Lead Scoring Case Study

By Saurav, Veenal and Mrityunjoy

Case Study – Summary

- X Education, an Education company is into selling of online courses to industry professionals.
- Company markets its courses on several websites and search engines like Google and visitors land on the website, browse the courses or fill up a form for the course or watch some videos.
- Visitors fill up a form providing their email address or phone number, they are classified to be a lead.
- The company also gets leads through past referrals and these leads are acquired, employees from the sales team start approaching the potential leads by making calls, writing emails, etc.
- As per company information lead conversion rate is around 30%.

Case Study –Business Goal & Requirement

Business Goal:

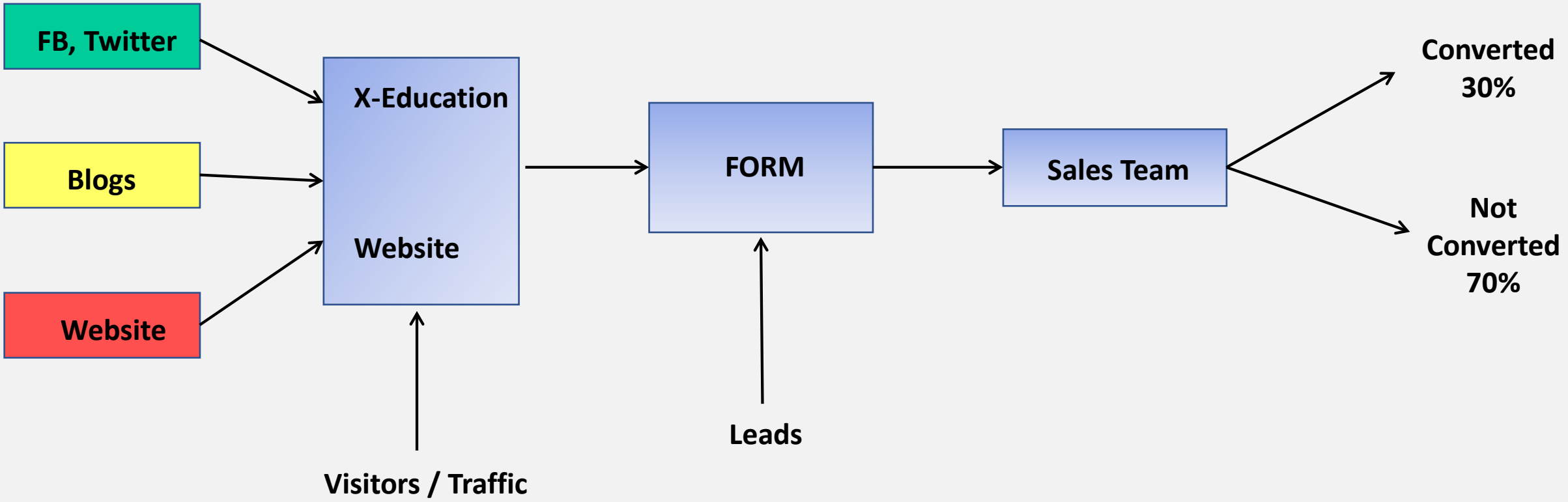
To find factors which would suffice the company's objective of identifying potential leads and their sources to increase their lead conversion, thereby adding to their overall revenue.

Requirement guideline:

A model with more than 80% conversion rate to assign was required to be built that would assign a lead score between 0 and 100 to each of the leads which would be used by the company to target potential leads.

X-Education Current Model

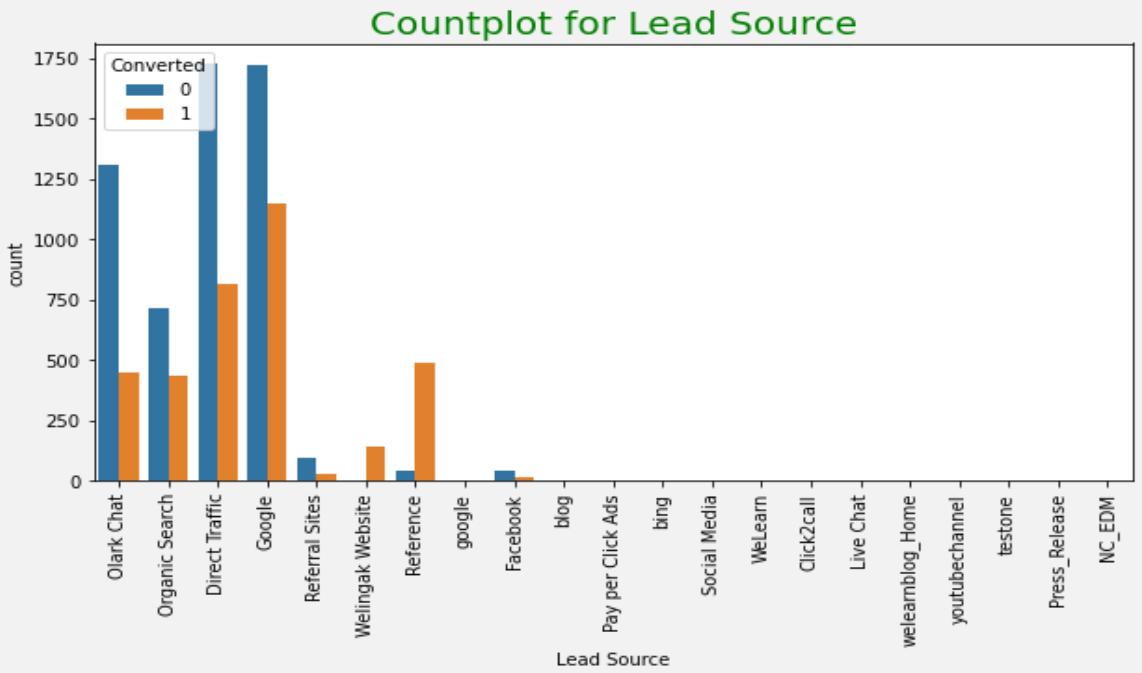
The diagram below shows that Visitors or traffic gets access to X Education website Landing page and fill up a Form. This Form is finally processed by the Sales Team based on the level of interest of each visitors towards the educational courses. Sales Team then finally makes a call based on the contact information shared and uses appropriate strategies to convert these leads into prospect.



