Lead Scoring Case Study

Increasing Lead conversion rate for X education

By,

Shruti Badiyal

Bhagyashree Shinde

Bhagyashree Awatagiri

Problem Statement

- 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. 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. The typical lead conversion rate at X education is around 30%.
- Now, although X Education gets a lot of leads, its lead conversion rate is very poor

STRATEGY

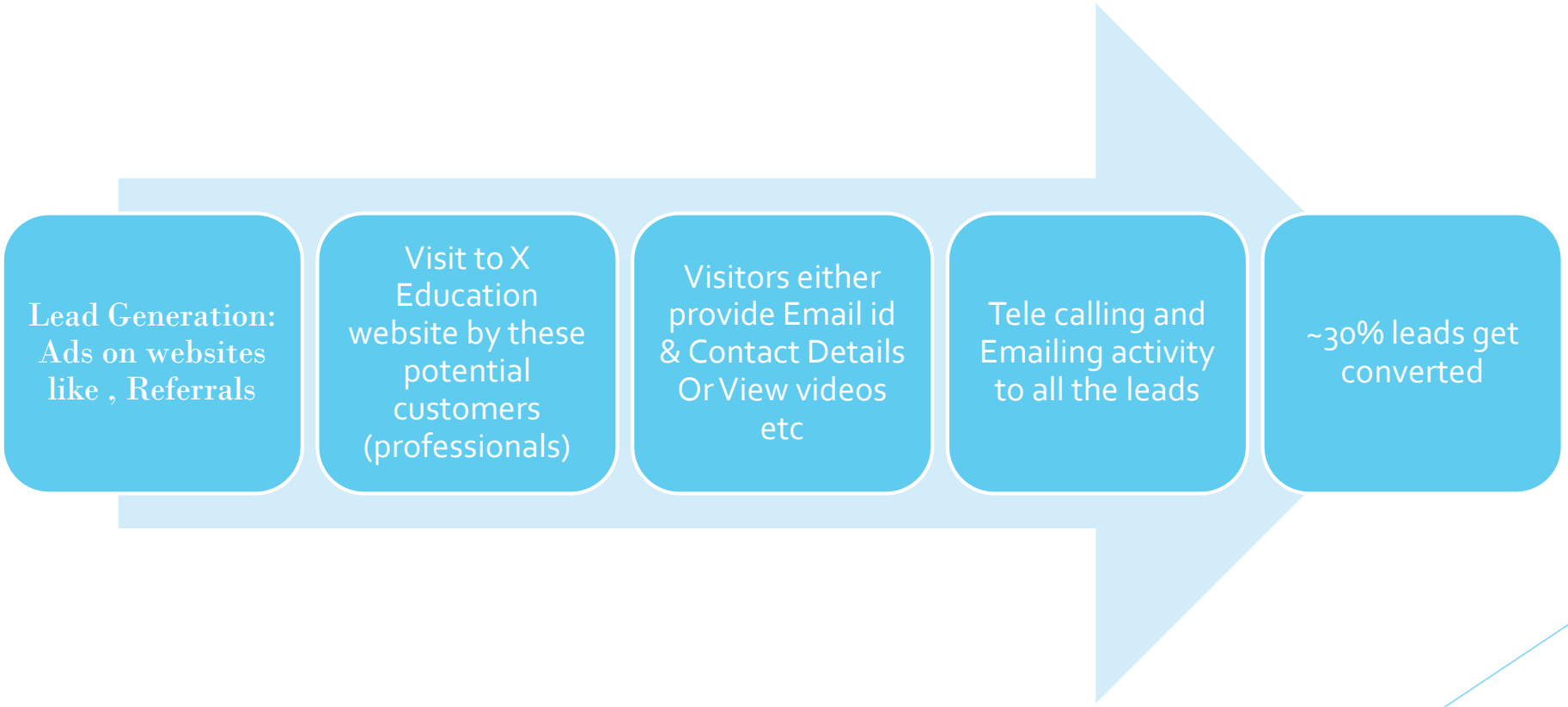
- Import the data
- Reading and understanding data
- Data cleaning and preparation for the analysis
- Preparing the data for modelling
- Dummy Variable creation
- Train-Test split
- Rescaling the variables
- Model building
- Model evaluation
- Prediction on the test se

Goals

There are quite a few goals for this case study:

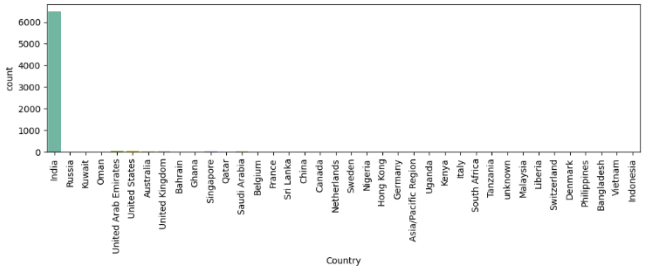
- Build a logistic regression model to assign a lead score between 0 and 100 to each of the leads which can be used by the company to target potential leads. A higher score would mean that the lead is hot, i.e. is most likely to convert whereas a lower score would mean that the lead is cold and will mostly not get converted.
- There are some more problems presented by the company which your model should be able to adjust to if the company's requirement changes in the future so you will need to handle these as well. These problems are provided in a separate doc file. Please fill it based on the logistic regression model you got in the first step. Also, make sure you include this in your final PPT where you'll make recommendations.

