Lead Score Case Study

Logistic Regression Assignment

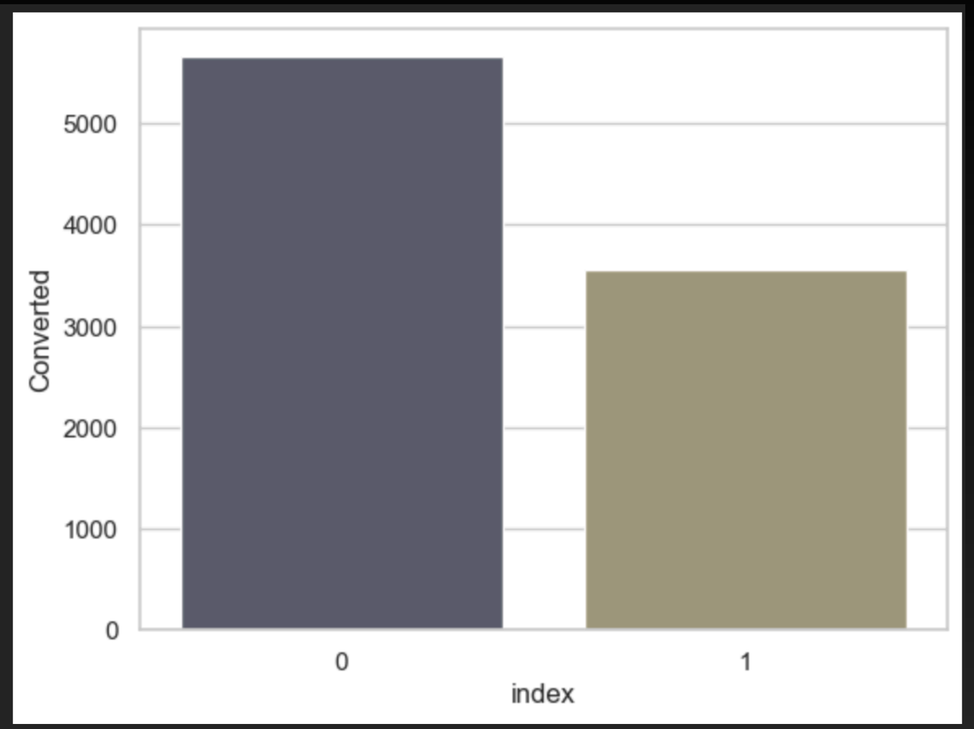
17.04.2023

Inference

- Target variable -‘Converted’
- All other columns are feature variables.
- Task of the analysis is to find a meaningful relationship between feature variables and target variable using Logistic Regression

