LEAD SCORE CASE STUDY

LOGISTIC REGRESSION

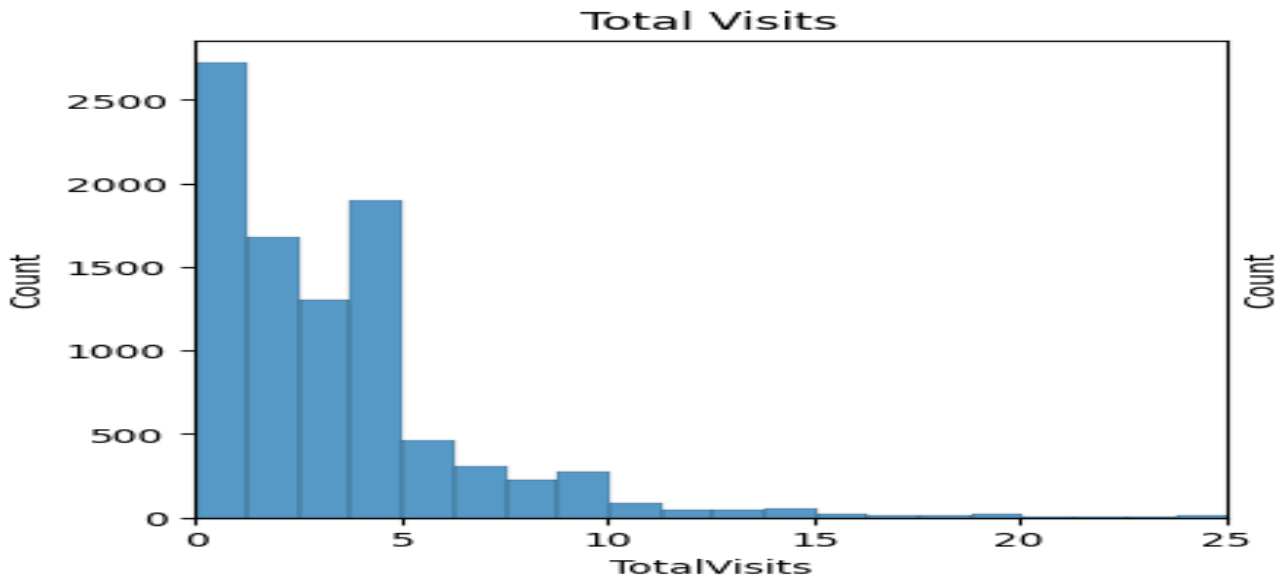
AGENDA:

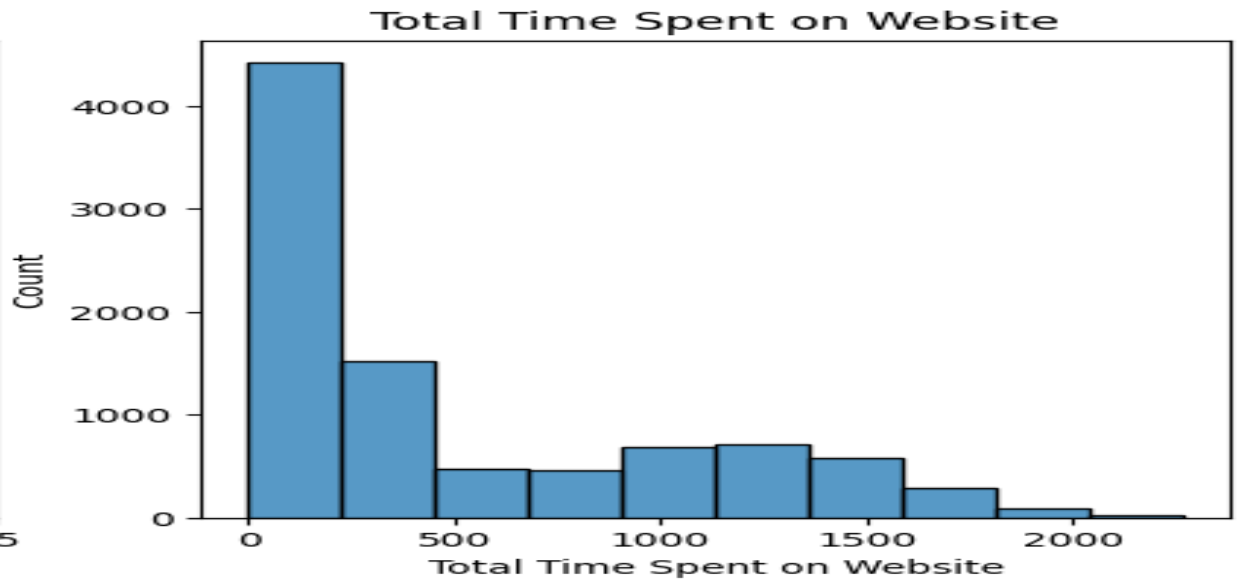
The purpose is to optimize the lead scoring mechanism based on their fit, demographics, behaviors, buying tendency etc. By implementing explicit & implicit lead scoring modelling with lead point systems.

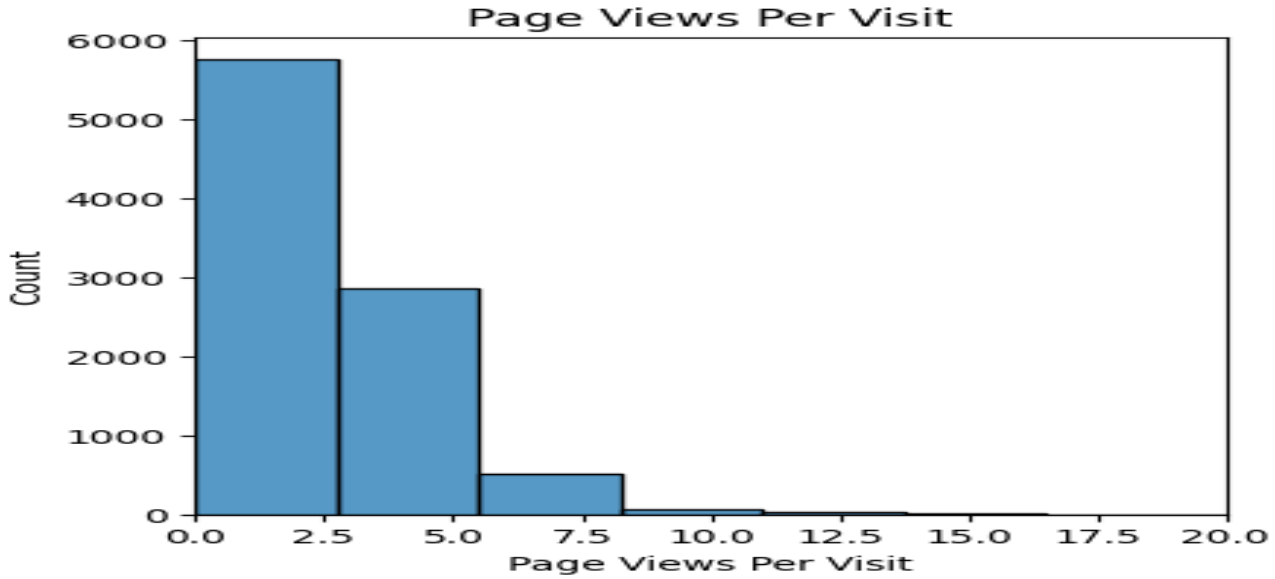
PROBLEM STATEMENT:

- 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. For example, if, say, they acquire 100 leads in a day, only about 30 of them are converted. To make this process more efficient, the company wishes to identify the most potential leads, also known as ‘Hot Leads’. If they successfully identify this set of leads, the lead conversion rate should go up as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone.