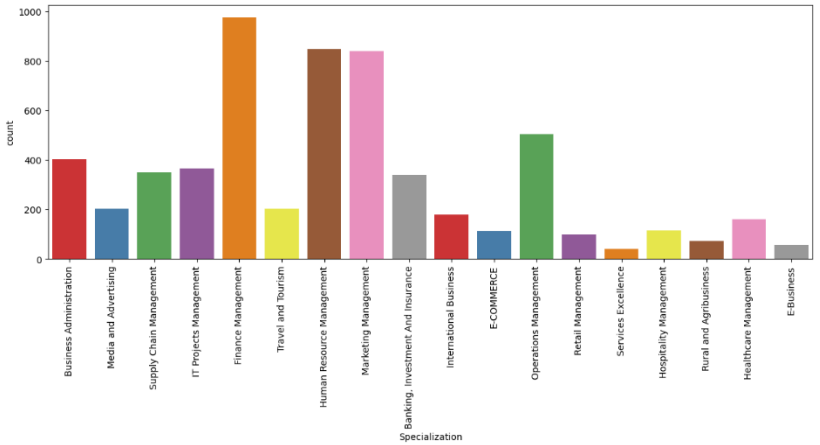
Lead – Conversion Process

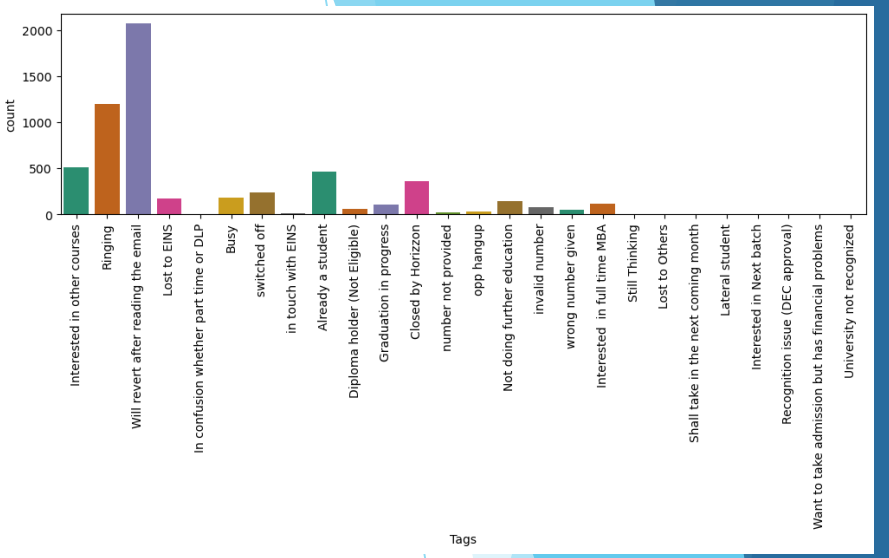


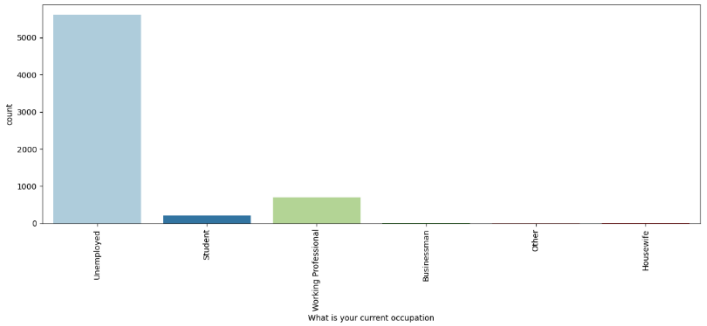
Exploratory Data Analysis

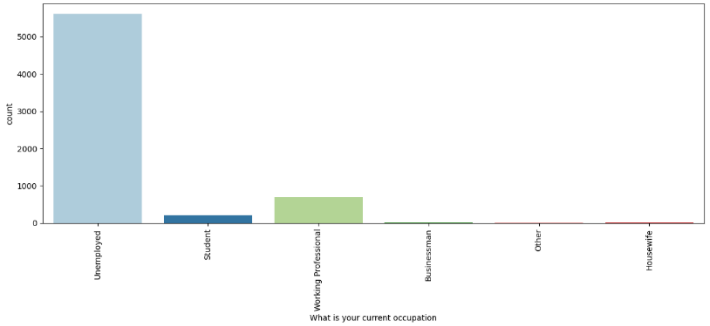
Data Visualizations

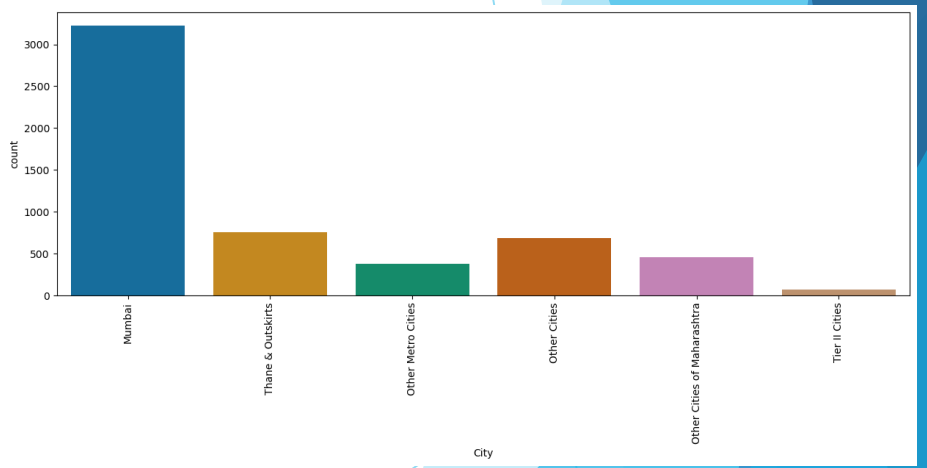












Dealing with missing values

Viewing the Specialization graph

There are numerous management courses available with a strong conversion rate; it would be advantageous to combine them into one entity for future study. Moreover, a few business courses can be merge

Viewing the Tags graph

There are numerous management courses available with a strong conversion rate; it would be advantageous to combine them