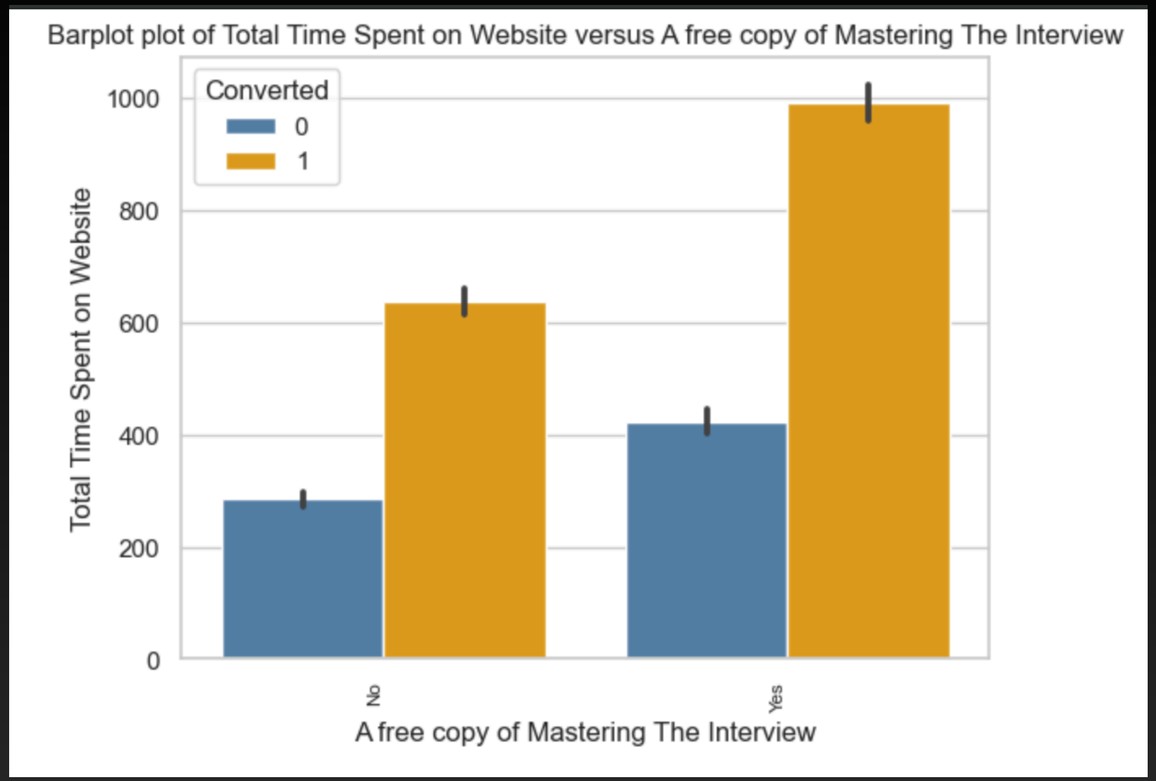
Point to deduce

- Out of the total number of leads, more than 5500 are not converted.
- Dataset is slightly imbalanced