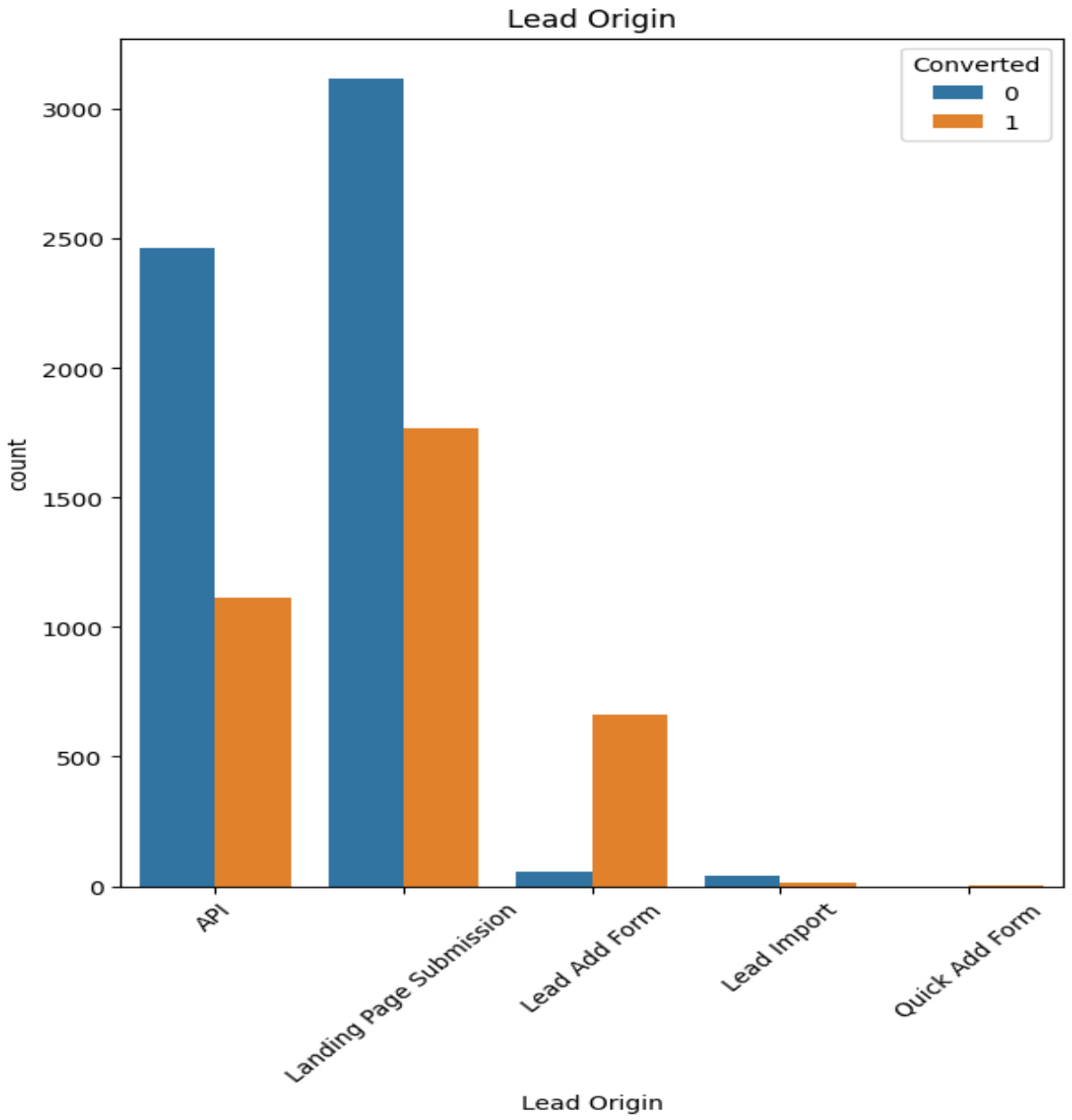
SUBPLOT FOR TOTAL VISITS, TOTAL TIME SPENT ON WEBSITE & PAGE VIEWS PER VISIT

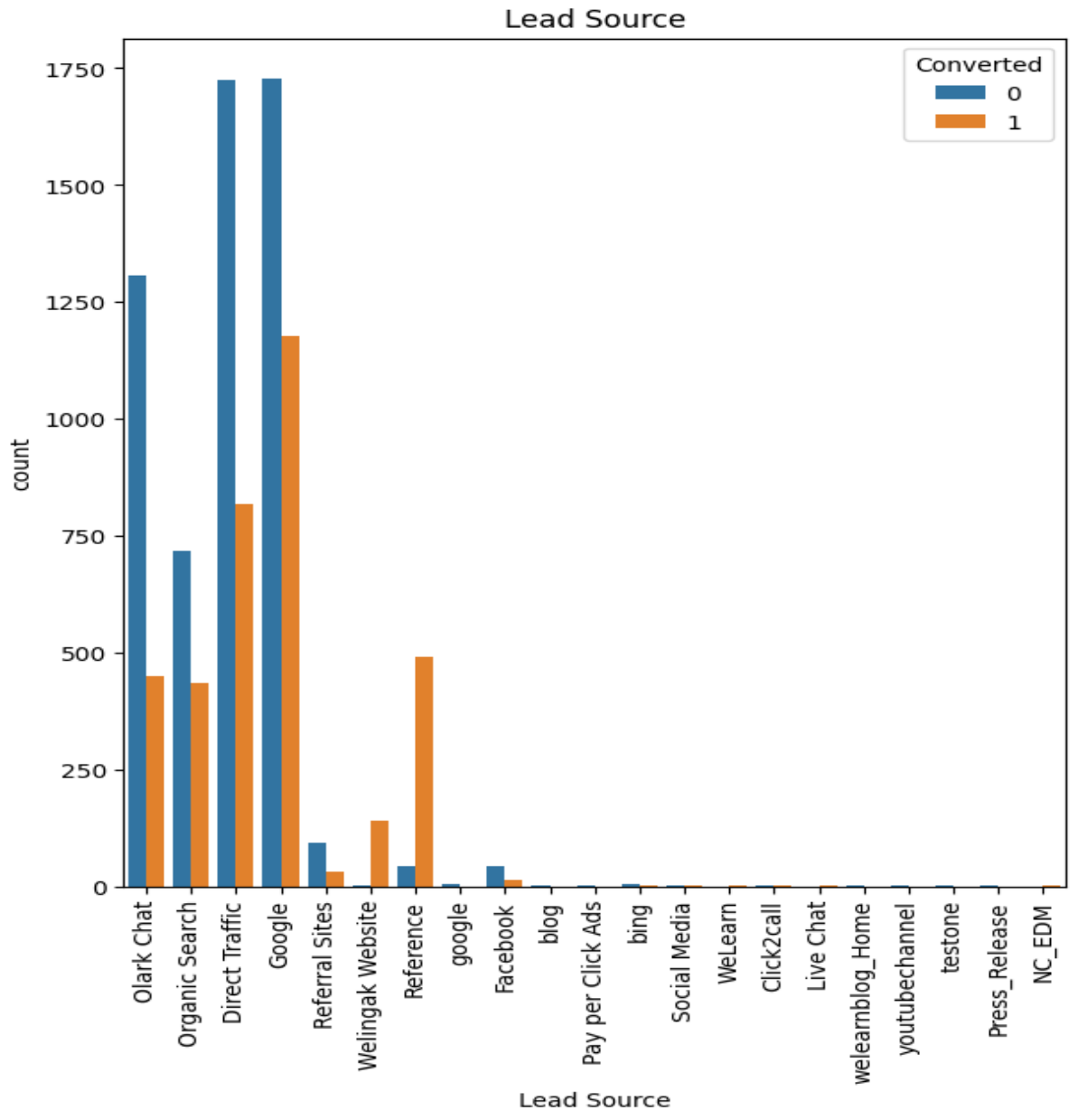




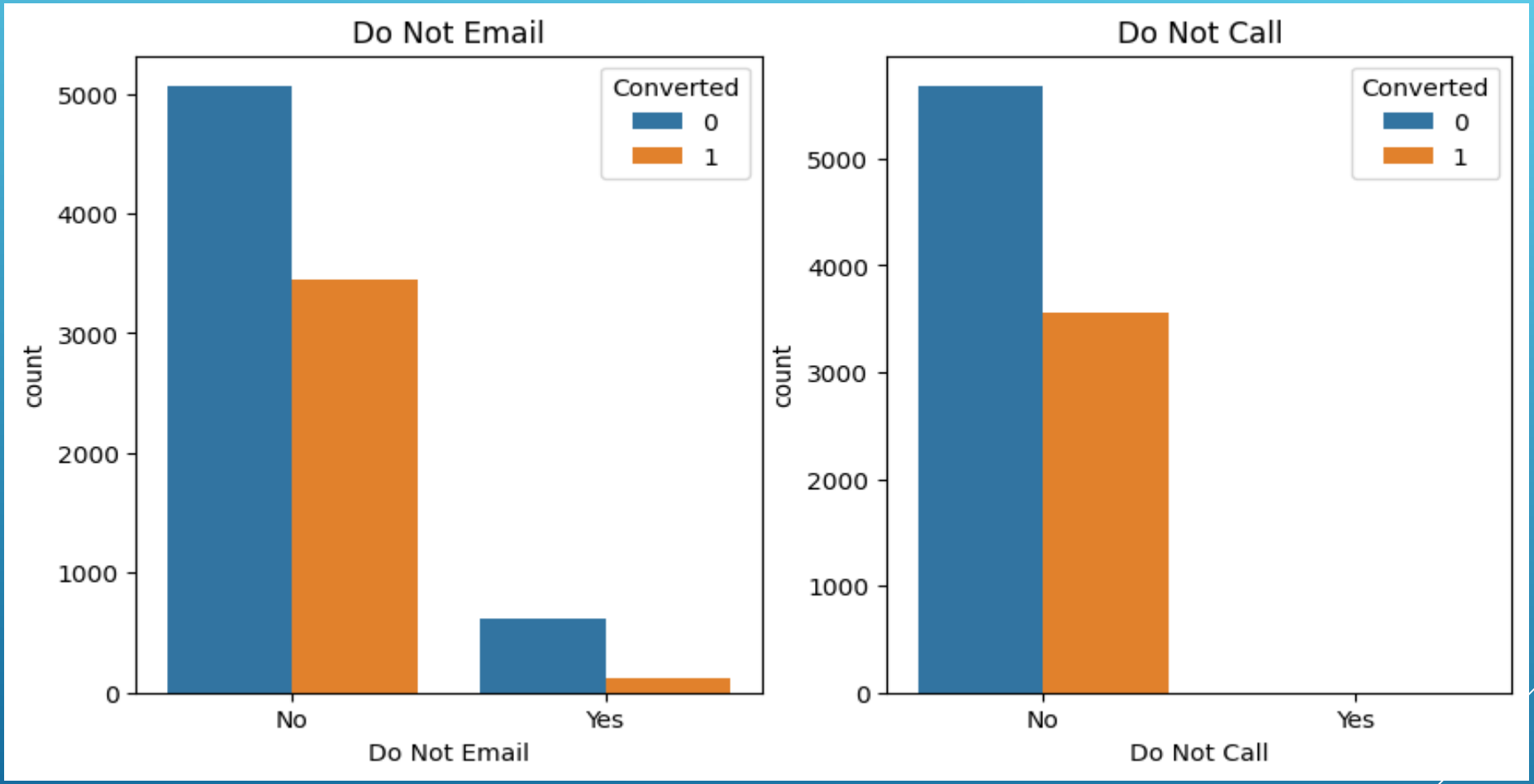