into one entity for future study. Moreover, a few business courses can be merged.

Observing the Lead Source graph

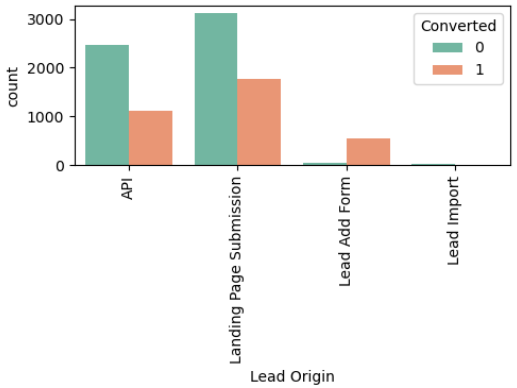
There are many categories which add no value to the conversion/non-conversion of the leads. Combining them as one entity would help
in further analysis while creating dummy variables.

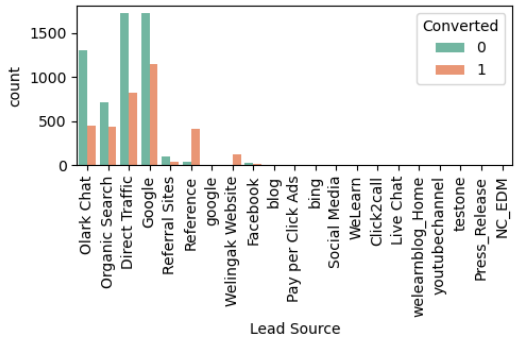
Observing the Last Activity

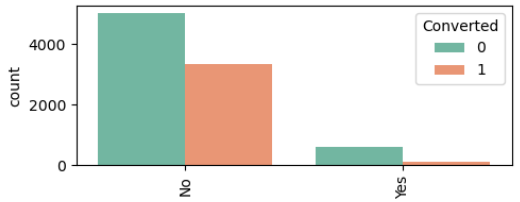
Combining the many Last Activity categories

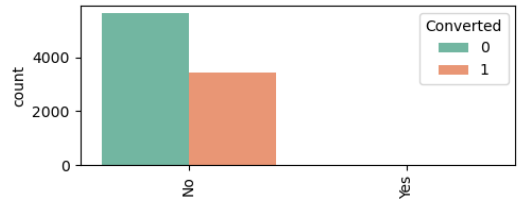
Observing Last Notable Activity graph

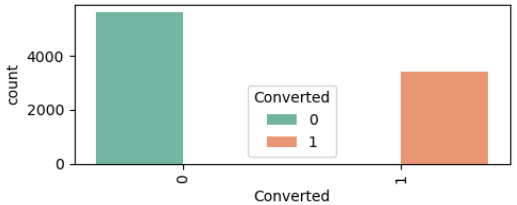
There are many Last Notable Activity categories which do not add any value to our analysis, so combining them will benefit
our analysis further