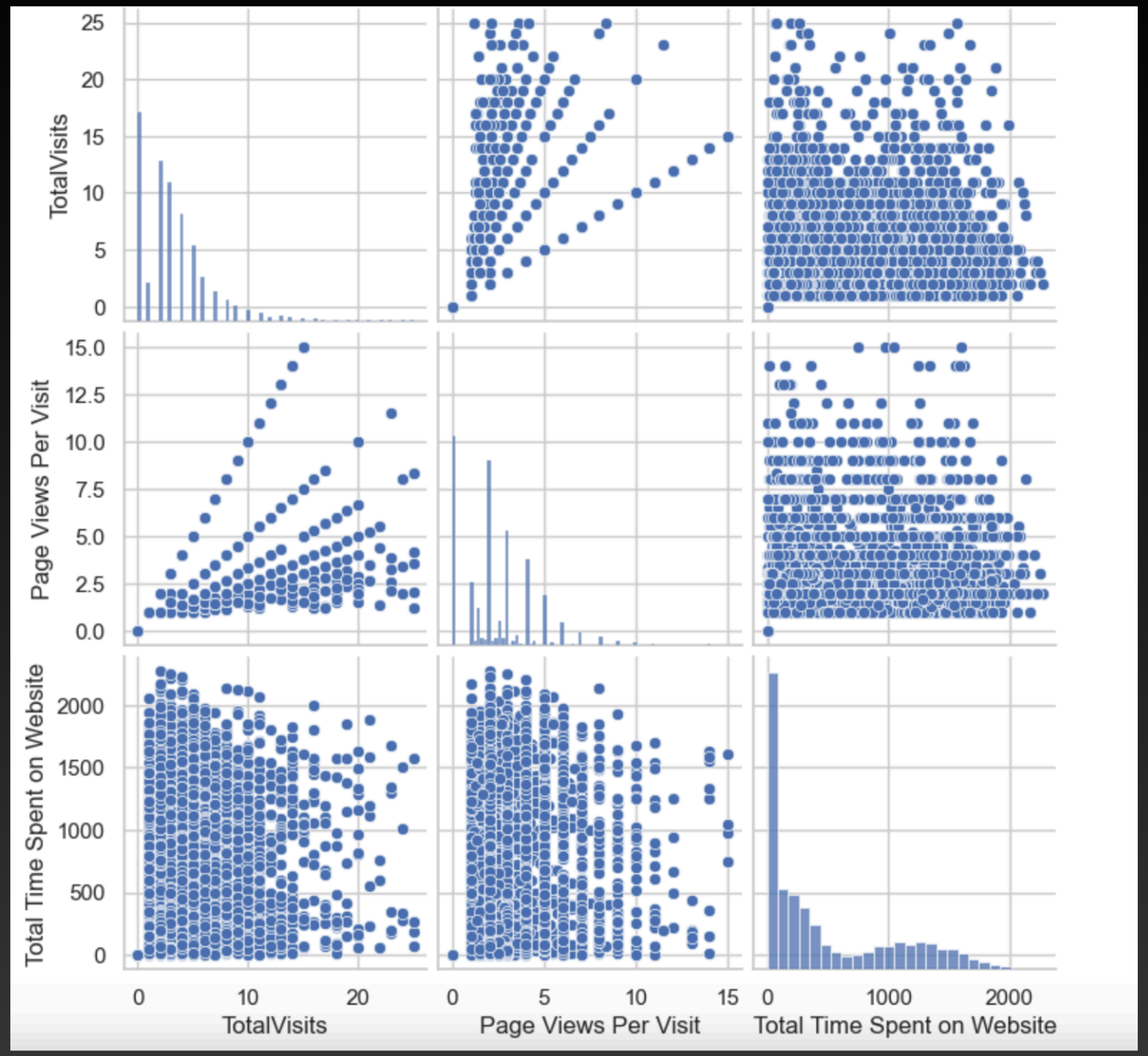
Point to deduce

- The target audience is people who spend more time exploring the website and downloading free guides.