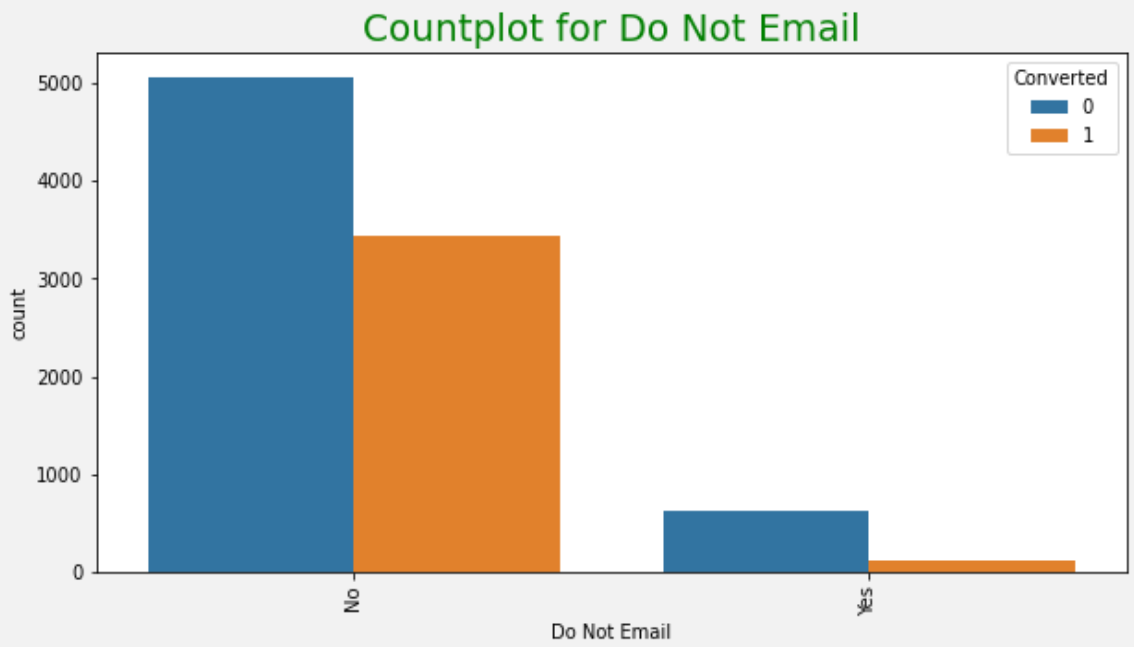
Lead Scoring – Strategies Applied

- Importing and reading the Data Frame
- Clean and prepare the Dataset for further Analysis like – missing values treatment etc.
- Exploratory Data Analysis to figure out most helpful attributes for conversion and advanced Data cleaning like – bucketing of attributes, imputing values, etc.
- Prepare the data for model building like creating Dummies for categorical variables, splitting data in to Test and Train sets, Feature Scaling using StandardScaler method etc.
- RFE for feature selection
- Logistic Regression Model building and Training the model
- Lead Score assignment for each lead
- Evaluating the model by applying different measures and metrics
- Testing the model on Test dataset
- Measuring the model accuracy and other metrics for evaluation

EXPLORATORY DATA ANALYSIS

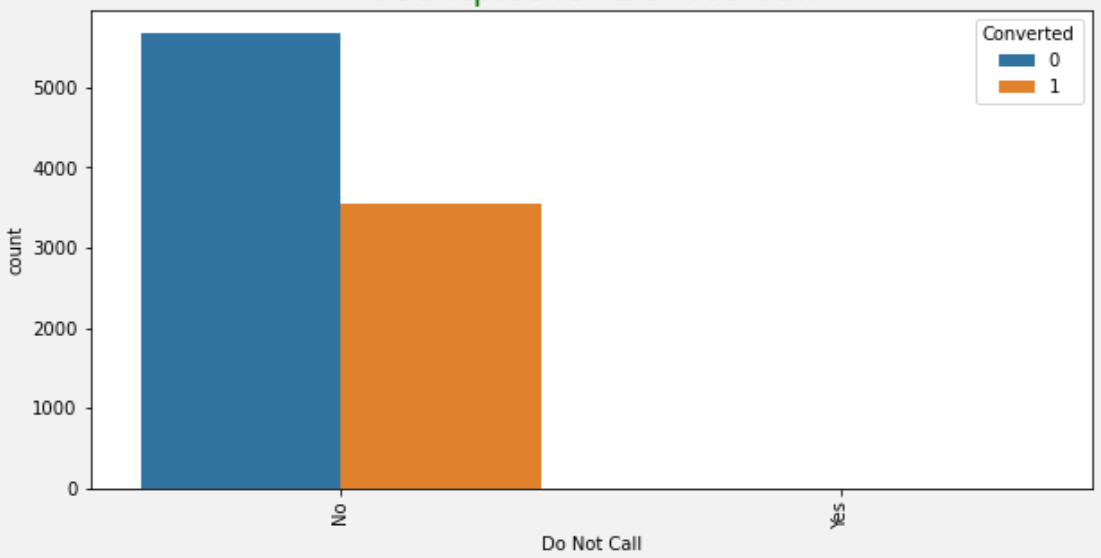


- Google search seems to provide a good number of leads
- Similar other platforms like Olark Chat, Direct Traffic and Organic Search
- Despite lower numbers the References and Welingak website, show high conversion rate.

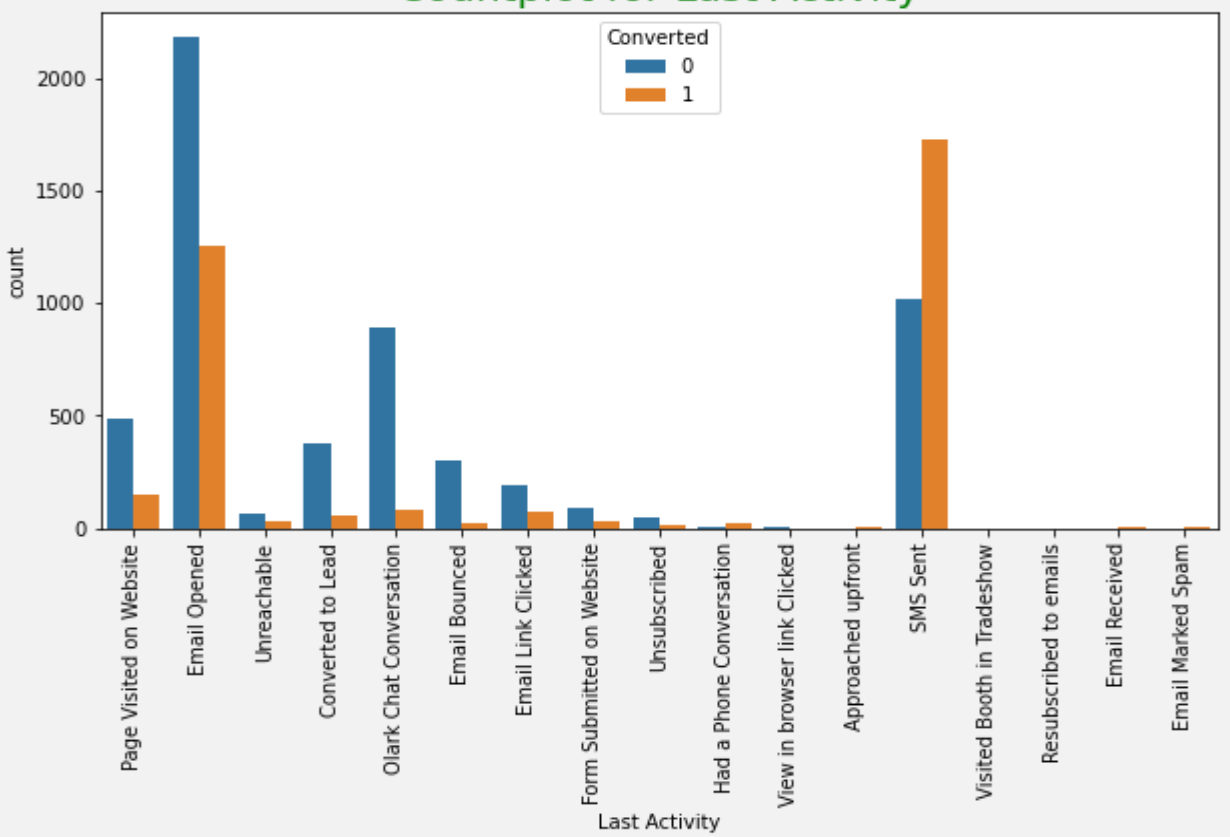


- The users that expressed their interests in getting email have higher chances of lead conversion

