LEAD SCORE CASE STUDY

MR. ANKUSH PATIL

PROBLEM STATEMENT

- Education sells online courses to industry professionals.

- 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.

BUSINESS OBJECTIVE

- Education wants to know most promising leads.

- For that they want to build a Model which identifies the hot leads.

- Deployment of the model for the future use

METHODS OF MODEL

- Data cleaning and data manipulation.

- 1. Check and handle duplicate data.

- 2. Check and handle NA values and missing values.

- 3. Drop columns, if it contains large amount of missing values and not useful for the analysis.

- 4. Imputation of the values, if necessary.

- 5. Check and handle outliers in data.

- EDA

- 1. Univariate data analysis: value count, distribution of variable etc.

- 2. Bivariate data analysis: correlation coefficients and pattern between the variables etc.

- Feature Scaling & Dummy Variables and encoding of the data.

- Classification technique: logistic regression used for the model making and prediction.

- Validation of the model. Model presentation.
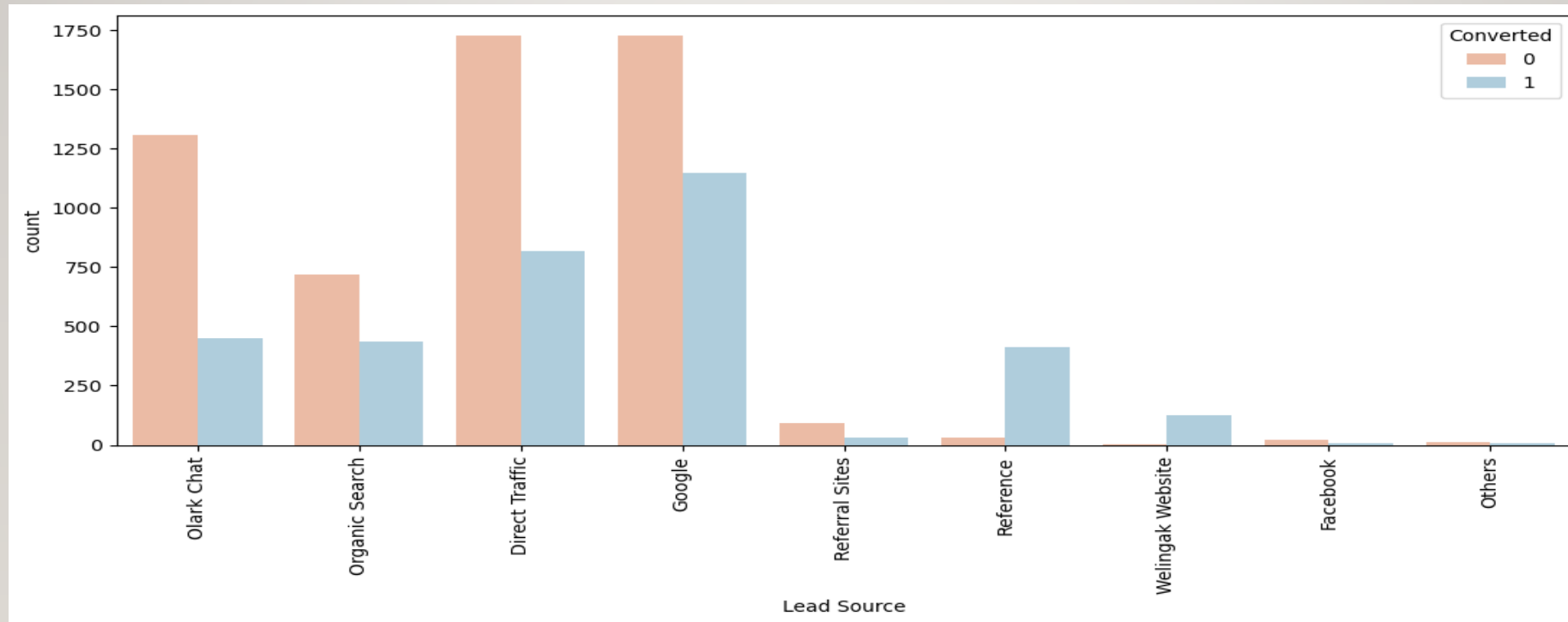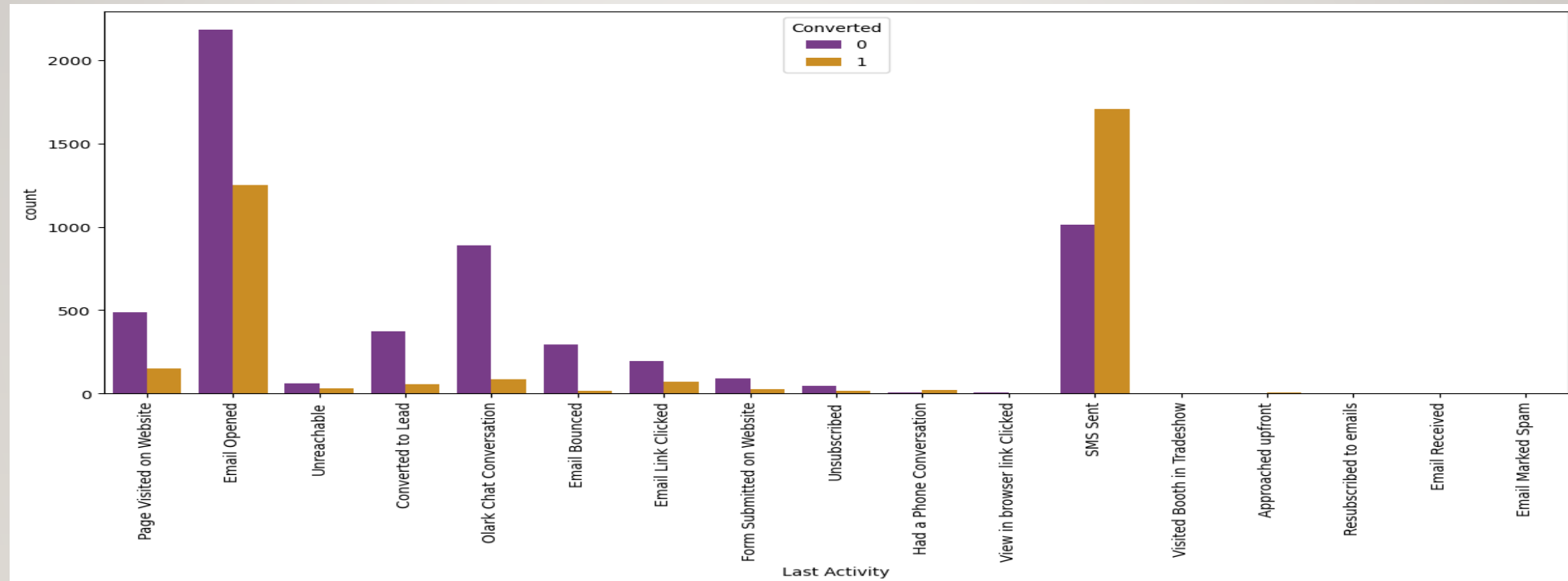
- Conclusions and recommendations