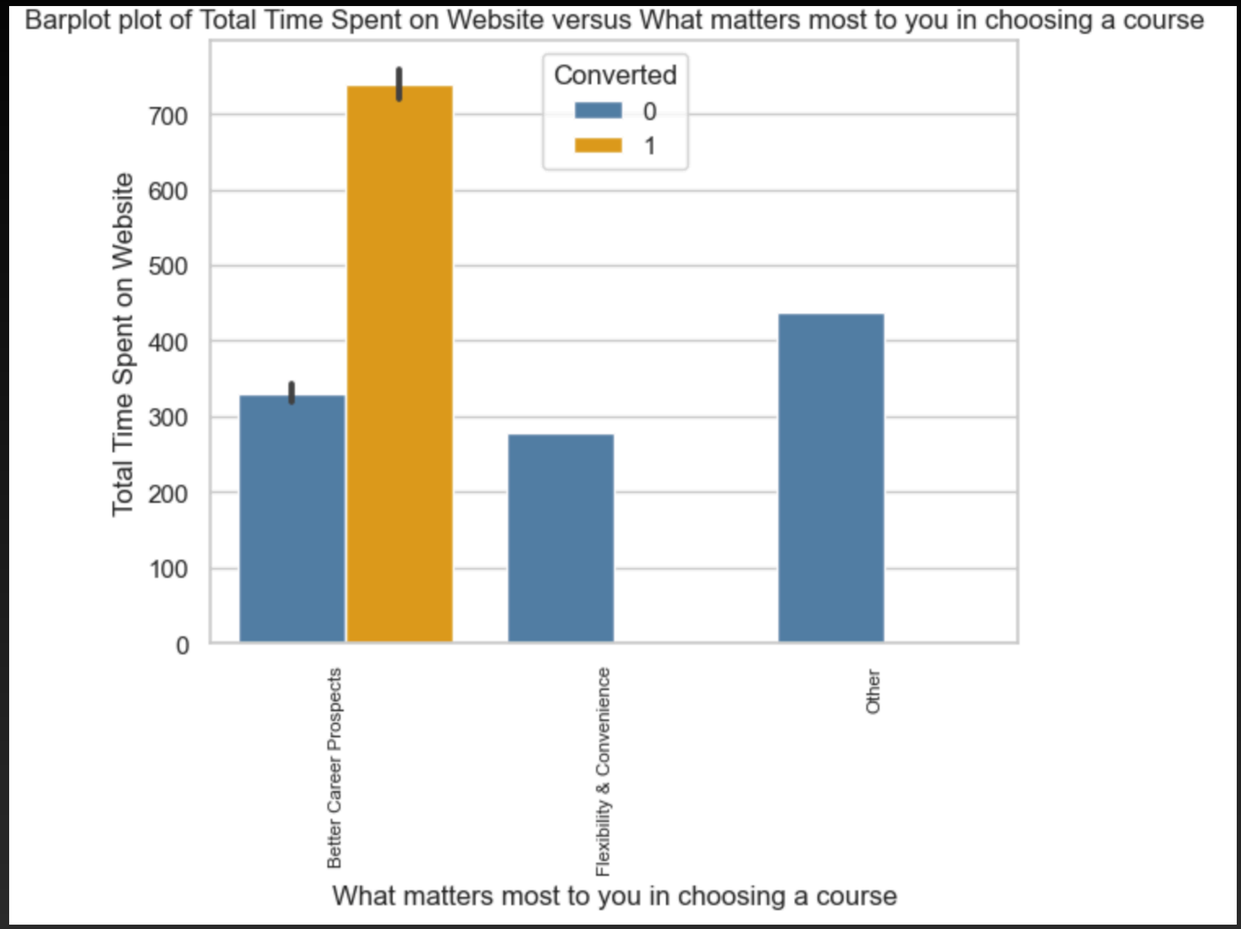
Point to deduce

- There is a linear relationship between columns 'Total Visits' and 'Page Views Per Visit'.



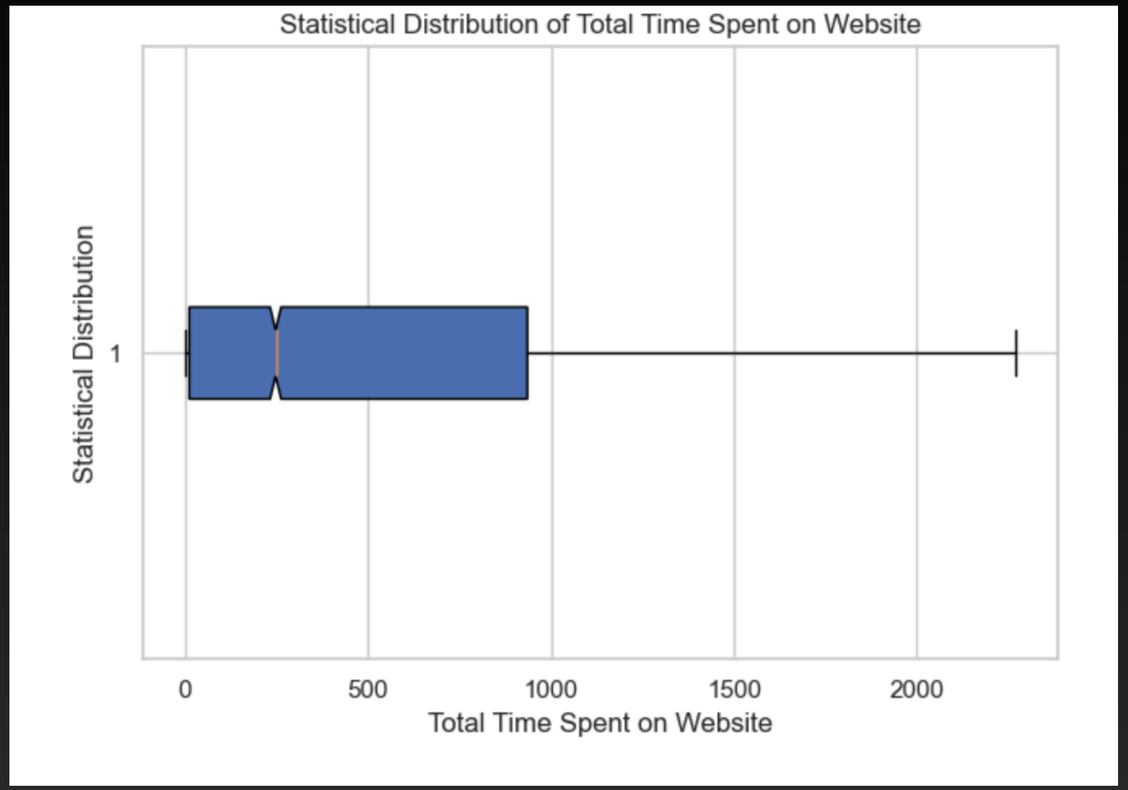
Point to deduce

- Most leads that spend more time on the website and are looking for better job opportunities are more likely to get converted.
- This group should be the educational company's primary emphasis.