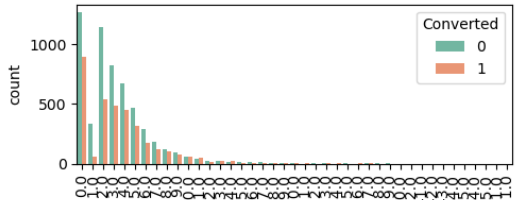


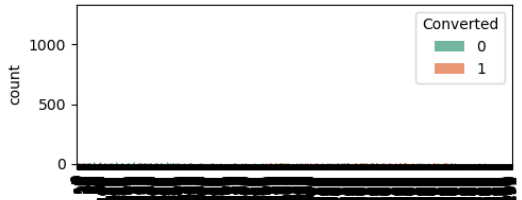


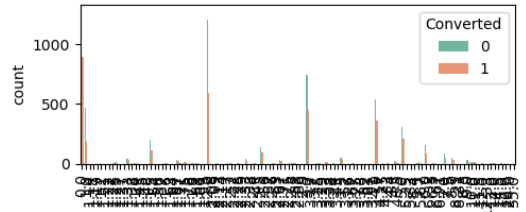


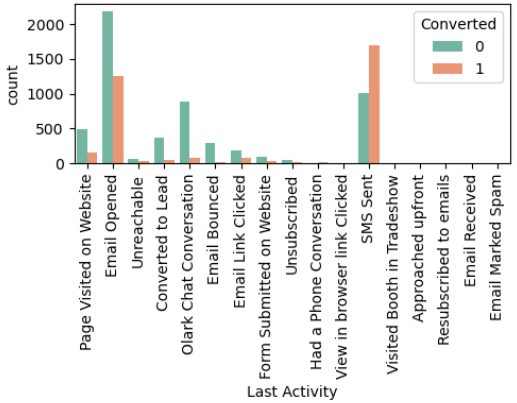


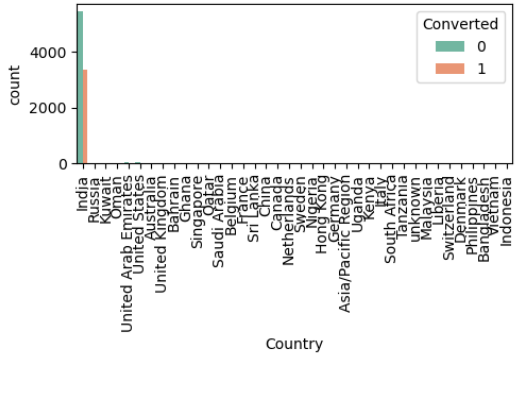


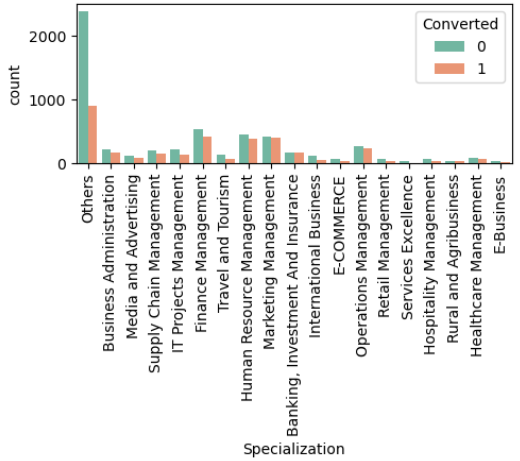


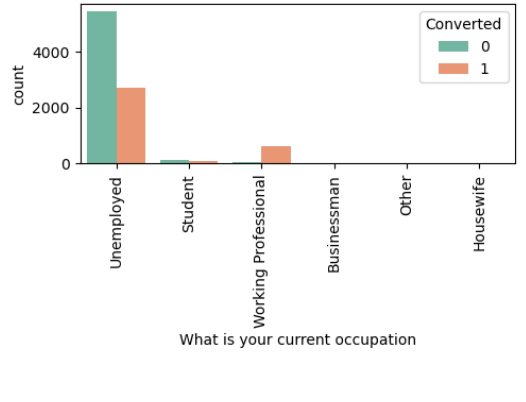










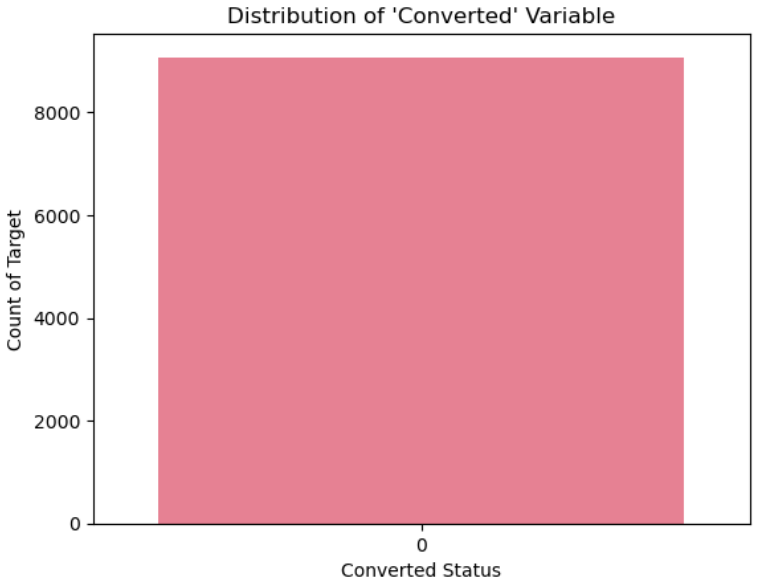


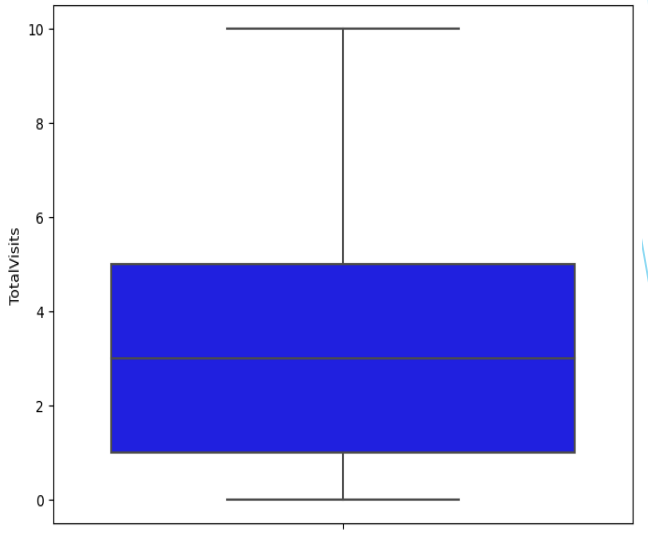
Dealing with categorical Variables

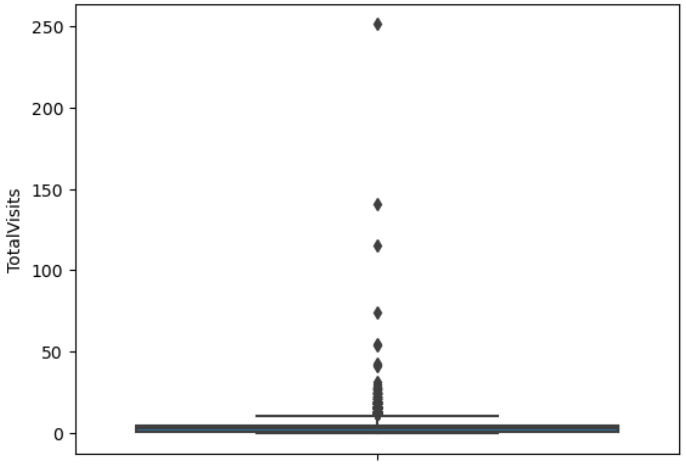
VISUALIZING CATEGORICAL VARIABLES

- Lead Origin [Lead Add Form, Landing Page Submission] have a high conversion rate.
- If the Lead Source is from Google, Direct Traffic, Reference they seem to be a potential lead with high conversion rates.
- Customers who choose 'Do Not Email' seem to have a higher conversion than the customers who opt for Email.
- If the Last Activity is SMS Sent or Email Opened they have a higher chance of being converted.
- Customers browsing for various management courses seem to be converted.
- Tags with 'Will revert after reading the email' are highest converted followed by Closed by Horizon tags.
- High number of Working Professionals and Unemployed leads get converted.
- Leads who don't opt for 'A free copy of mastering The Interview' seem to convert more than who opt for it.
- If the Last Notable Activity is SMS Sent or Email Opened then they have a higher chance of being converted