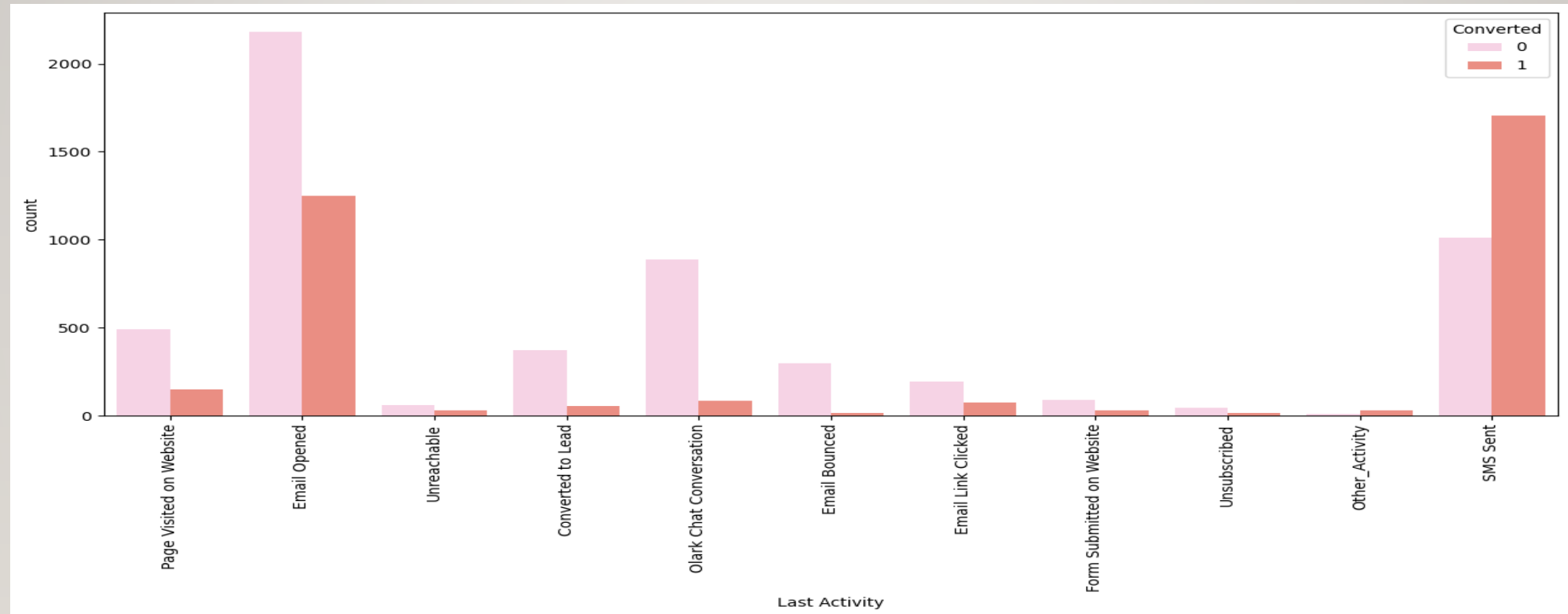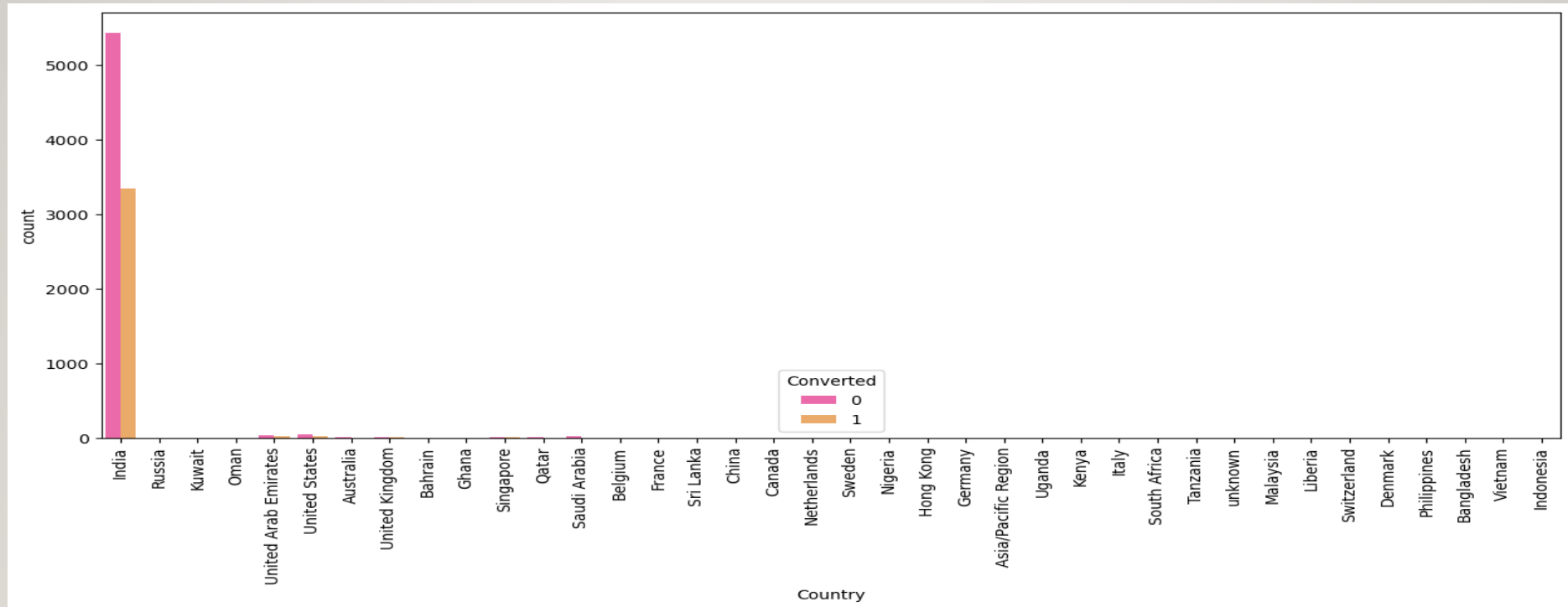
EDA
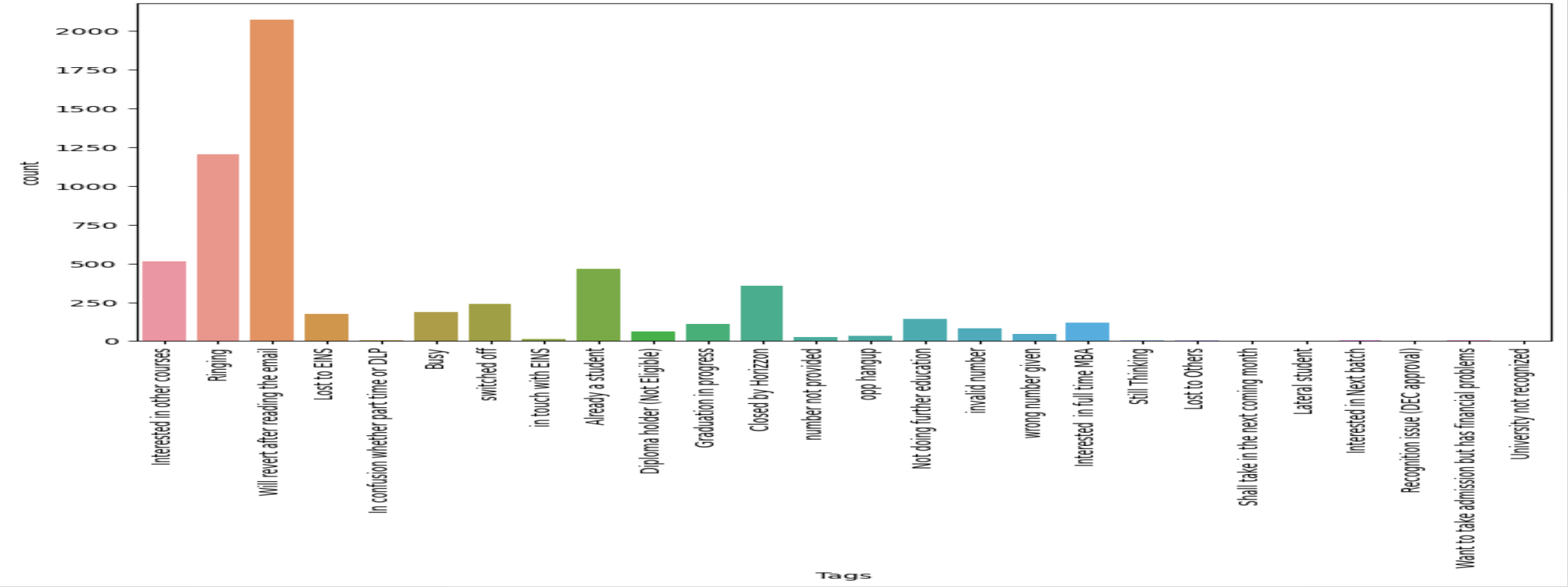
SPECIALIZATION
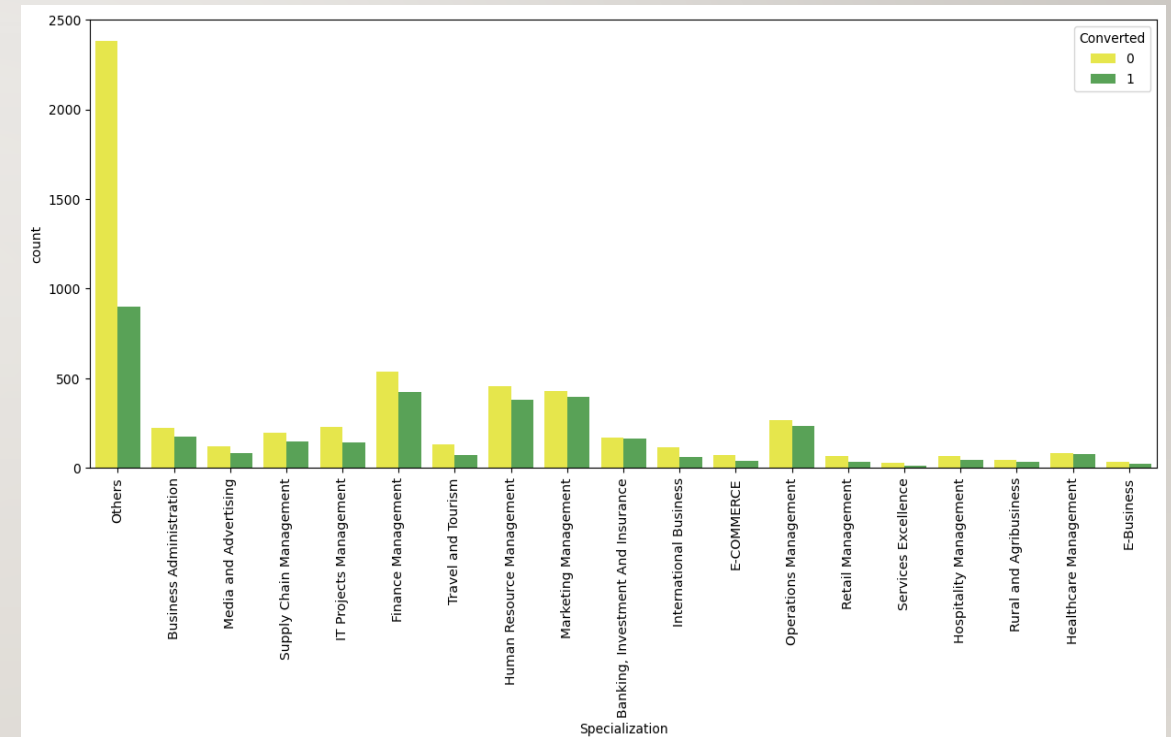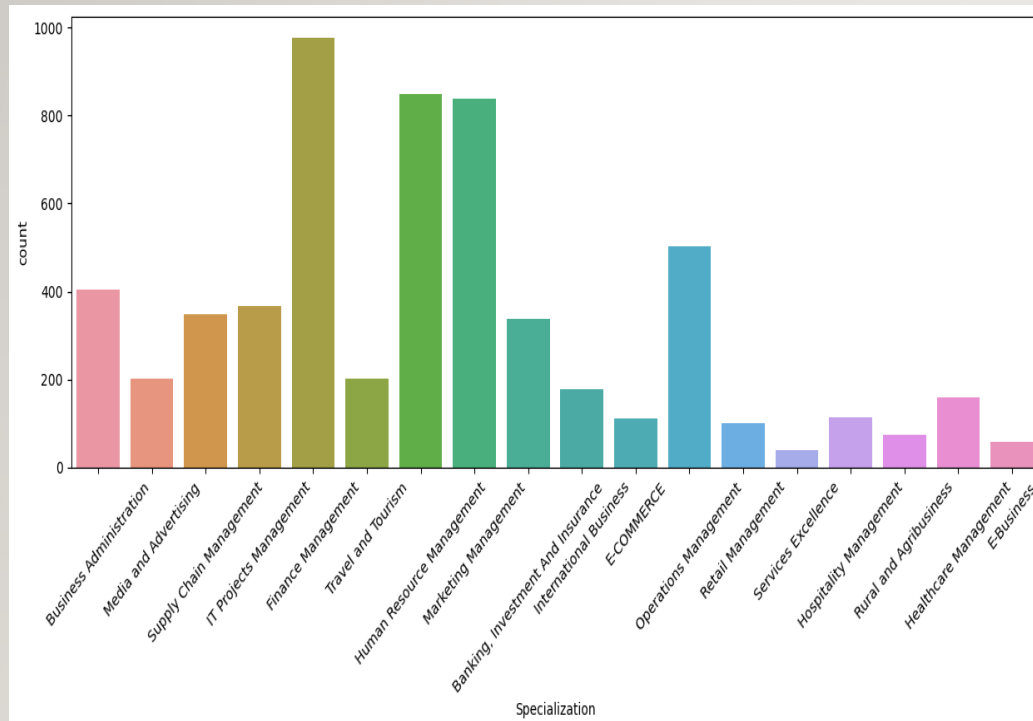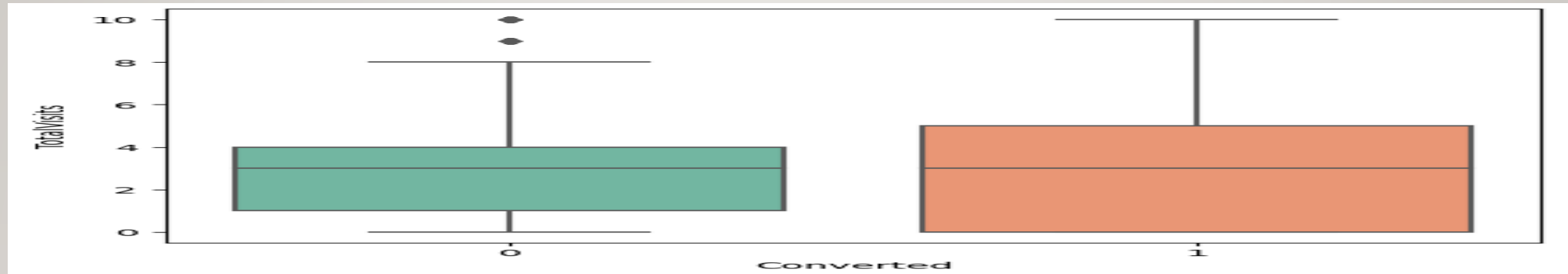
LEAD SCORE