SUBPLOT FOR LEAD ORIGIN & LEAD SOURCE

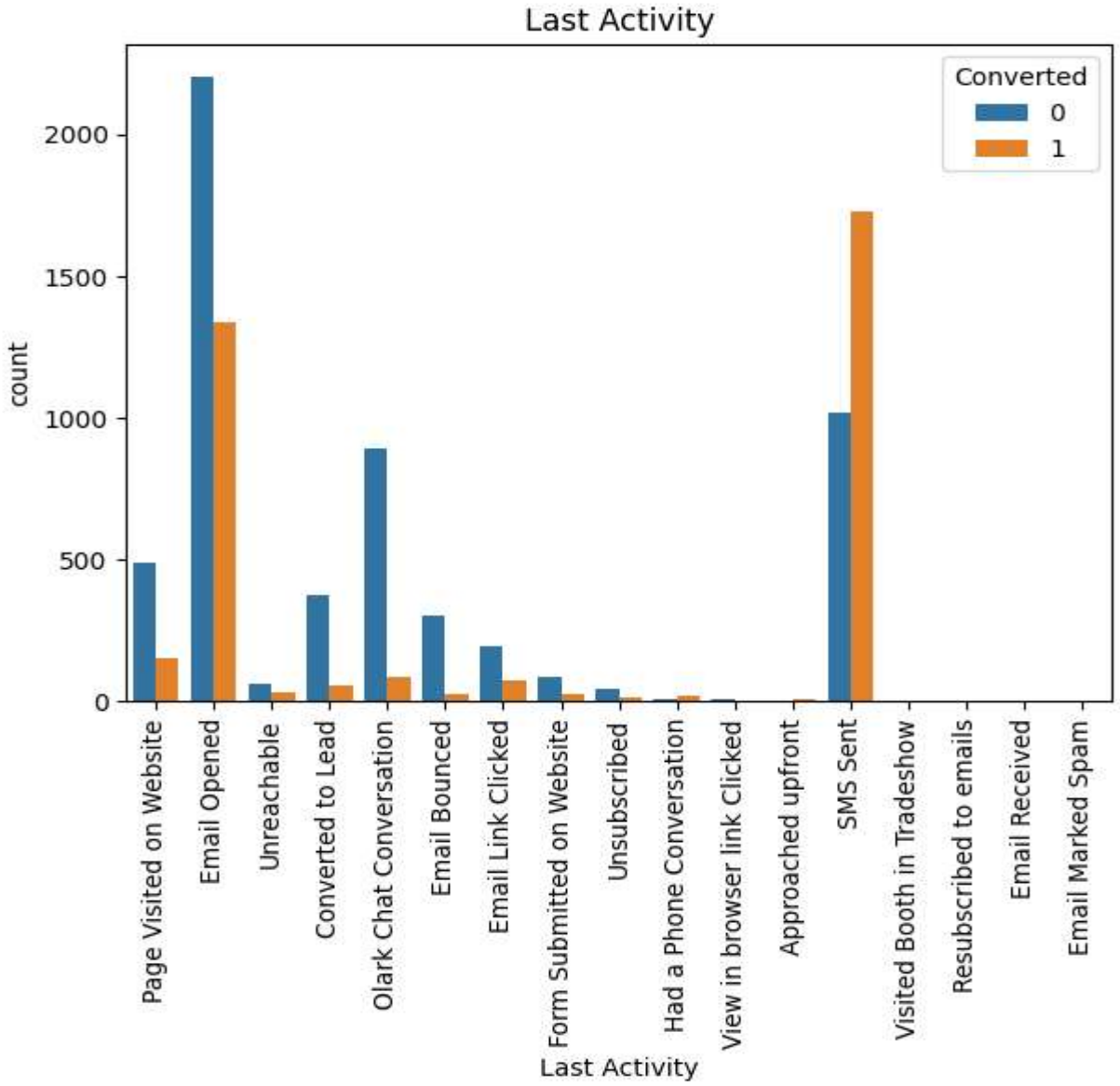


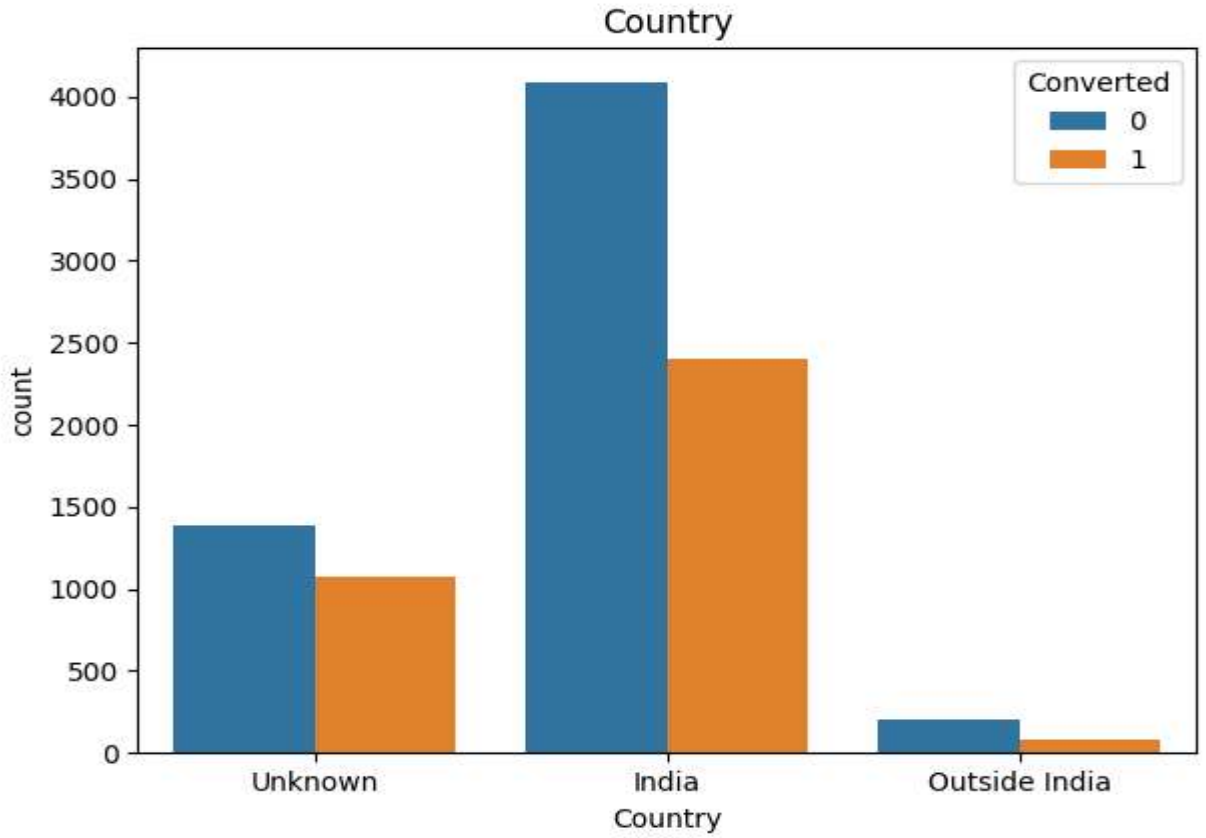


SUBPLOT FOR DO NOT EMAIL VS DO NOT CALL