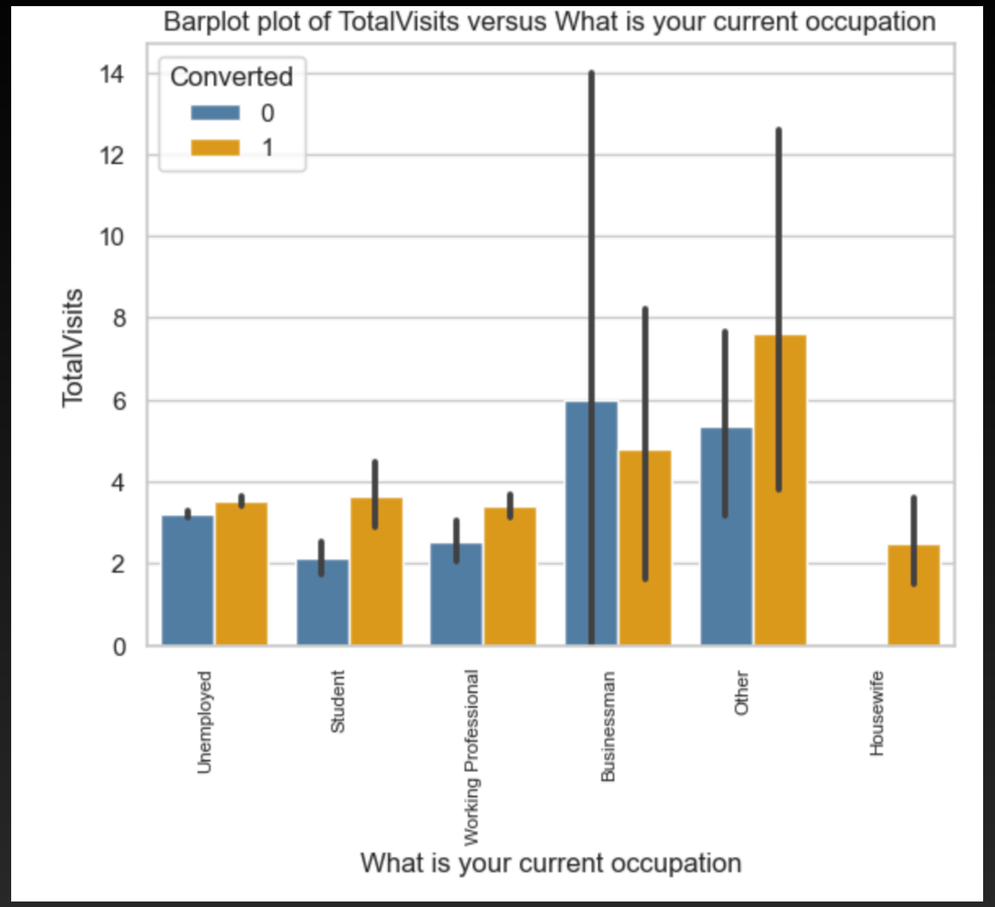
Point to deduce

- On an average, the total time spent by people exploring the website is around 250-350.



