contacted by SMS.
3. Leads who are unemployed can be targeted because they are more likely to choose the course because it will improve their career chances.
4. The working professionals are the next category that can be targeted because they are more likely to enroll in the course in order to increase their employment prospects.
5. People who use email are more inclined to choose the courses.