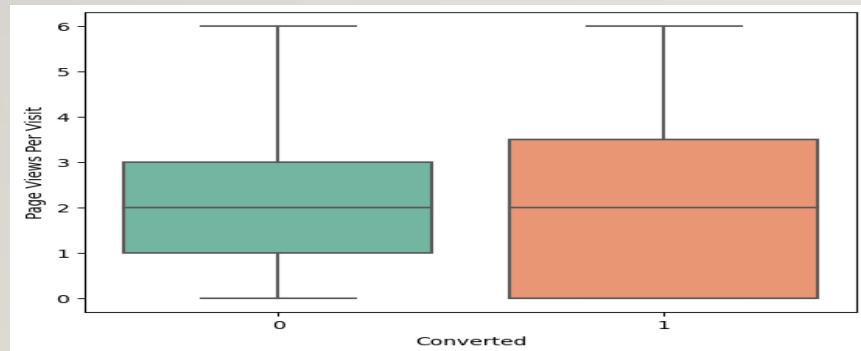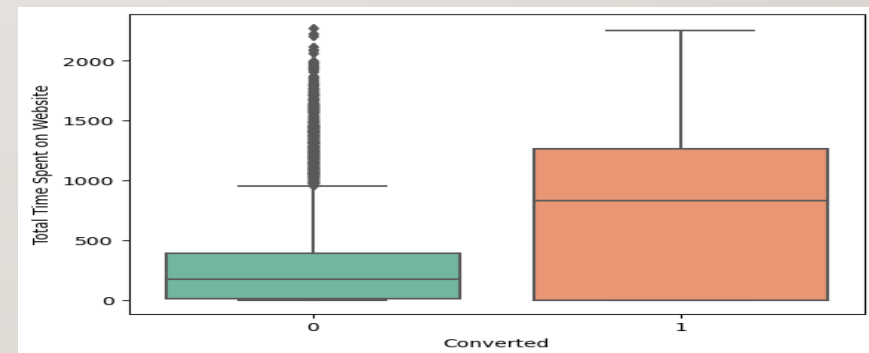
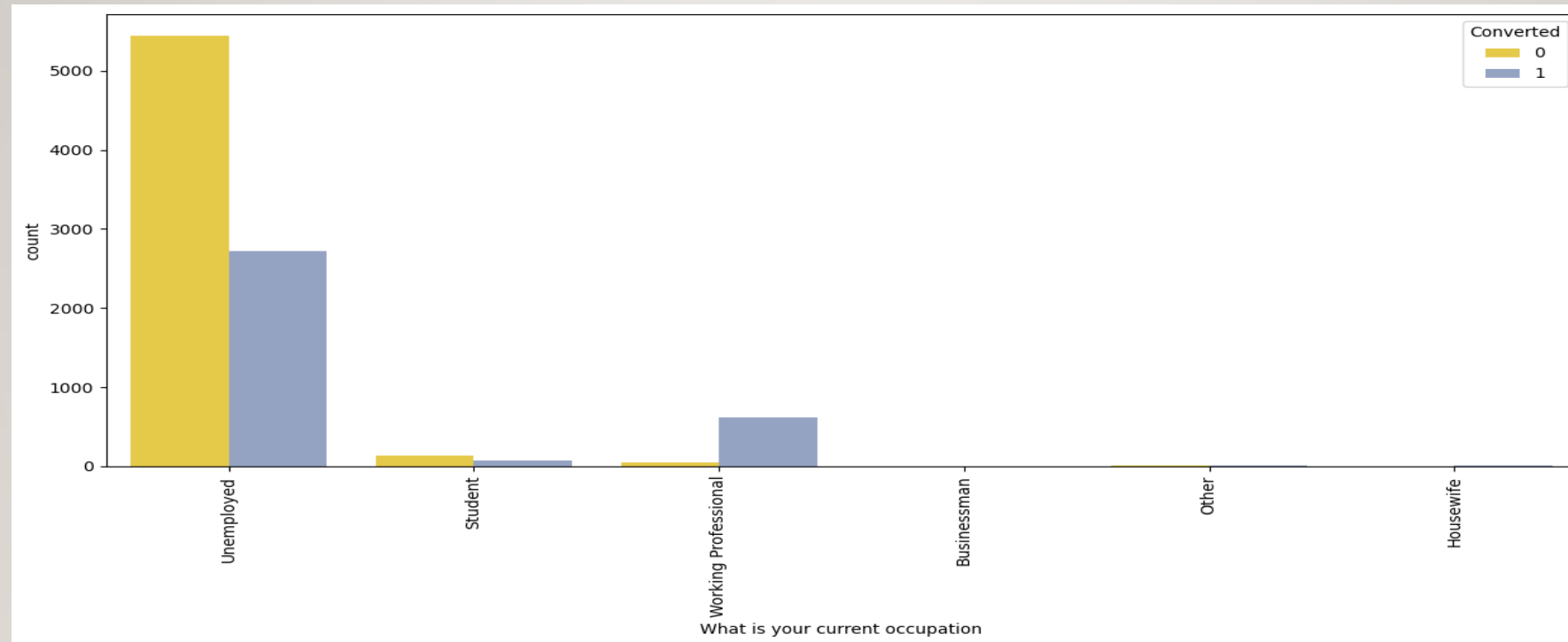# LEAD ACTIVITY VISUALIZATION
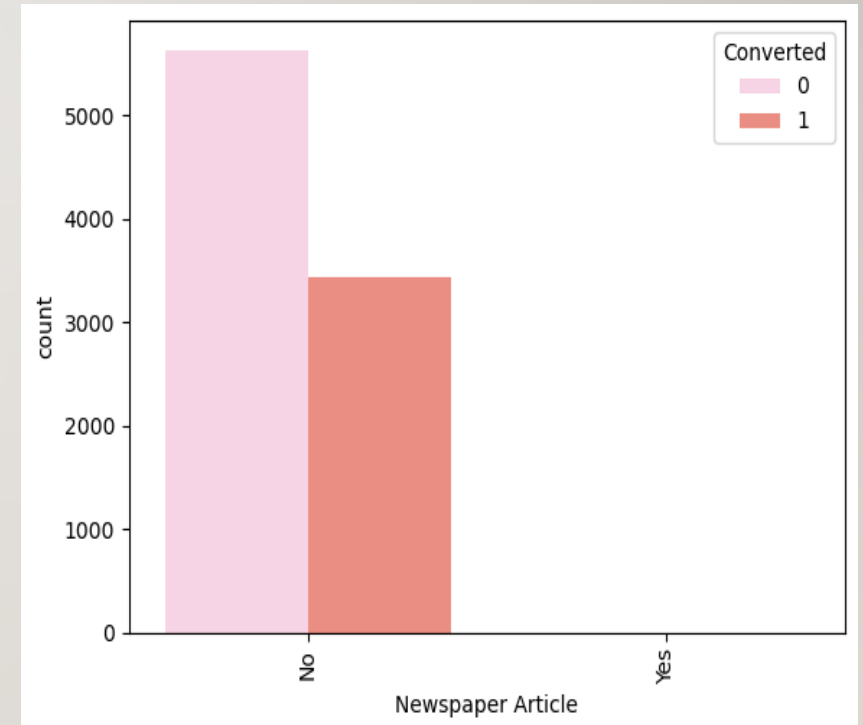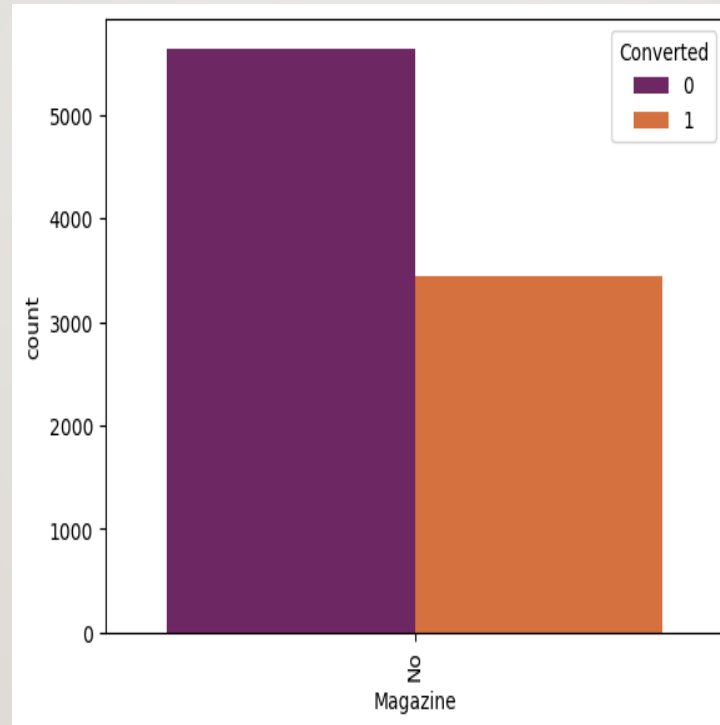
LAST ACTIVITY