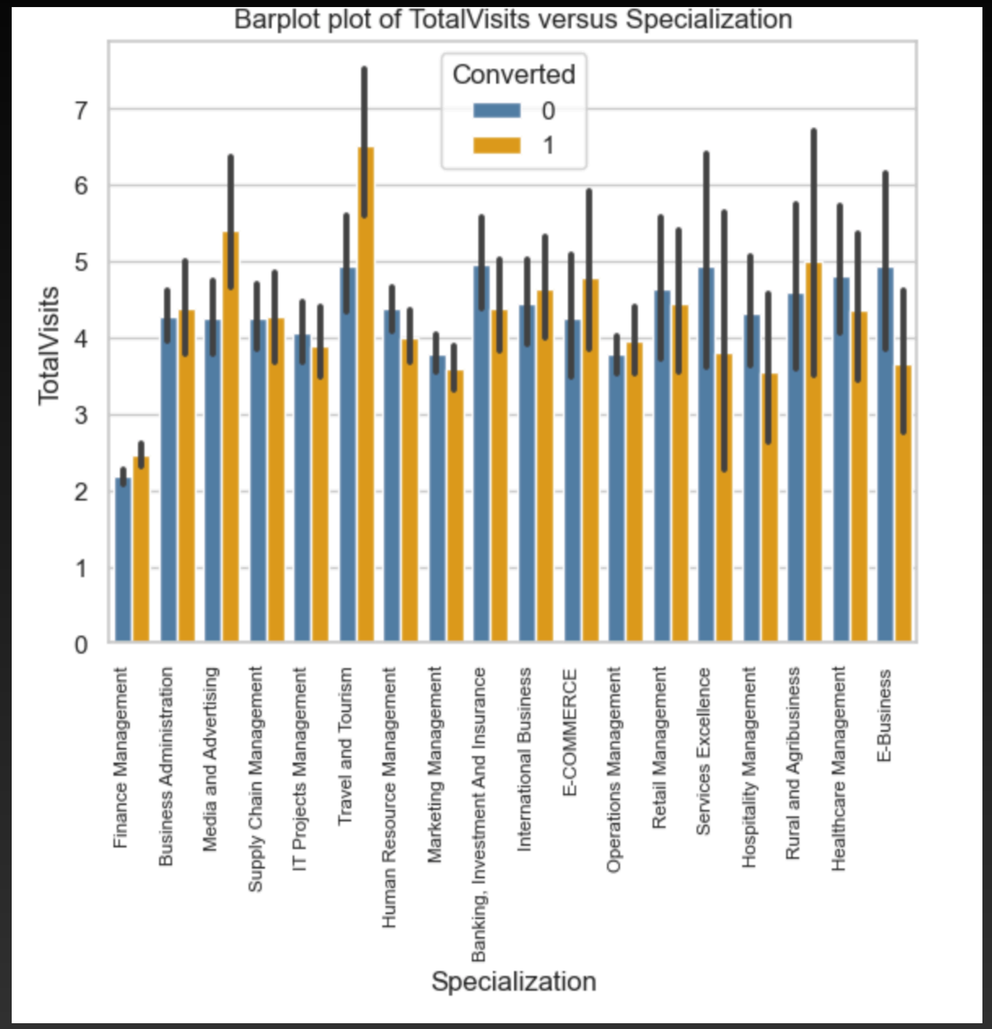
Point to deduce

- The conversion rate among students and working professionals is low.It can be due to the poorly developed programme curriculum or the tediousness of filling out forms.
- The likelihood that housewives seeking a second inning will be converted is higher.