Bivariate & Univariate Analysis

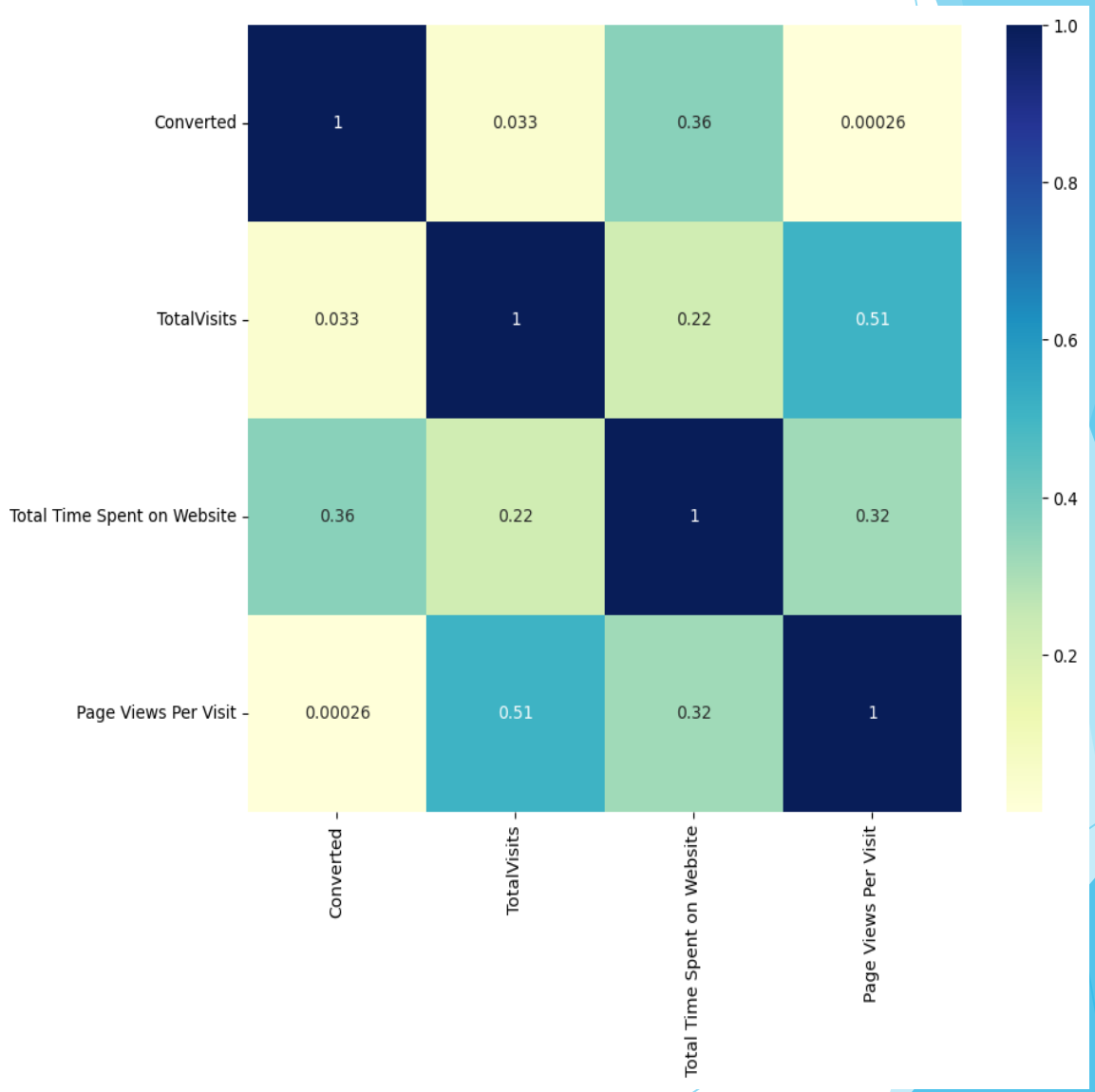




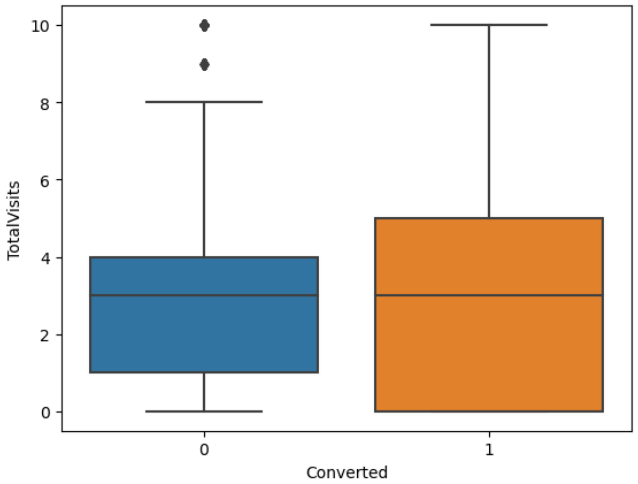


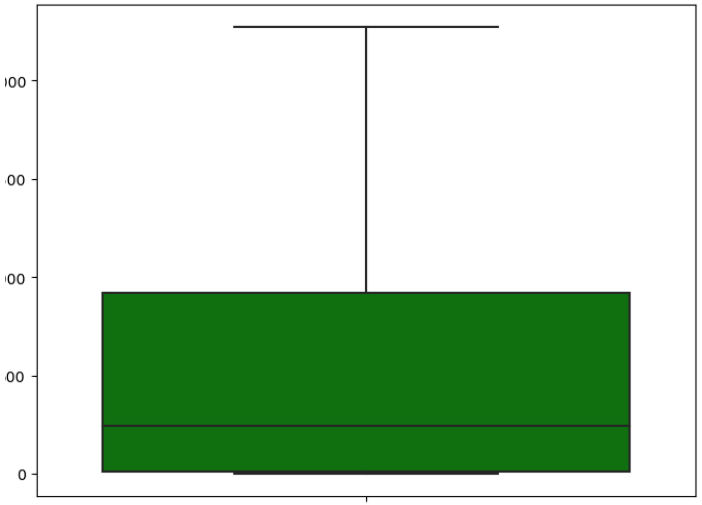
VISUALIZING THE NUMERICAL VARIBALES

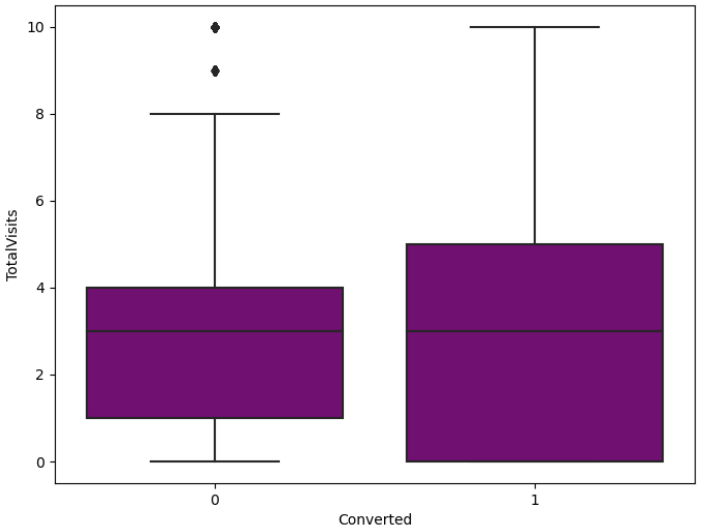
- Page Views Per Visit and Total Visits appear to be slightly correlated.
- Overall Time Spent on Website Seems Proportional to the
- Conversion.
- The linear association between these variables is not particularly strong.