EXPLORATORY DATA ANALYSIS

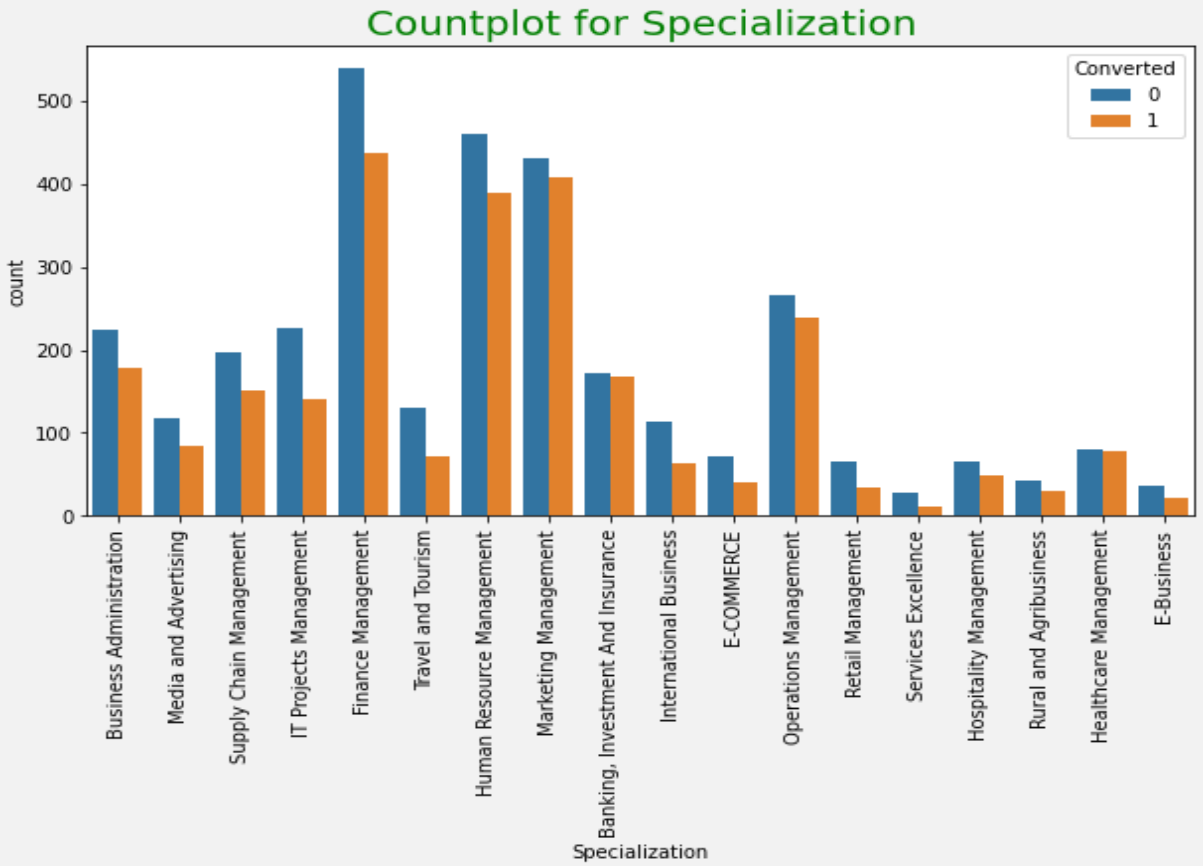


Most Leads preferred to get call via phone.

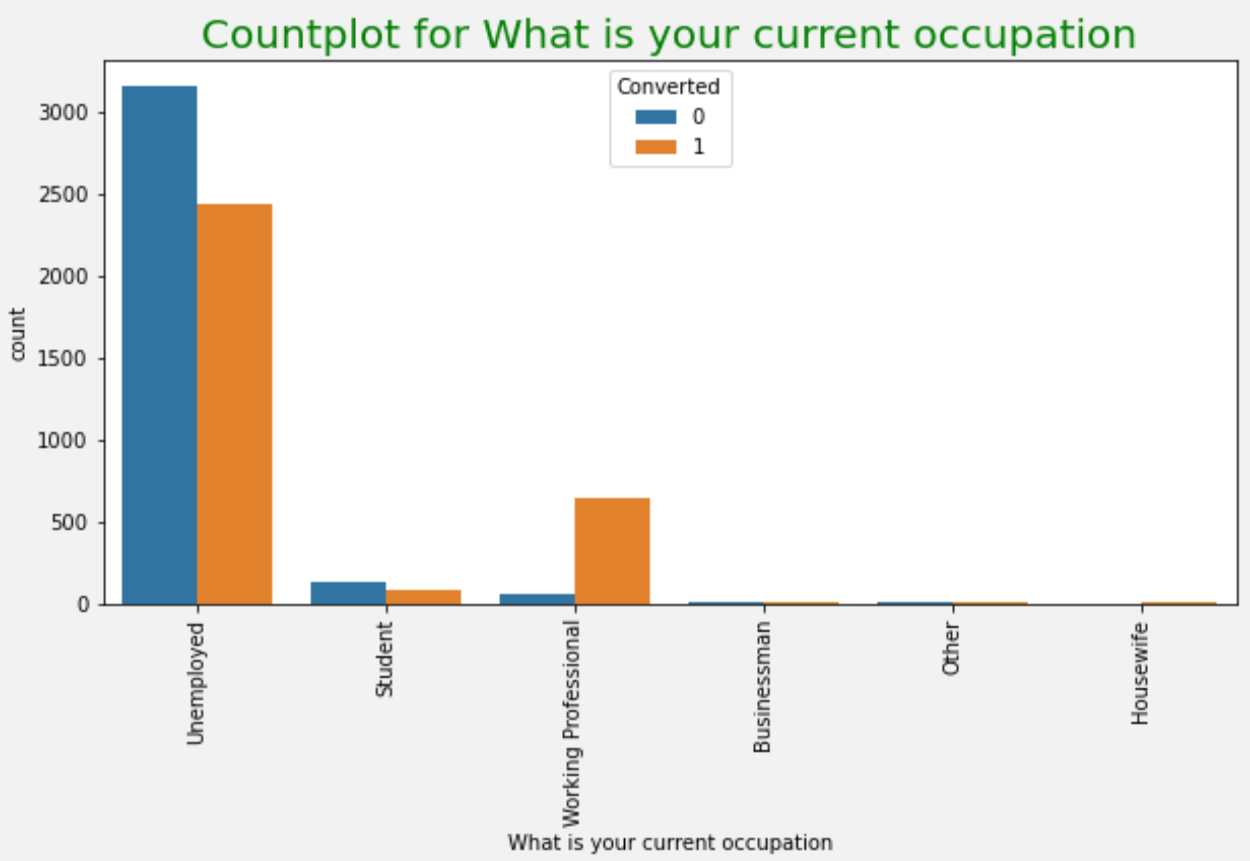


SMS seems to be the most preferred method followed by Emails, however users getting SMS could be identified as potential lead