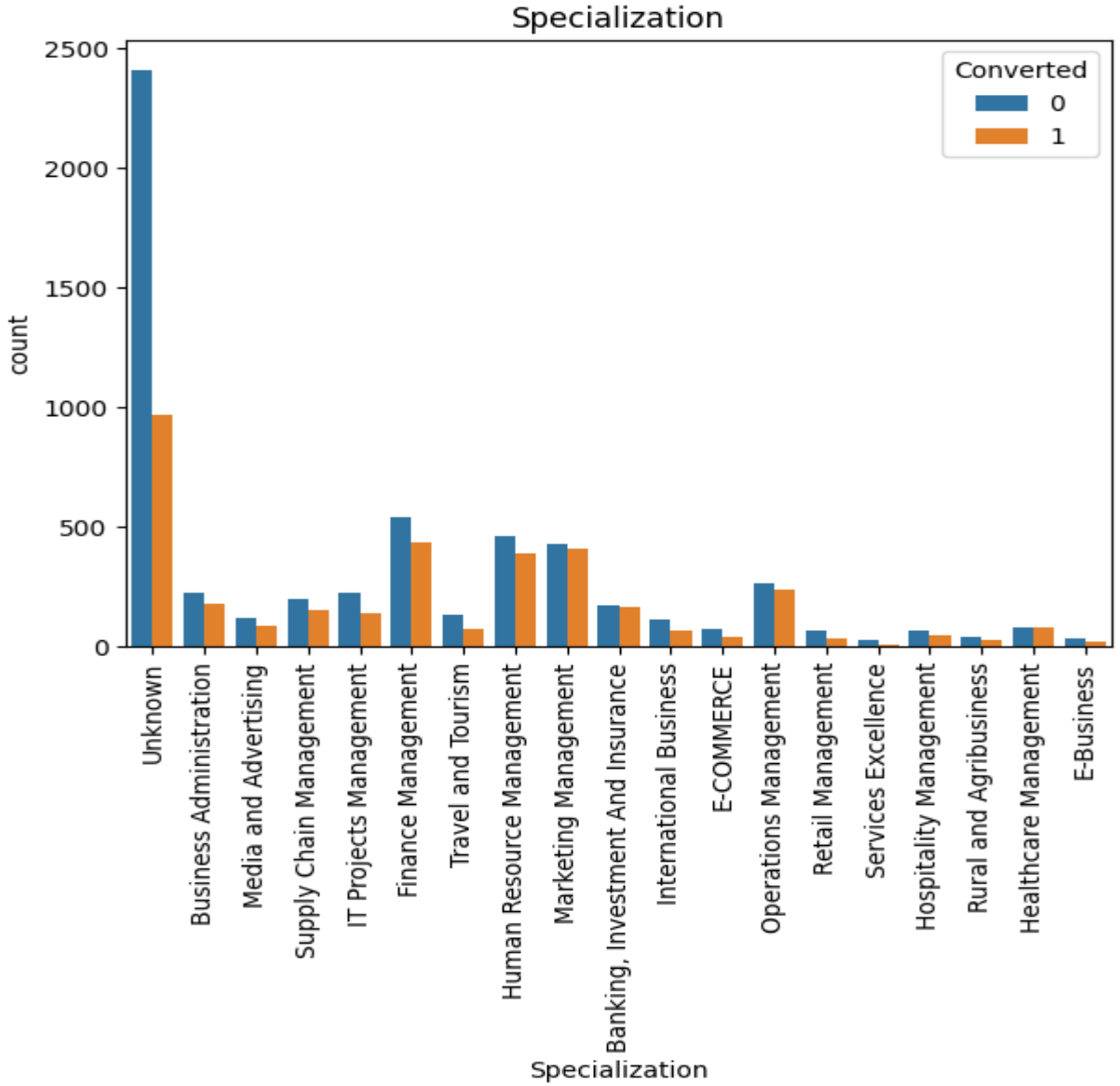


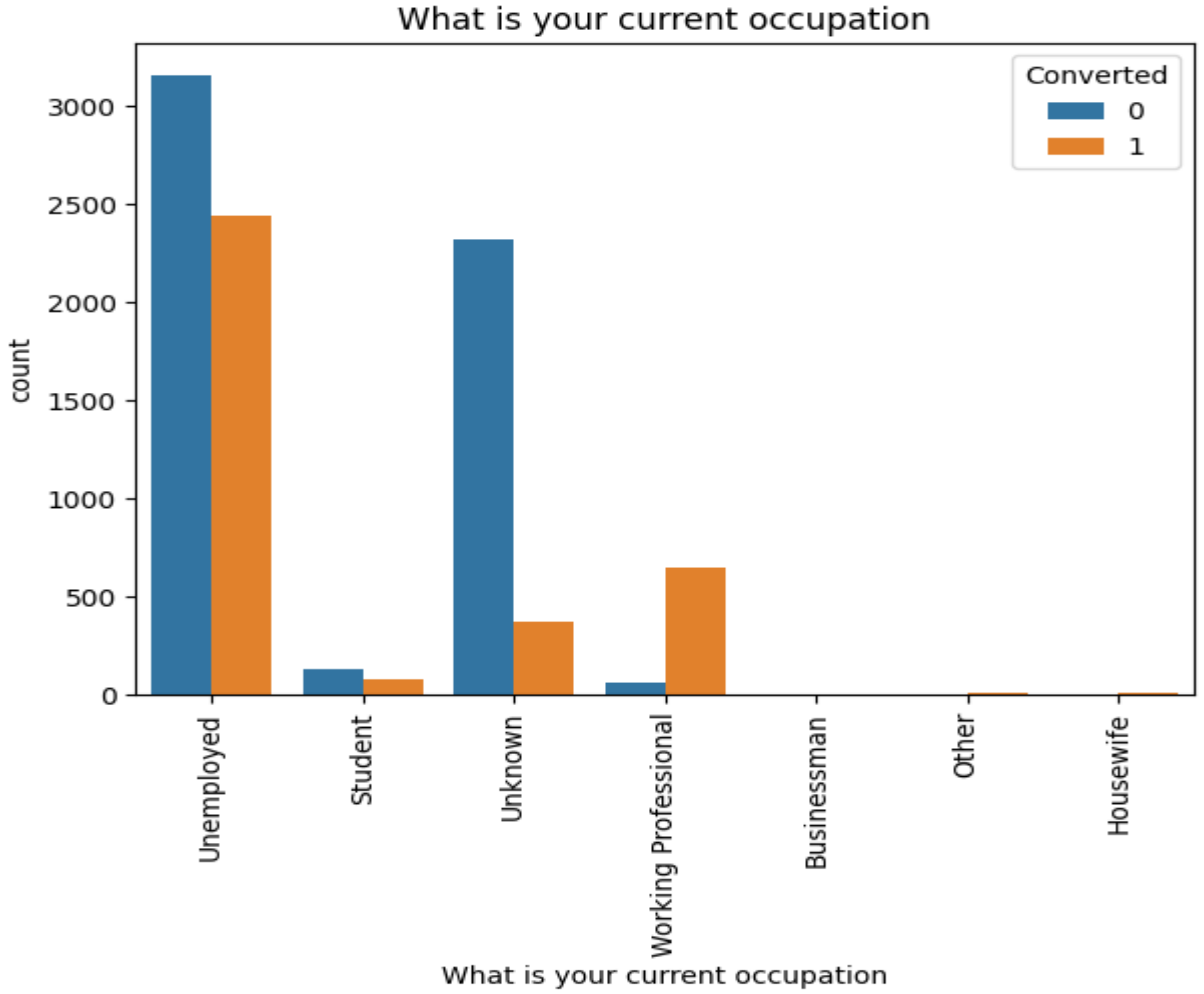
SUBPLOT FOR LAST ACTIVITY & COUNTRY



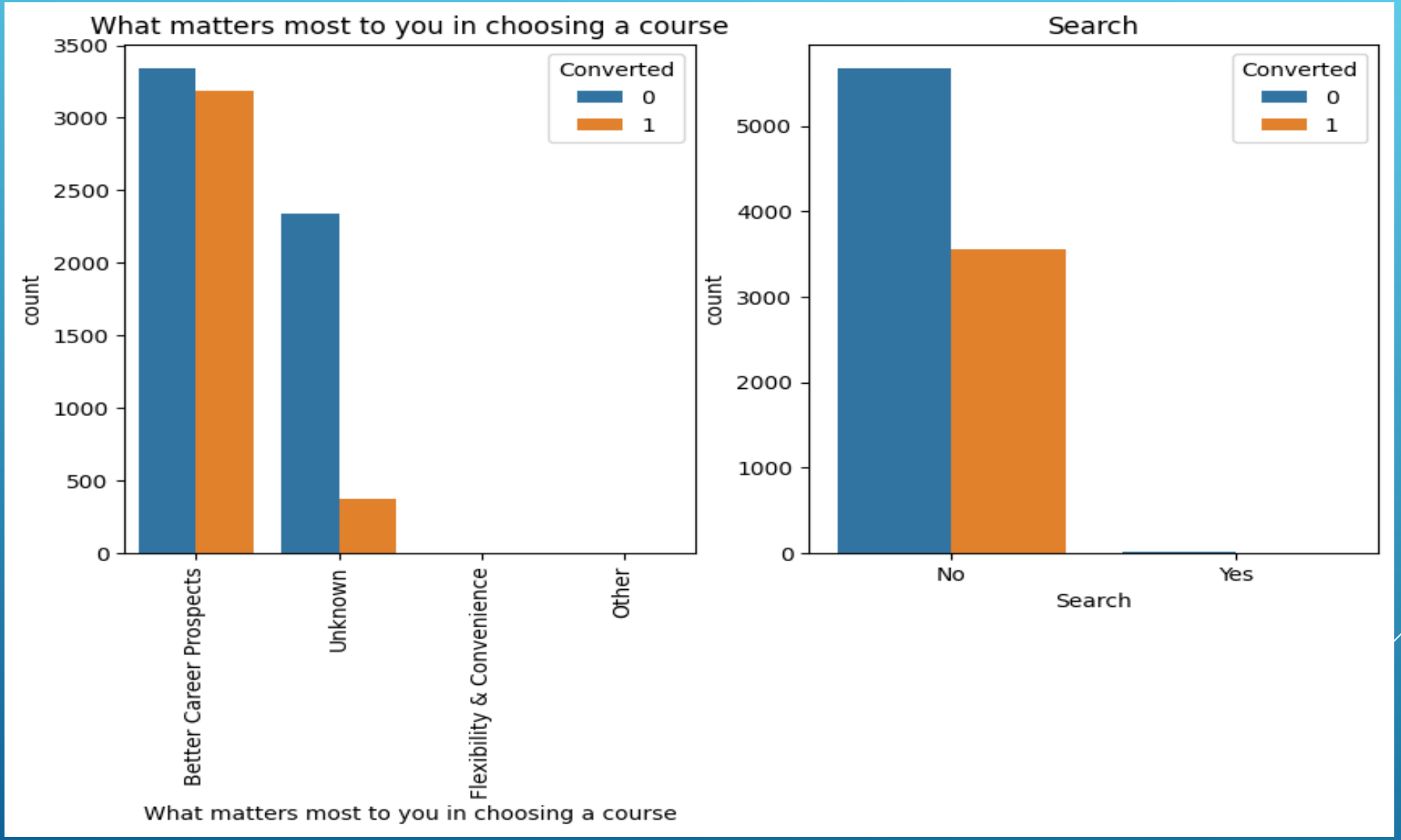