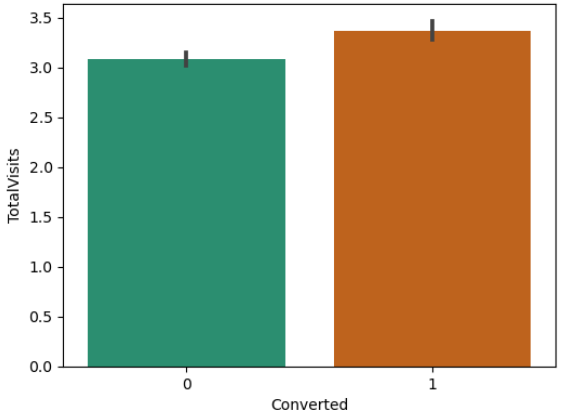


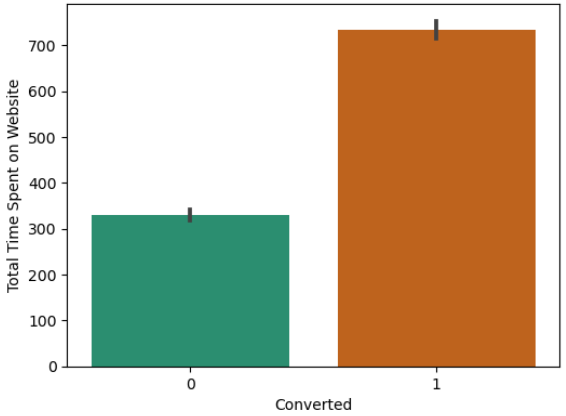
Heat map for numerical variables

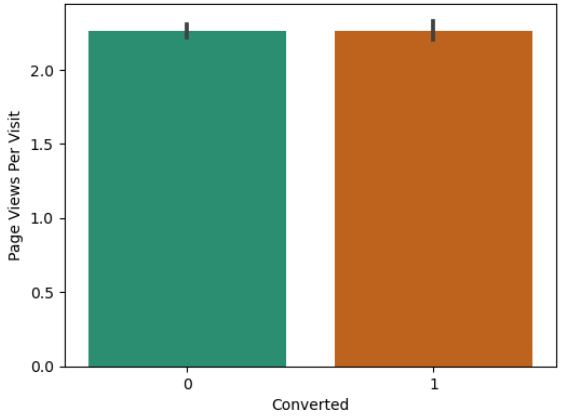

TotalVisits
Converted
0
1

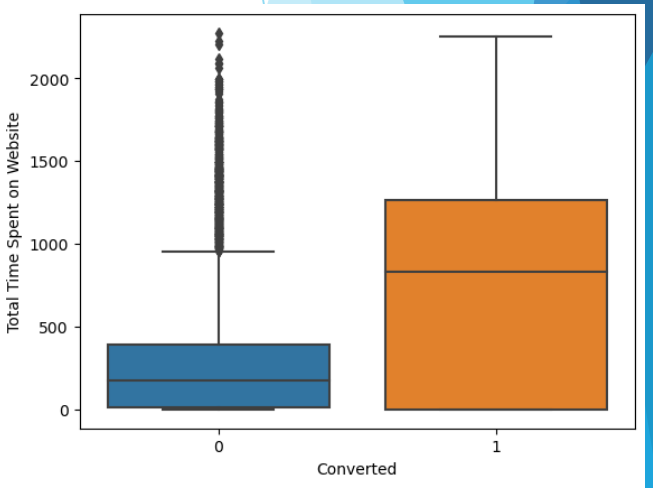



TotalVisits
Converted
0
1

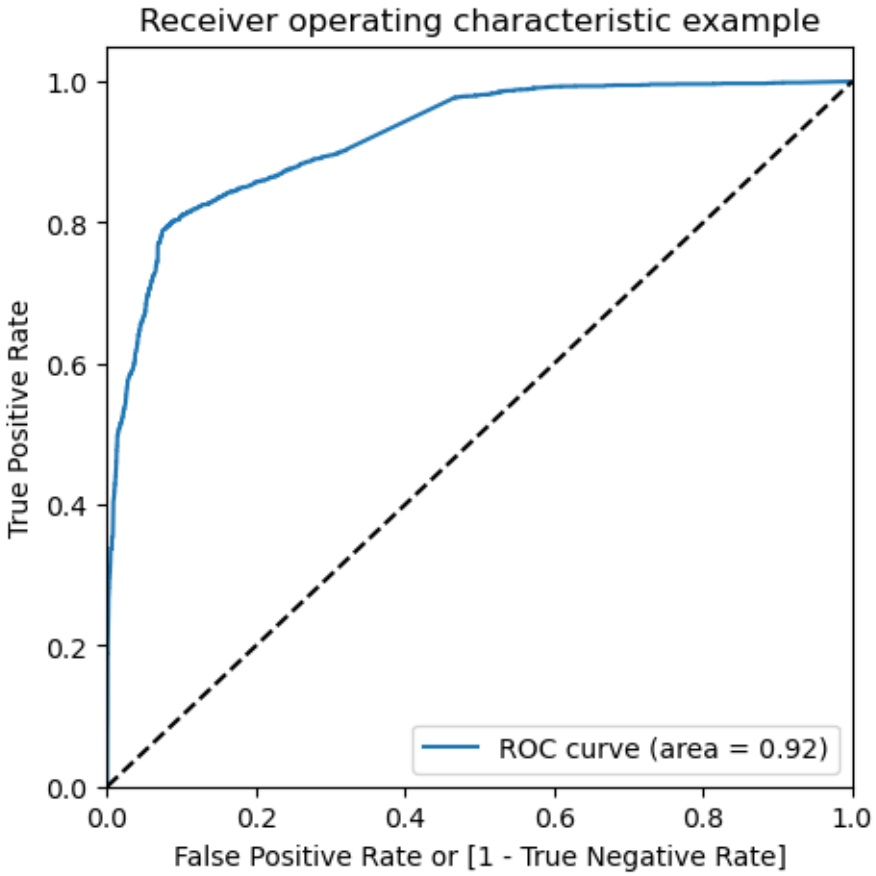

TotalVisits
Converted
0
1


Total Time Spent on Website
Converted
0
1

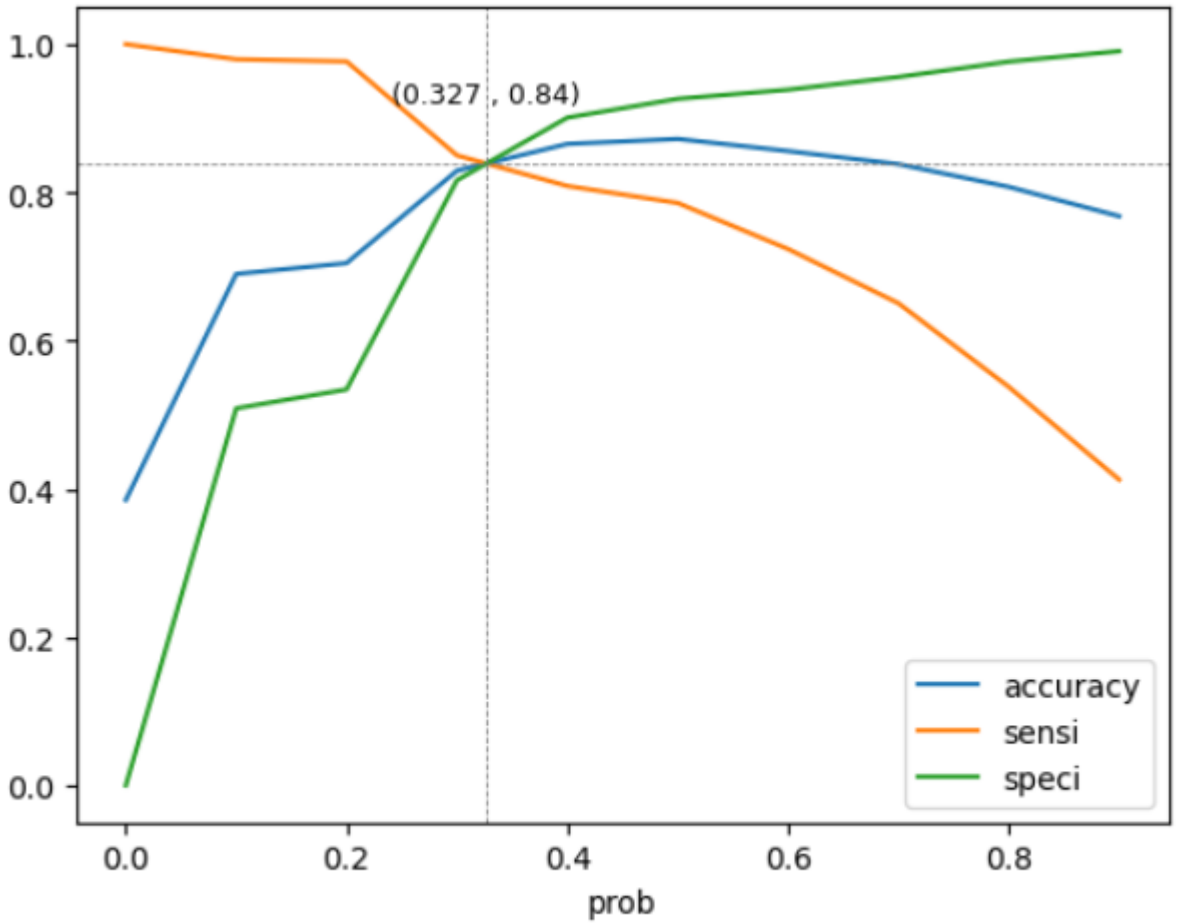

Page Views Per Visit
Converted
0
1


Total Time Spent on Website
Converted
0
1

Model building and Evaluation



The area under the ROC curve is 0.92. Hence, this is a good model.



From the above graph a cut off point is 0.327

Conclusion

- These results are similar to the results obtained on the training data, which suggests that the model is performing well.
- The important features that contribute to the probability of a lead being converted are:
  - Lead Origin: Lead Add Form
  - What is your current occupation: Working Professional
- Total Time Spent on Website:
  - These features can be used to target marketing campaigns more effectively. For example, if a company is trying to sell a product or service to working professionals, they may want to focus their marketing efforts on leads that have those characteristics.

Here are some additional insights that can be gained from this information:

- The model is more likely to correctly predict that a lead will convert if the lead came from the Lead Add Form.
- The model is more likely to correctly predict that a lead will convert if the lead is a working professional.
- The model is more likely to correctly predict that a lead will convert if the lead spent a longer time on the website.
- This information can be used to improve the conversion rate of a company's marketing campaigns.

Final Observations

Train Data:

- Accuracy : 84.29%
- Sensitivity :83.69%
- Specificity : 84.66%¶

Test data

- Accuracy: 82.51%
- Sensitivity: 85.44%
- Specificity: 80.85%

THANK YOU