EXPLORATORY DATA ANALYSIS

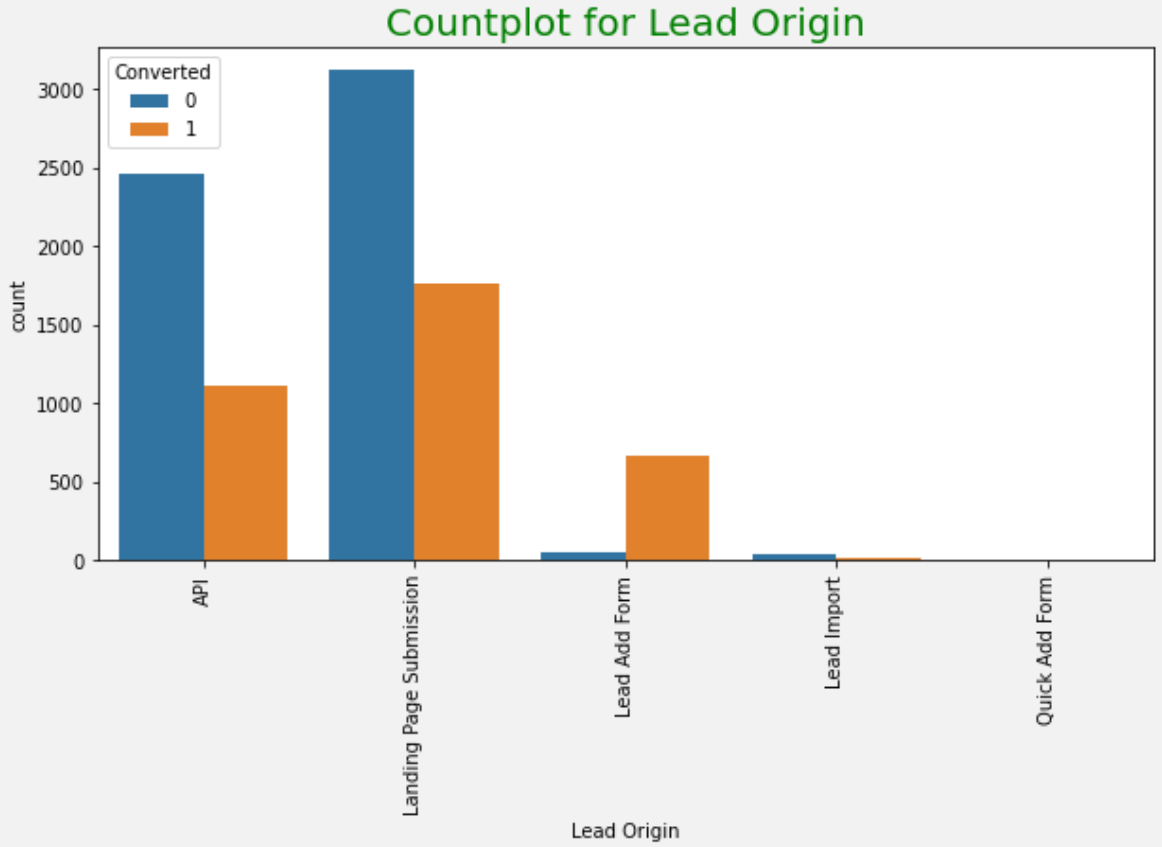


Individuals from Financial management made up the highest number of leads followed by Human Resource Management and Marketing Management.

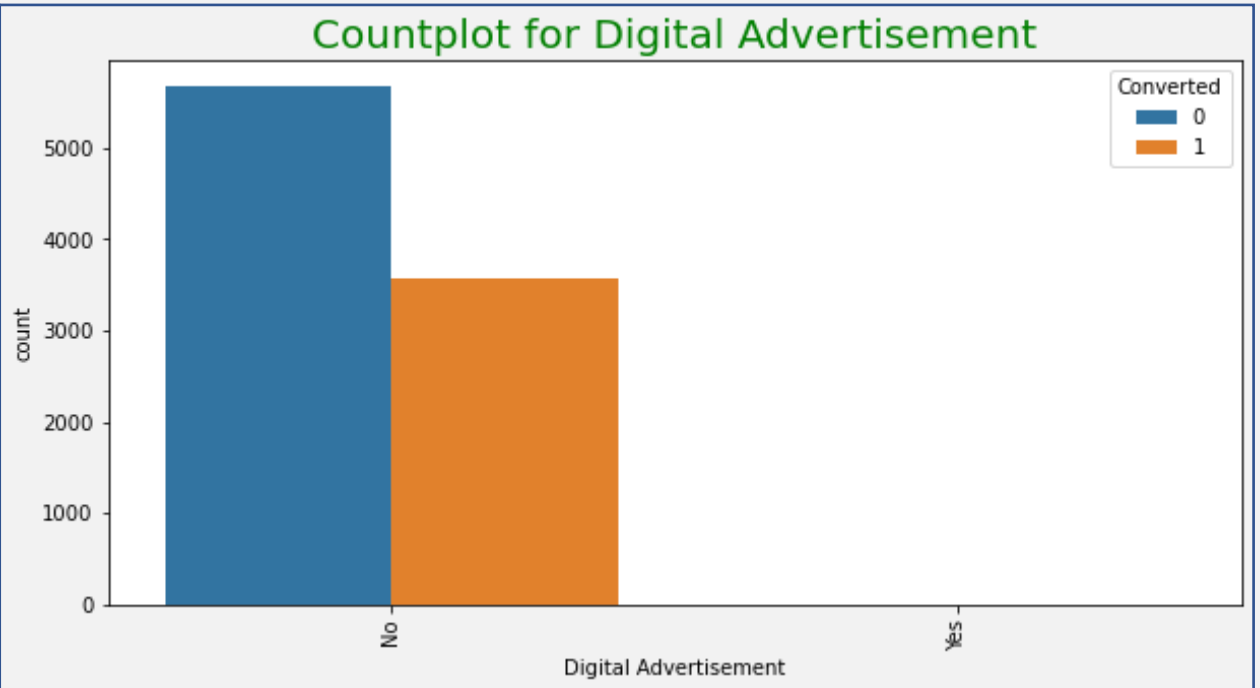


Unemployed forms the highest number of leads whereas Working Professionals could be targeted for better conversion.

EXPLORATORY DATA ANALYSIS

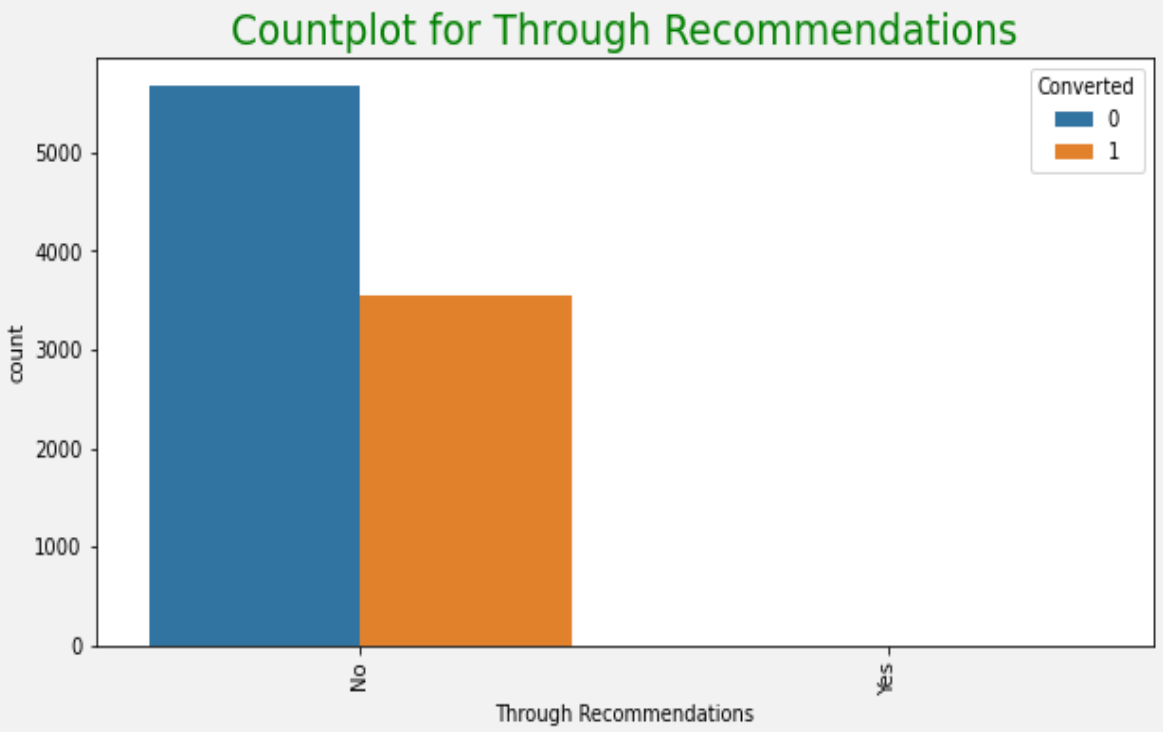


- Landing page submissions is most common Lead Origin, followed by API.
- Lead Add Form is having highest conversion rate.