SUBPLOT FOR SPECIALIZATION & WHAT IS YOUR CURRENT LOCATION

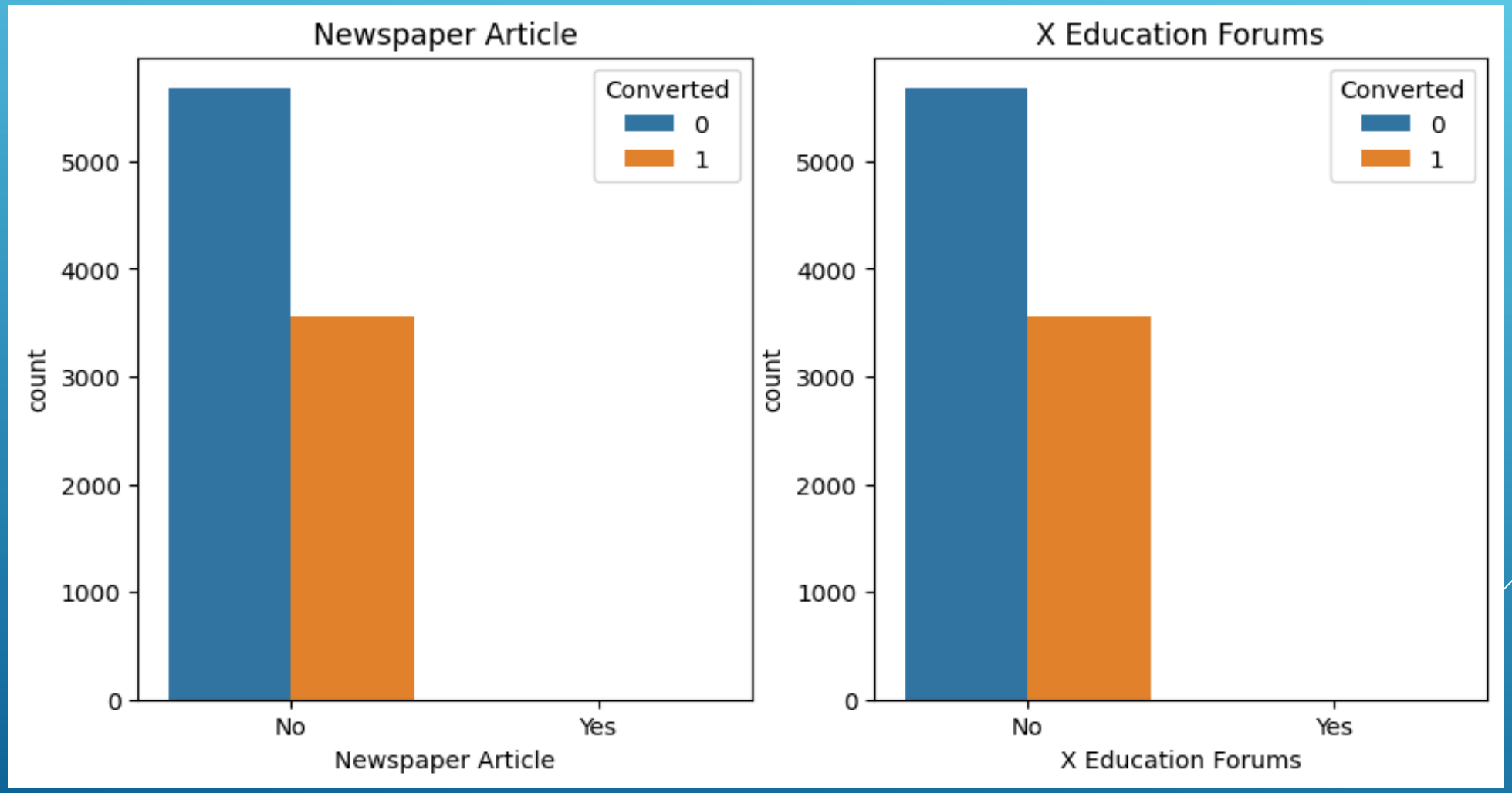




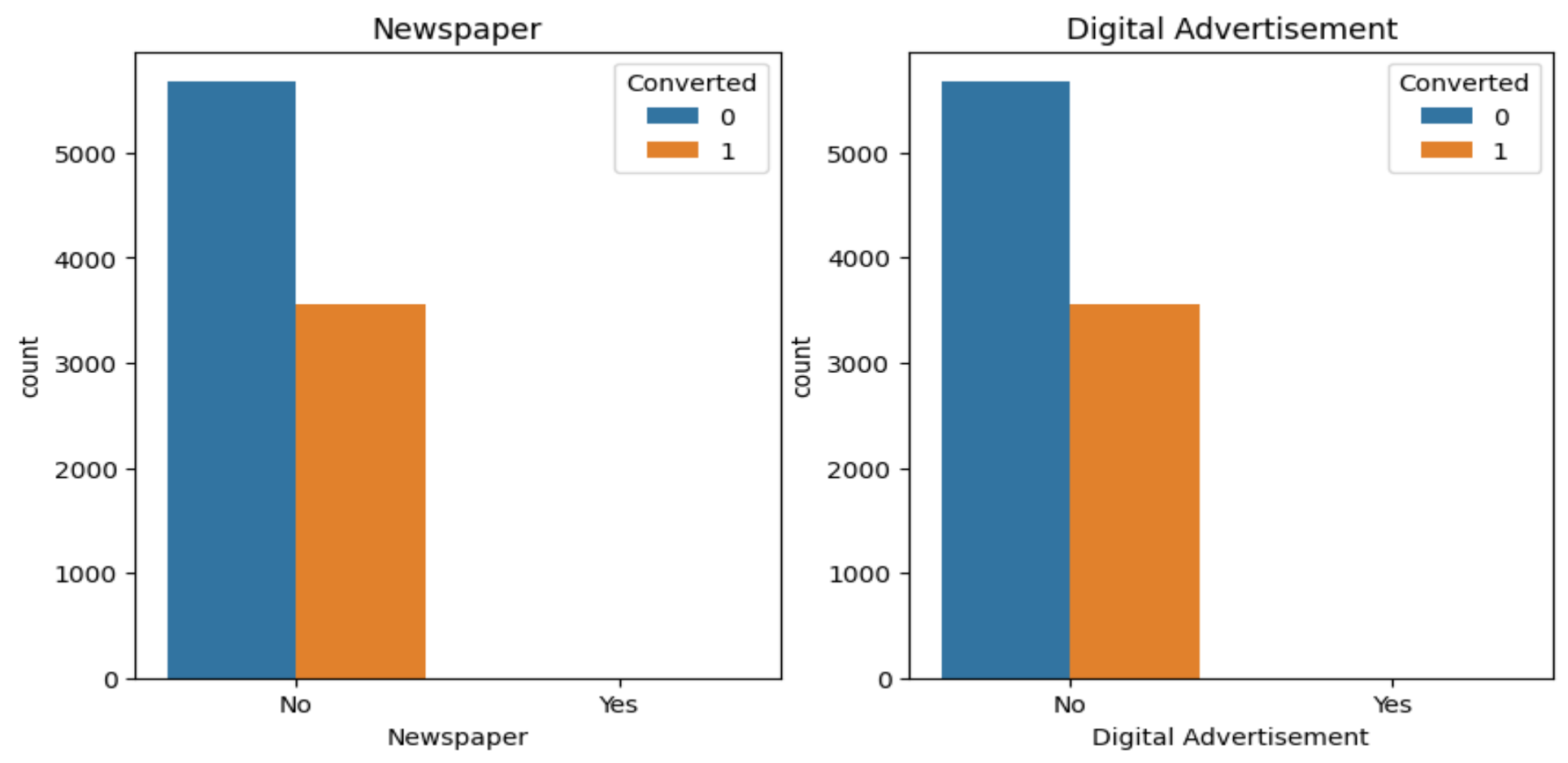
SUBPLOT FOR WHAT MATTERS MOST TO YOU IN CHOOSING A COURSE & SEARCH