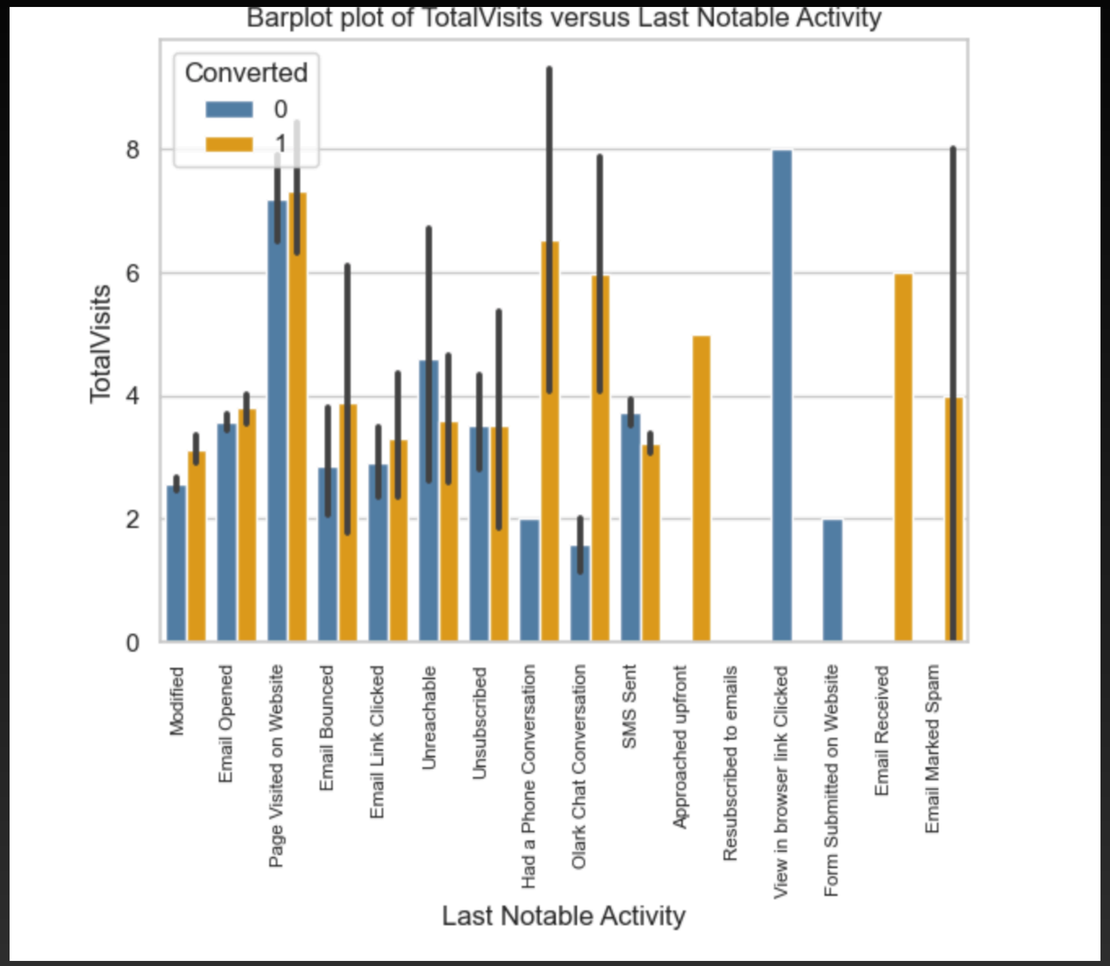
Point to deduce

- People from "Travel and Tourism" and "Media and Advertising" have better conversion rates.It can be as a result of the curriculum being in line with their area of specialisation.
- For individuals falling under specialisations, such as "IT Project/ Operations/HR Management,"there is a slight difference between the conversion and non-conversion rates.Feedback and surveys completed prior to leaving the page can assist us in understanding the issues and problems.



Point to deduce

- The conversion rate of leads who are engaged in phone or online chat conversations is greater.This suggests that assigning a lead partner to every lead can aid in boosting conversion.