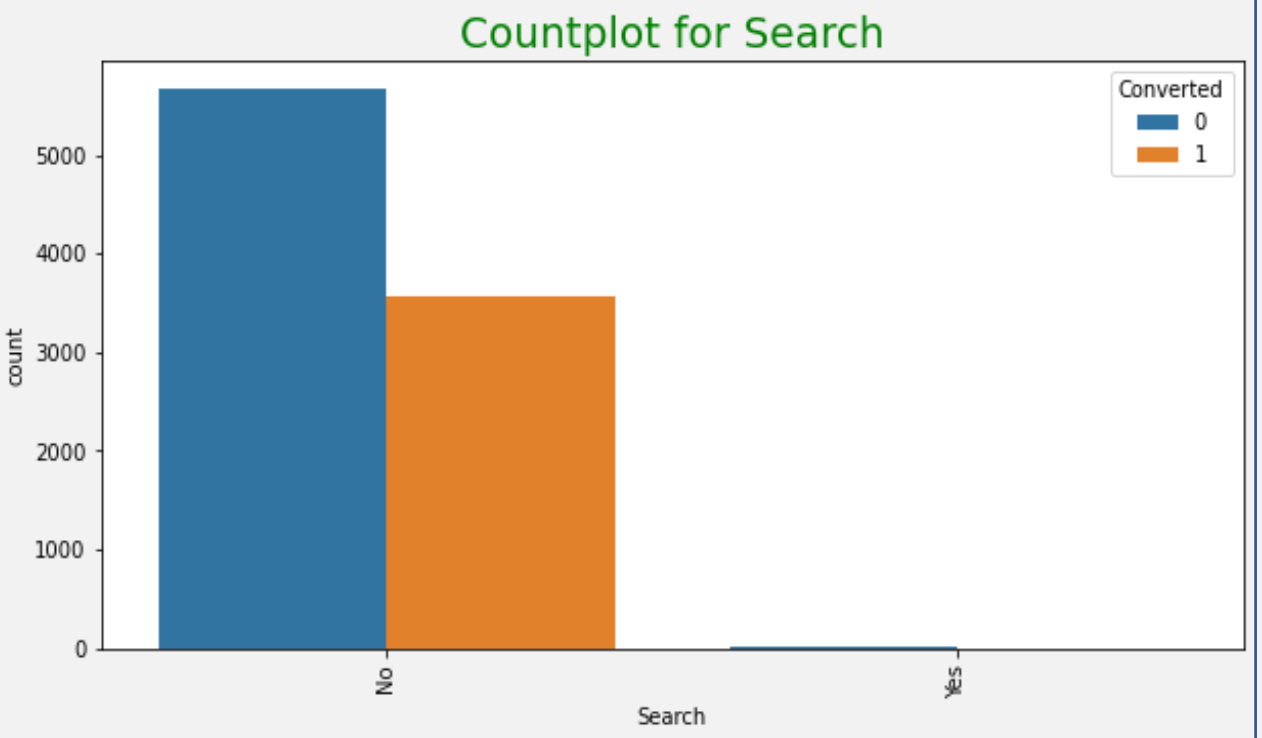


Digital advertisement doesn't seem to contribute to promising leads as there are negligible leads that could be identified.

EXPLORATORY DATA ANALYSIS

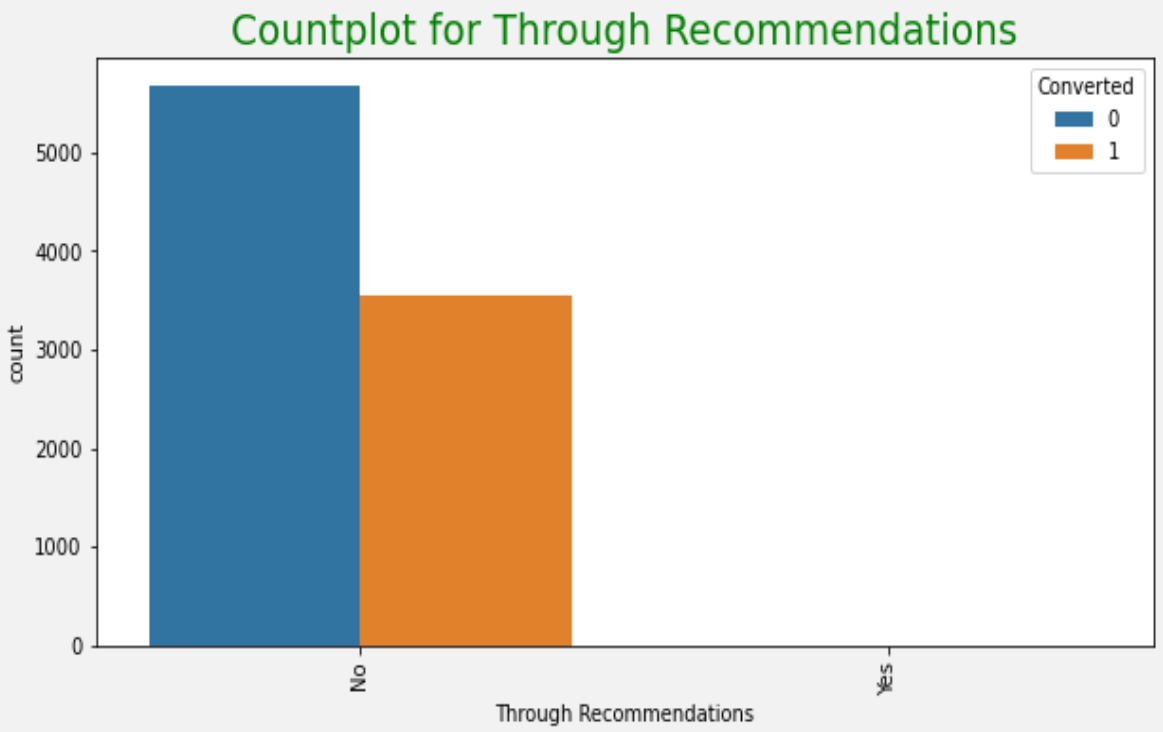


Above graph shows that Recommendations is not a good source for lead conversion.

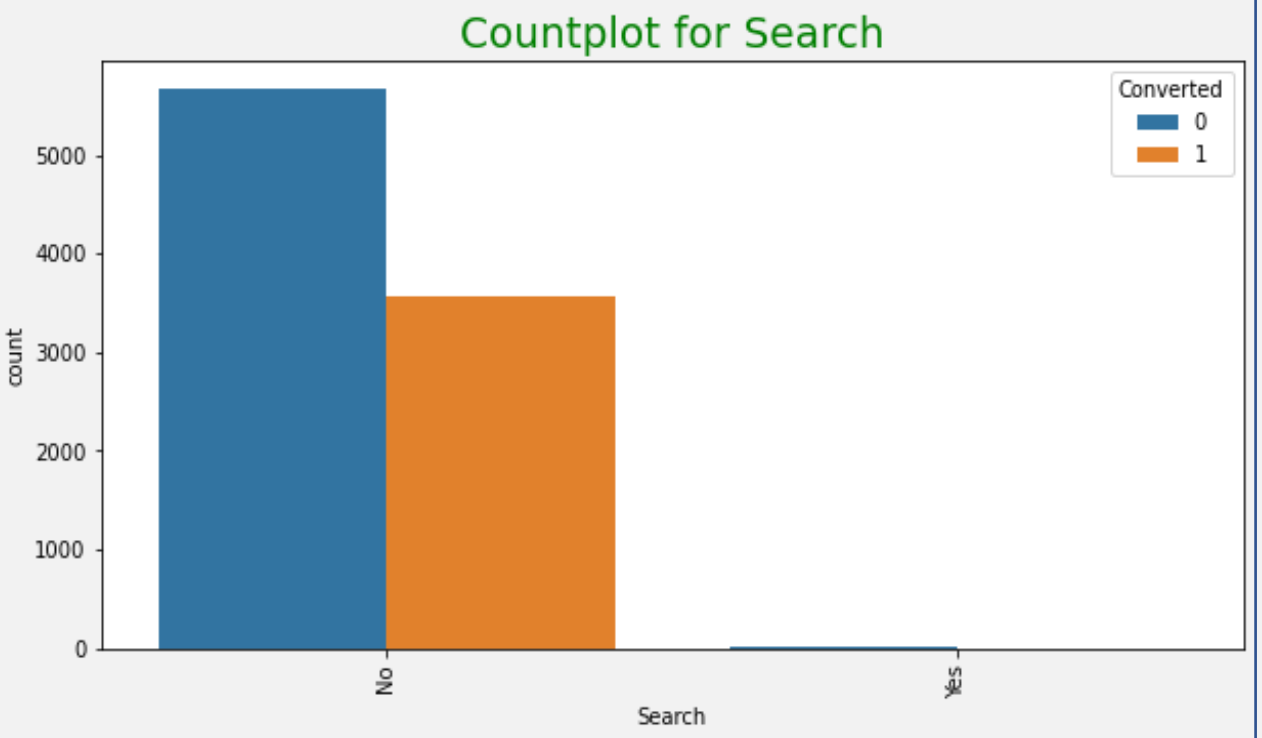


Search also doesn’t seem to be a good source for lead conversion.