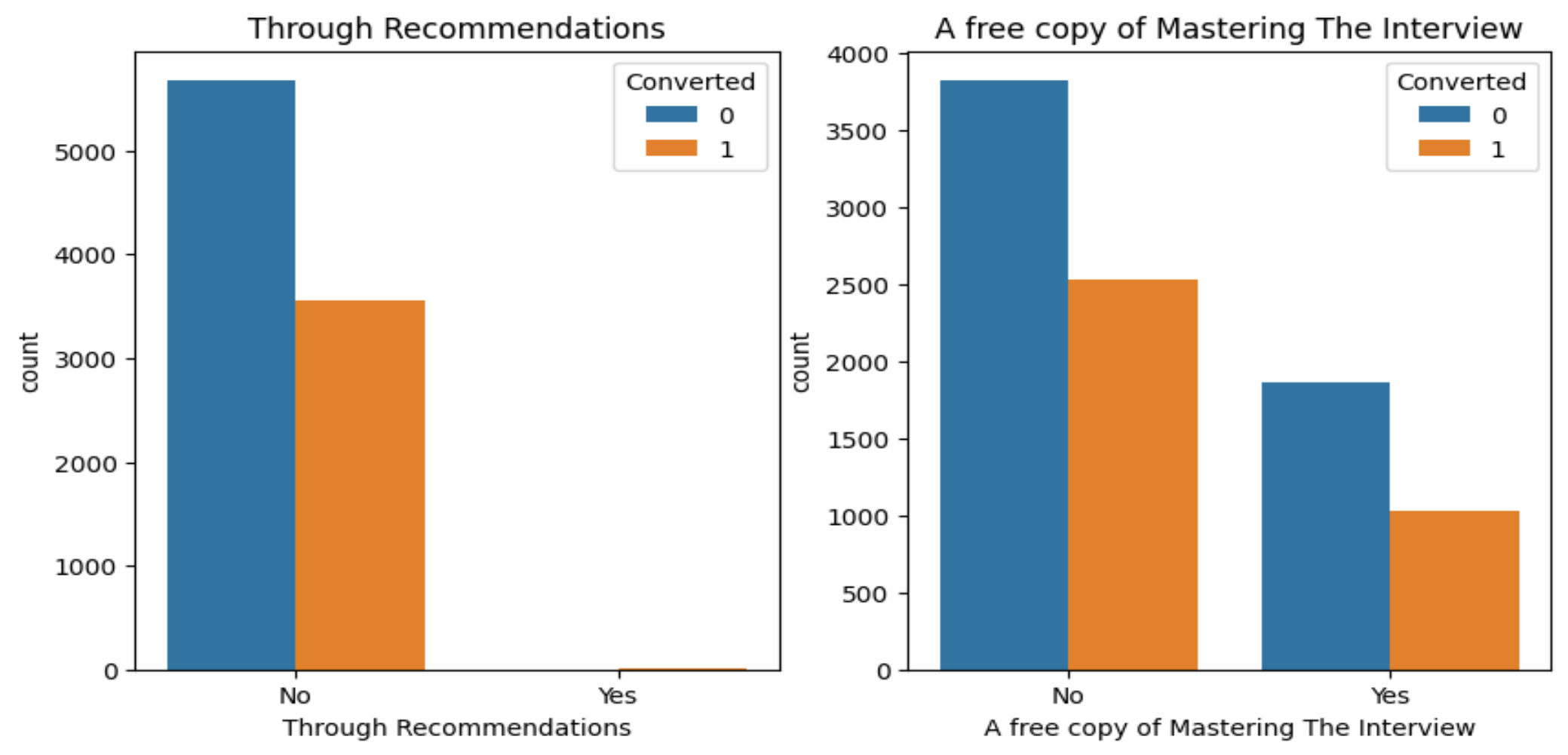
SUBPLOT FOR NEWSPAPER ARTICLE & X EDUCATION FORUMS



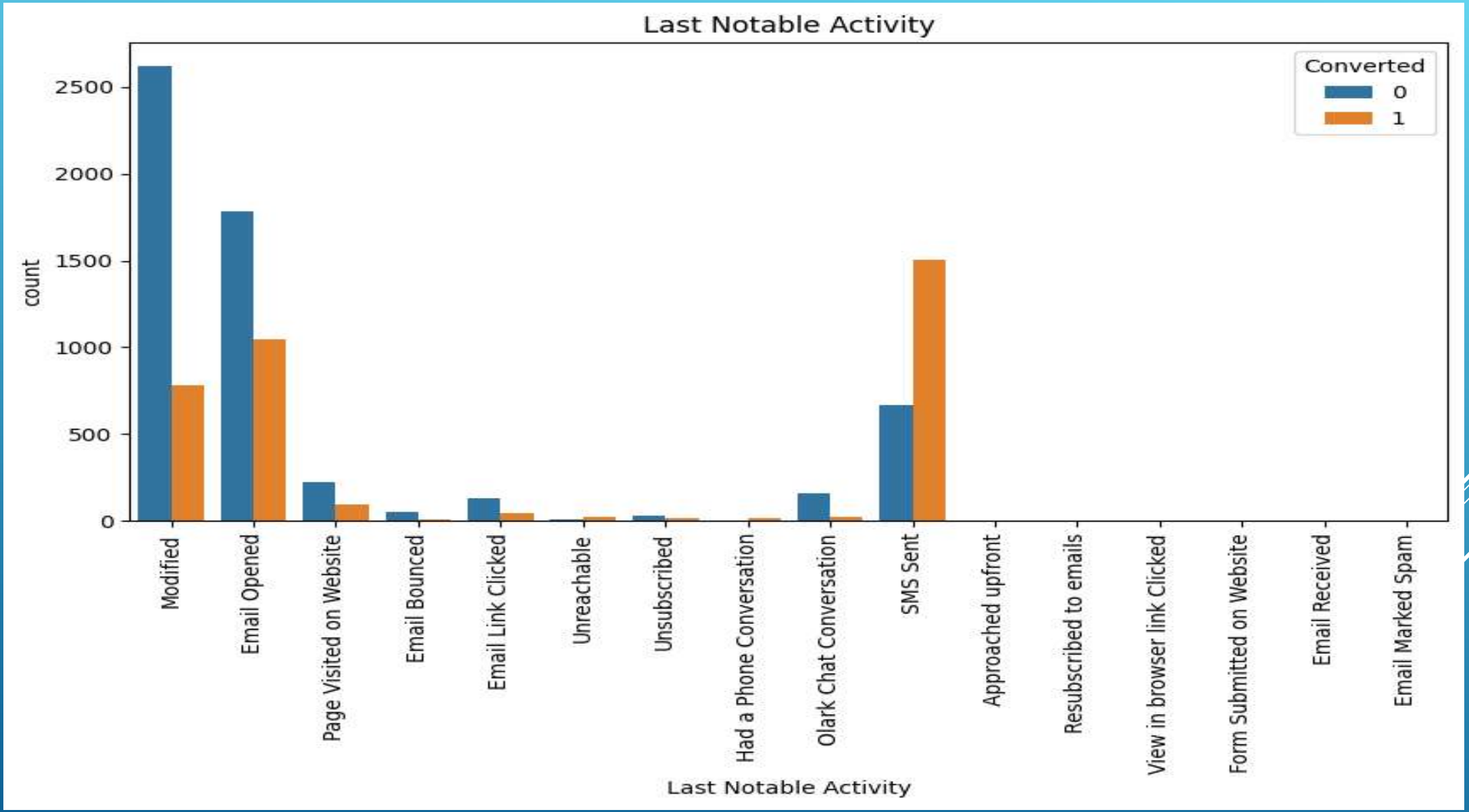
SUBPLOT FOR NEWSPAPER VS DIGITAL ADVERTISEMENT