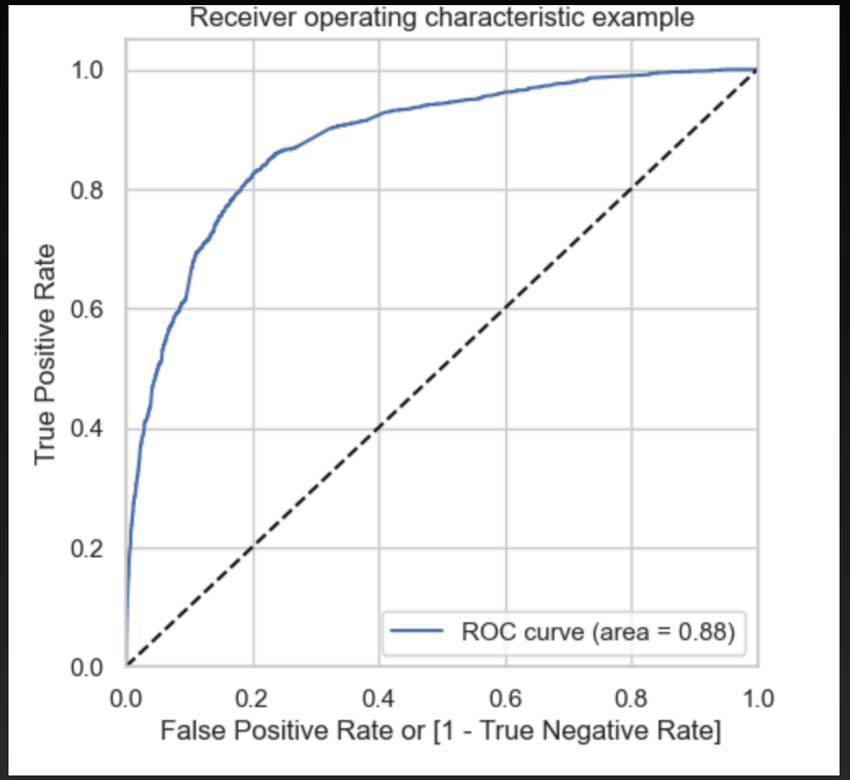


Point to deduce

- A value of -1 indicates strong negative correlation
- A value of +1 indicates strong positive correlation
- A value near 0 indicates weak correlation
- Eg-‘TotalVisits‘ with a correlation value of 0.69 can be said to be correlated with feature ‘Page Views Per Visit’

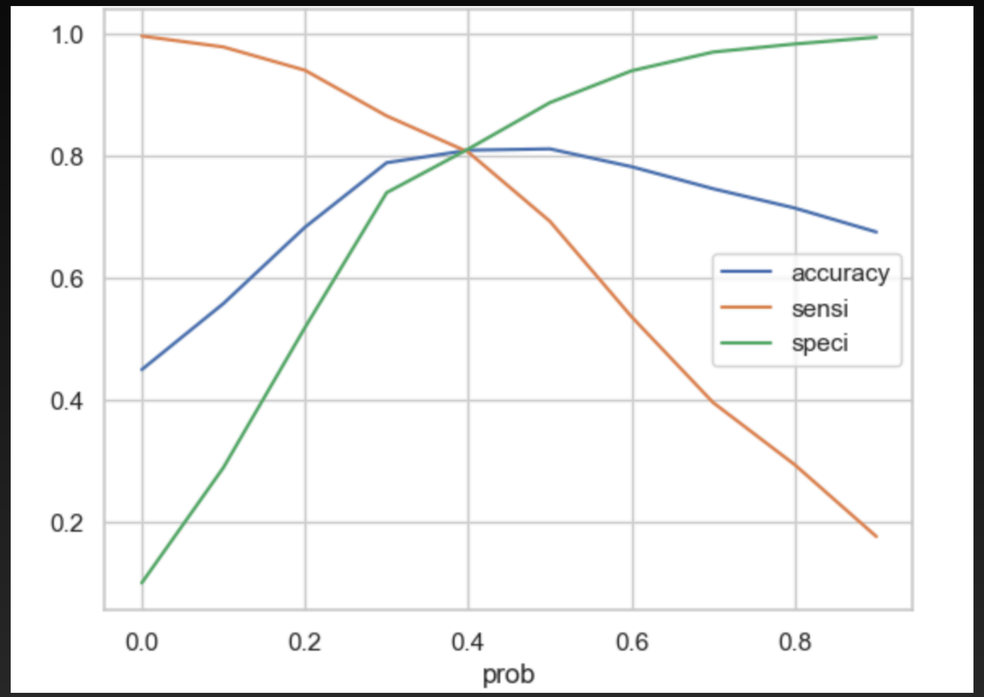
Point to deduce

- The trade-off between True Positive Rate and False Positive Rate, which is essentially a trade-off between Sensitivity and Specificity, is depicted by the ROC curve.
- By determining the area under the curve (AUC) of an ROC curve, you can determine how good the model is.A value of 0.88 indicates a decent model.



Point to deduce

- As the chart illustrates, Sensitivity rises when Specificity falls and vice versa.
- Optimal cut off point is where Accuracy =Specificity=Sensitivity i.e 0.4 in this case.



Conclusion

- Pay attention to leads that have been converted. Engage leads in question-and-answer sessions to gather the pertinent data you require about them.
- Release feedback forms and surveys to those who opted out despite speaking with the lead partner or mentor.This can help to fairly assess their worries/concerns.
- From your prospect leads, sort the best candidates first. The three metrics that have the biggest effects on how likely it is that a lead will be converted are "Total Visits," "Total Time Spent on Website," and "Page Views Per Visit." Then, have a list of leads handy so you can alert them to new initiatives, services, employment opportunities, and impending further education.