EXPLORATORY DATA ANALYSIS

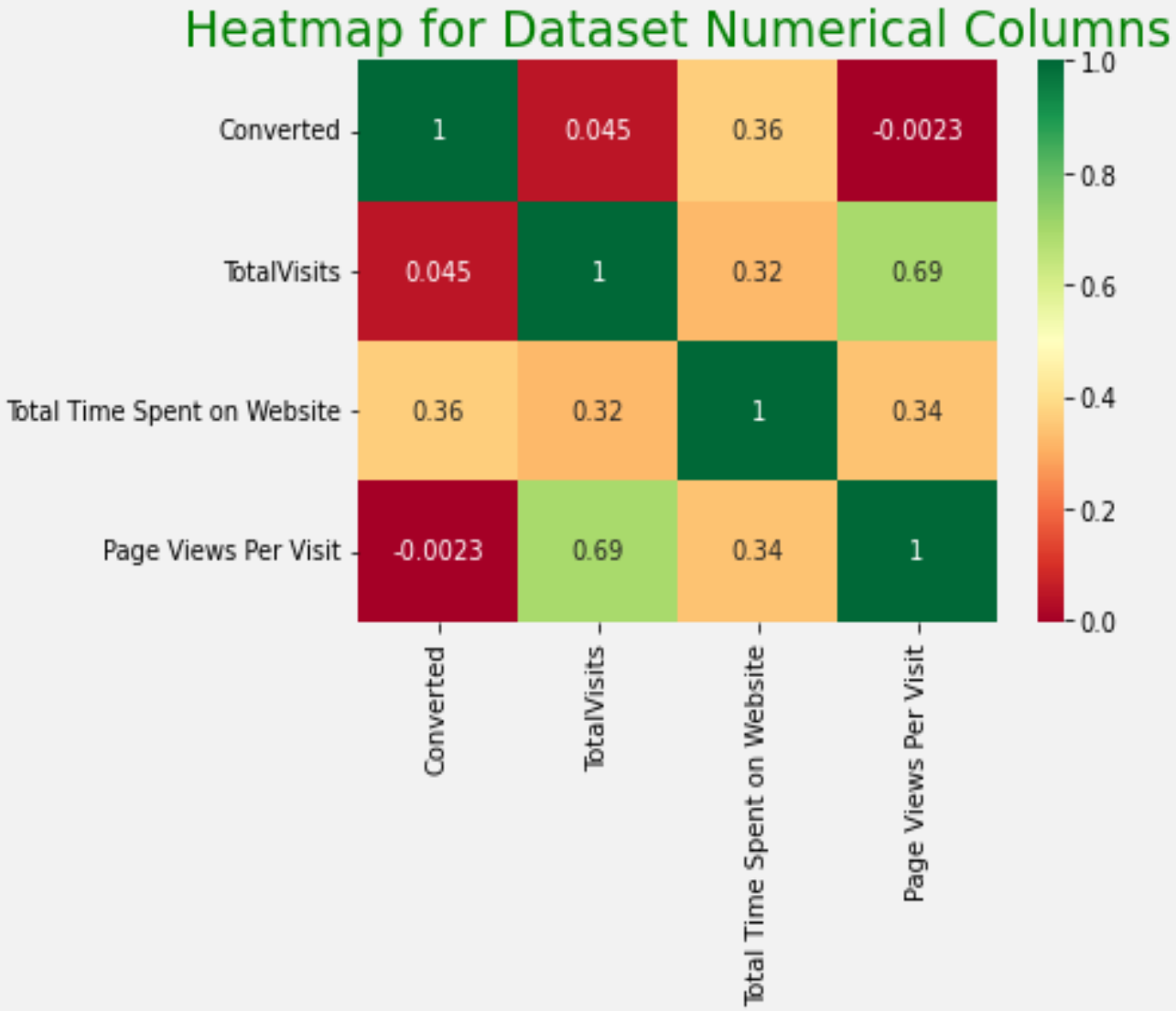


Above graph shows that Recommendations is not a good source for lead conversion.



Search also doesn’t seem to be a good source for lead conversion.

EXPLORATORY DATA ANALYSIS

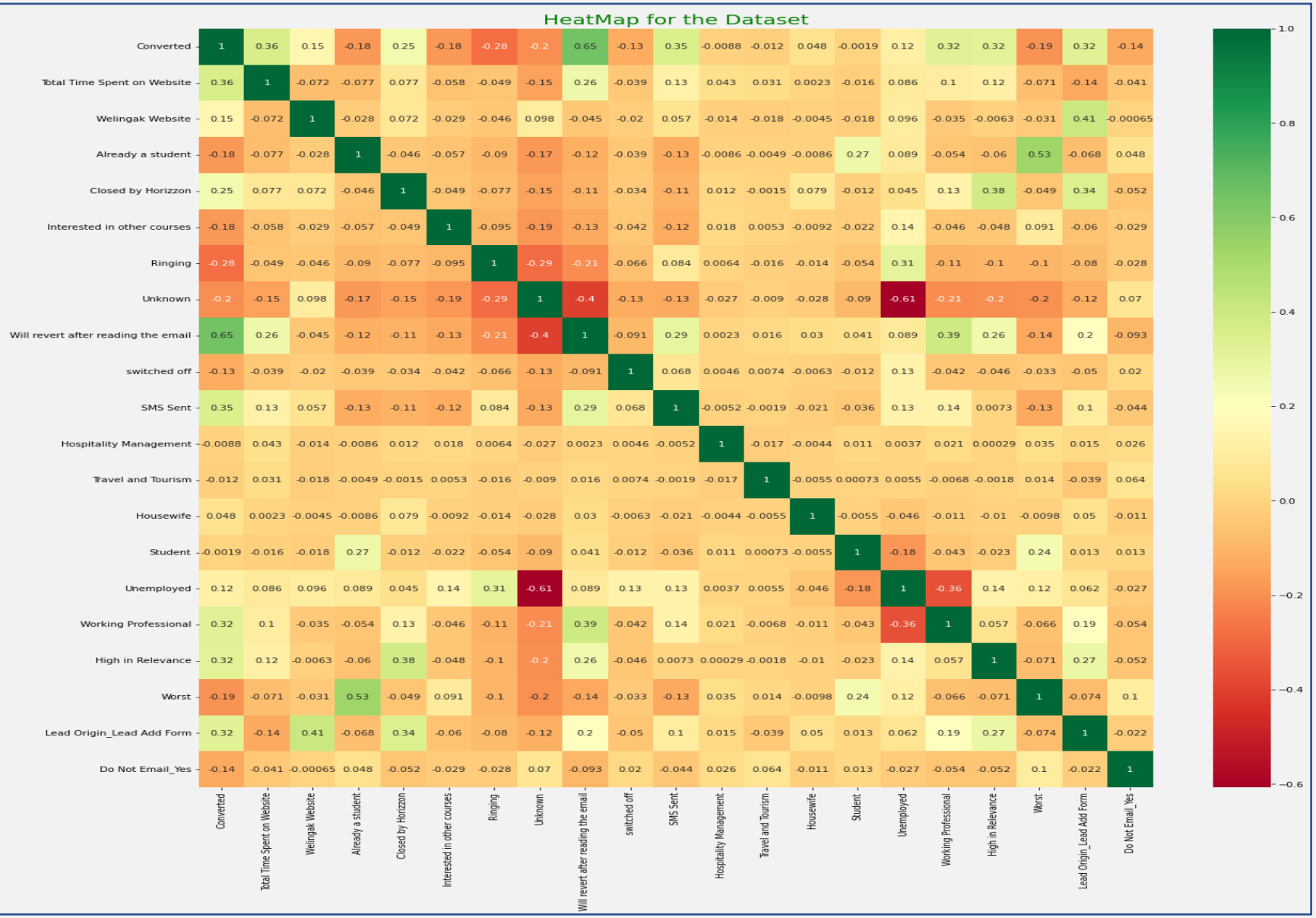


- There is a higher positive correlation between Page views per visit and Total Visit.