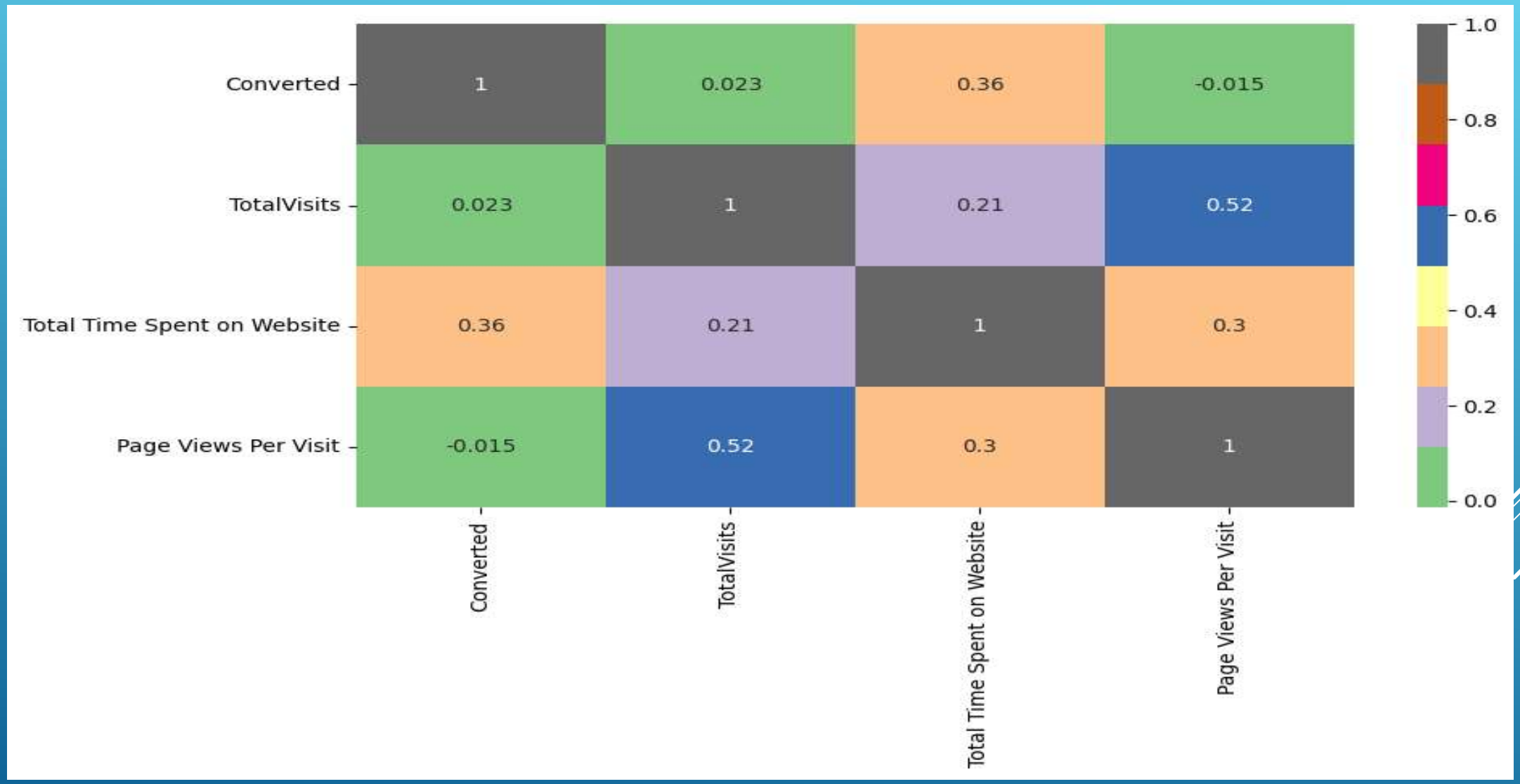
SUBPLOT FOR THROUGH RECOMMENDATIONS & A FREE COPY OF MASTERING THE INTERVIEW



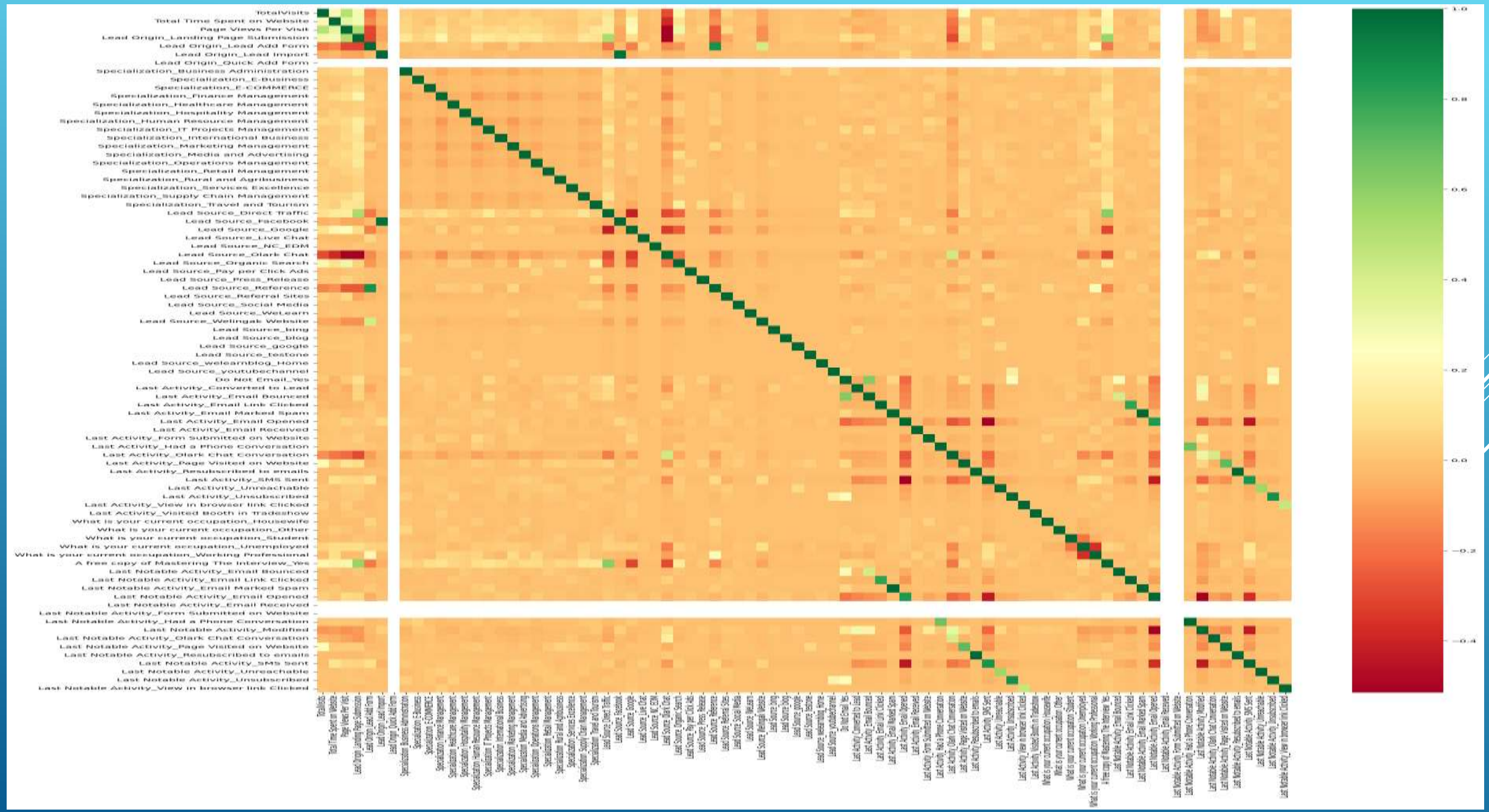
SUBPLOT FOR LAST NOTABLE ACTIVITY



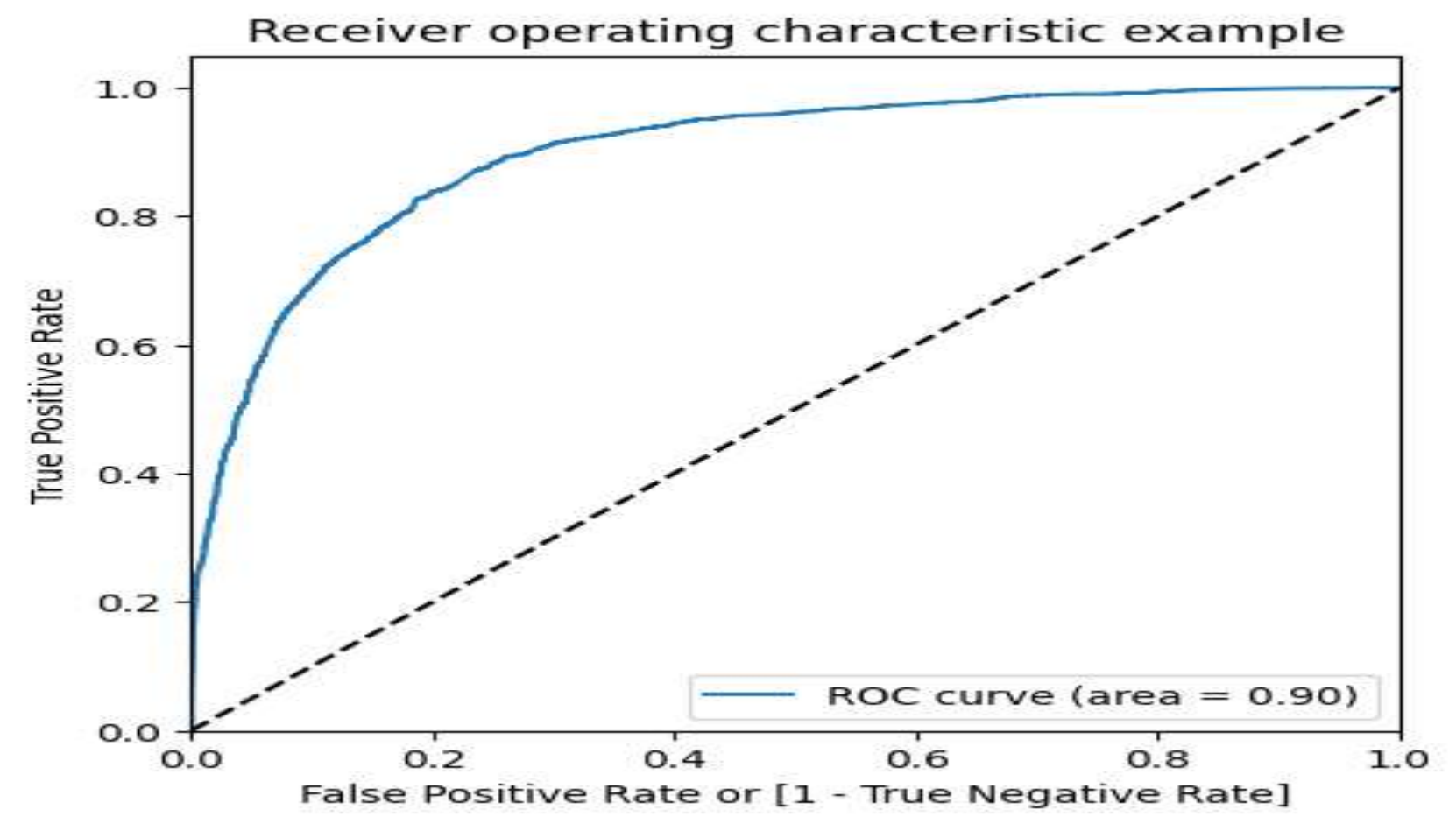
HEATMAP TO CHECK THE CORRELATIONS AMONG VARIABLES



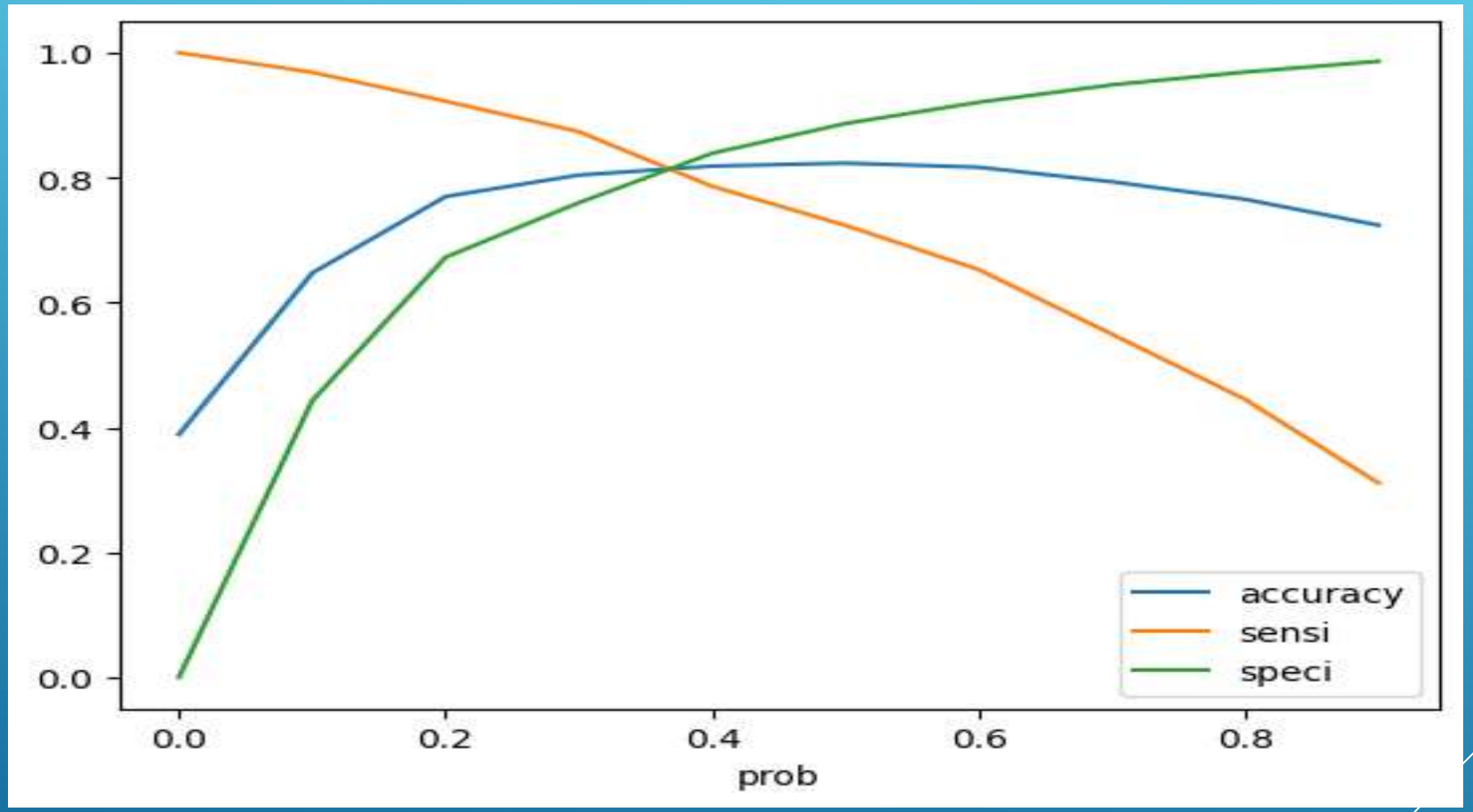
HEATMAP TO CHECK FOR CORRELATIONS AMONG VARIABLES



PLOTTING ROC CURVE



PLOTTING ROC CURVE FOR SENSITIVITY AND SPECIFICITY FOR VARIOUS PROBABILITIES



CONCLUSION

- While we have checked both sensitivity-specificity as well as precision & recall metrics, we have considered the optimal cut off based on sensitivity & specificity for calculating the final prediction.
- Accuracy, Sensitivity & Specificity values of test set are around 78%,81%,76%which are approximately closer to values calculated using trained test data.
- Lead score calculated for the conversion rate final model on Train & test dataset is 82.7% & 80.8% respectively.
- Hence, Overall Model seems to be good.