MODEL BUILDING

1. Pre-requisite steps:

- Splitting into Train and Test Dataset at 70% and 30%
- Scale Variables in Train set
- Use RFE technique to eliminate the less relevant variables.

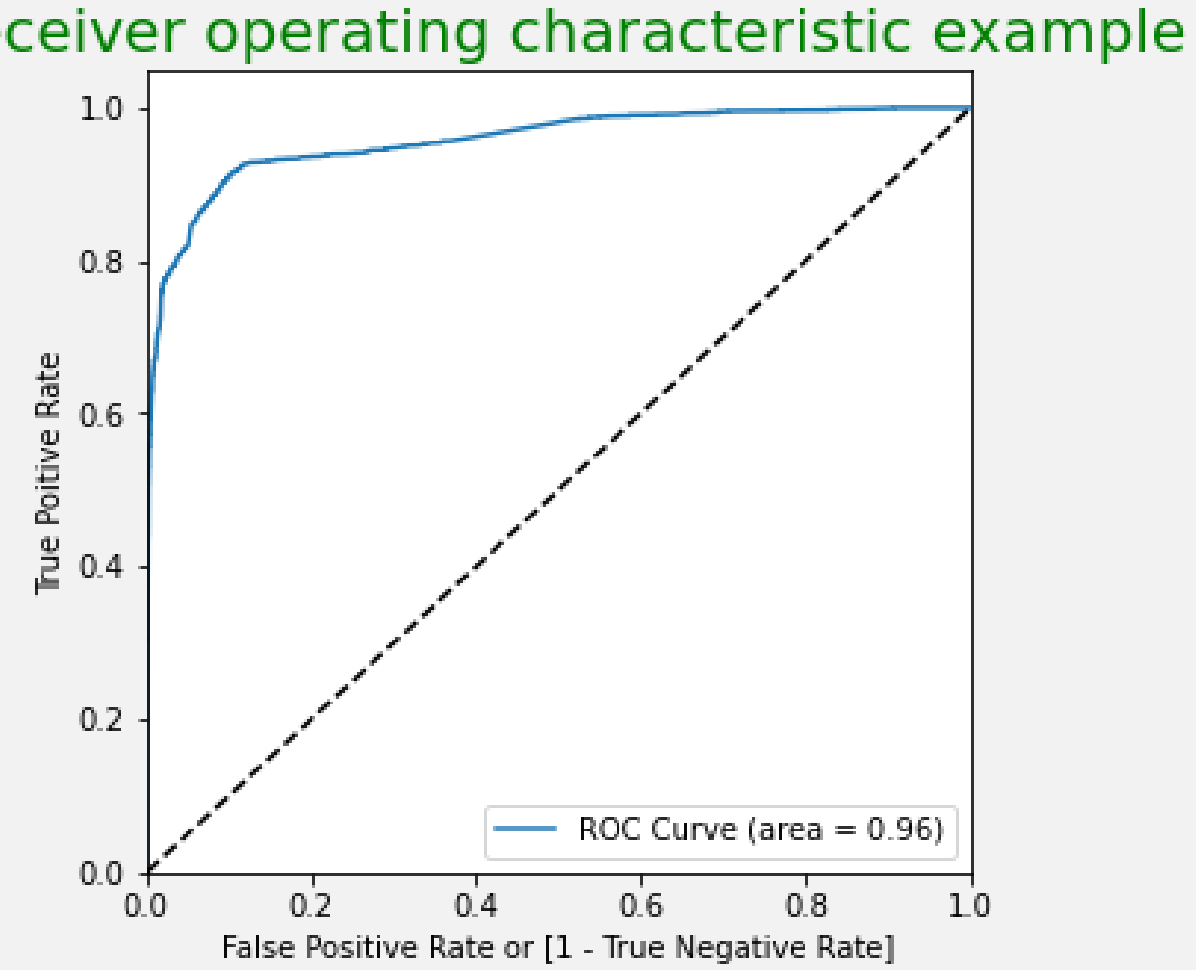


Correlation of the Top 20 features selected by RFE are depicted in the Heatmap

MODEL BUILDING

2. Model Building steps:

- Build the model
- Iterative Elimination of variables based on high p-values and VIF
- Predict using train set
- Above steps are repeated until a stable model was found ROC curve of the model predicted at 0.5 probability is as seen on the right.



The area of the ROC curve indicates the performance of the model. The model value is better as it is close to 1.

MODEL BUILDING

3. Post Model Evaluation steps:

- Accuracy, Sensitivity and Specificity were evaluated to determine the optimal cutoff threshold for the test dataset.
- Precession and recall analysis on test predictions

Accuracy, Sensitivity and Specificity

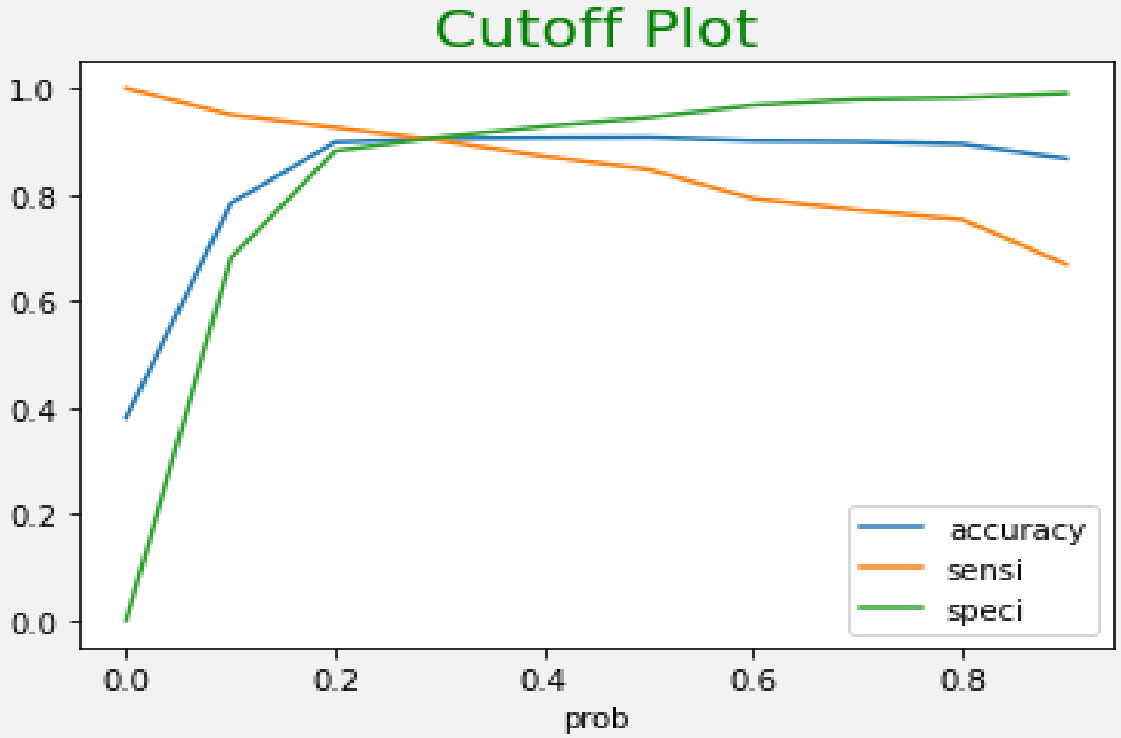
1515	**162**
93	**1002**

- 90.8 % Accuracy
- 92 % Sensitivity
- 90 % Specificity

Precession and Recall

- 86.1 % Pression
- 91.5 % Recall

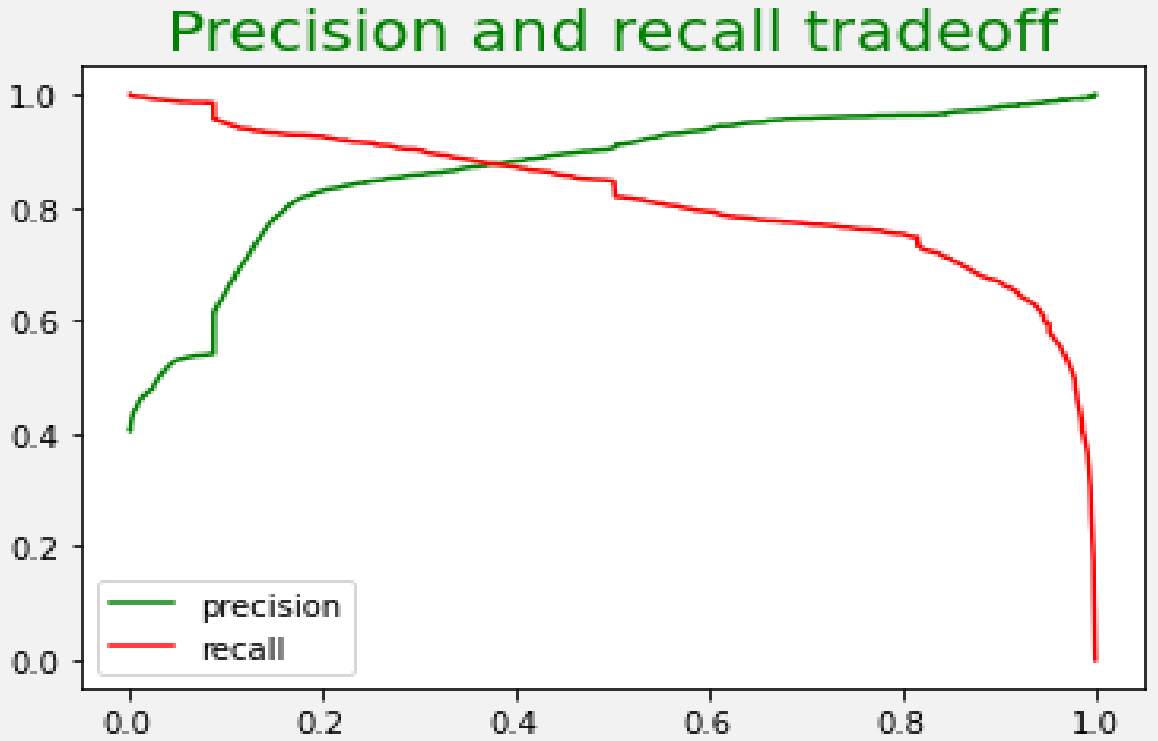
MODEL EVALUATION (TRAIN)



Accuracy, Sensitivity and Specificity

3633	**369**
242	**2224**

- 90.6 % Accuracy
- 90.2 % Sensitivity
- 90.8% Specificity

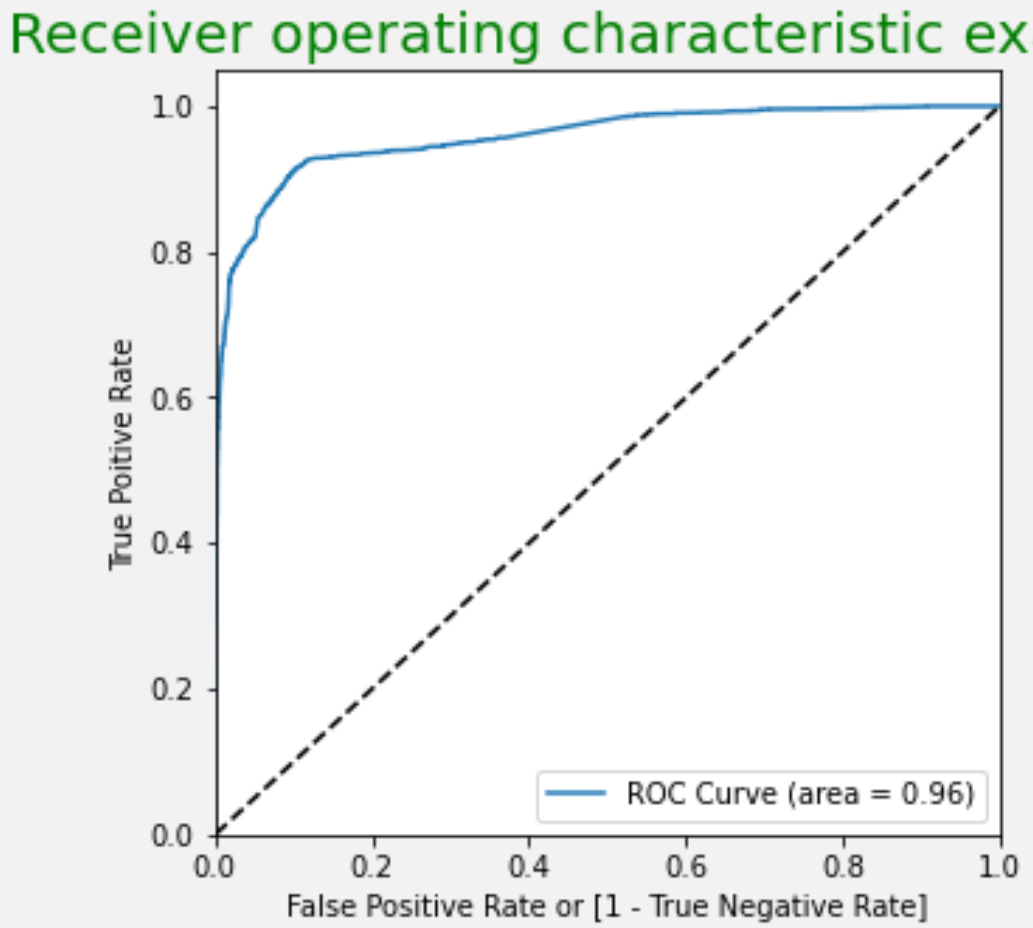


Precession and Recall

- 90.5 % Pression
- 84.7 % Recall

0.37 is the tradeoff between Precision and Recall. However, the threshold was chosen at **0.30** so as to favor recall

MODEL EVALUATION (TEST)



Accuracy, Sensitivity and Specificity

1515	**162**
93	**1002**

- 90.8 % Accuracy
- 92.0 % Sensitivity
- 90.0 % Specificity

Precession and Recall

- 86.1 % Pression
- 91.5 % Recall

CONCLUSION

- As the final model is derived, it can be observed that the probability of conversion will highly be influenced by the leads closed by Horizzon.
- Welingak website seems to be a source for hot leads as it has a strong positive impact on the model.
- Students or potential customers who chose to revert after reading the email were more likely to be converted as leads.
- Leads whose phones were generally switched off or ringing were less likely to be converted as leads.
- It was also anticipated that potential leads who are already students or interested in other courses had a lesser chance of becoming strong leads.
- In general, the Tags column seems to be an important column as most of the variables from the model are converted to dummy variables with high coefficient magnitudes.
- Using this model, the sales team of X education can easily filter models using the lead score for identifying and converting high-potential leads.

RECOMMENDATION

- As Google Search is a preferred method of Marketing the conversion rate is not as high as Welingak website. Hence, apart from using optimum SEO options for Google, it is also imperative that the X Education company needs to use key words for their page to be more searchable and generate more traffic to their website.
- Unemployed forms the highest volume in leads, and could be targeted for better conversion. However, the most focus area for the company is to concentrate more on the working professionals as they are usually converted.
- The company also needs to focus more on their page content so that it becomes attractive for the users to explore more options and increase their duration of visit on their website.
- Digital advertisement doesn’t seems to contribute to promising leads, should company should focus more on the same and make attractive offerings and user friendly webpage content to boost the users conversion rate.