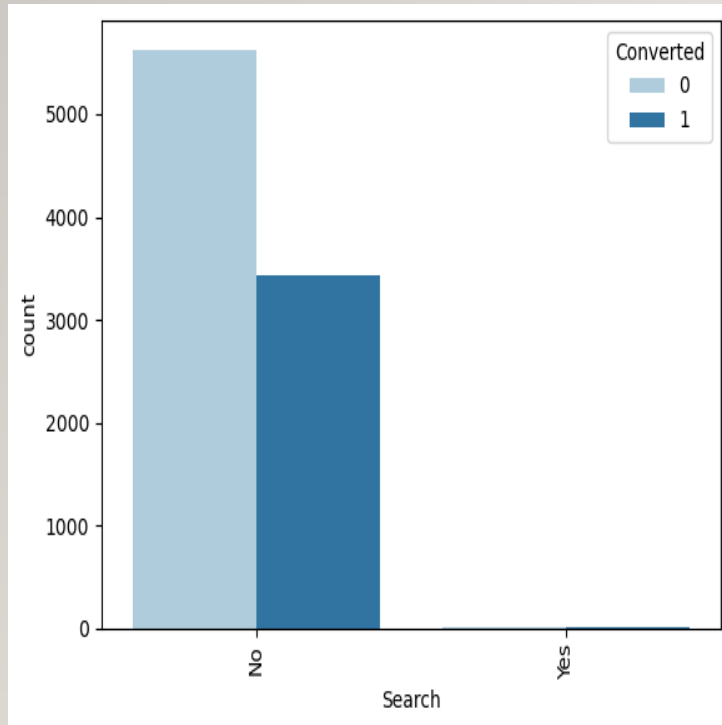# COUNTRY DATA

TOTAL VISITS

PAGE VIEWS PER VISIT
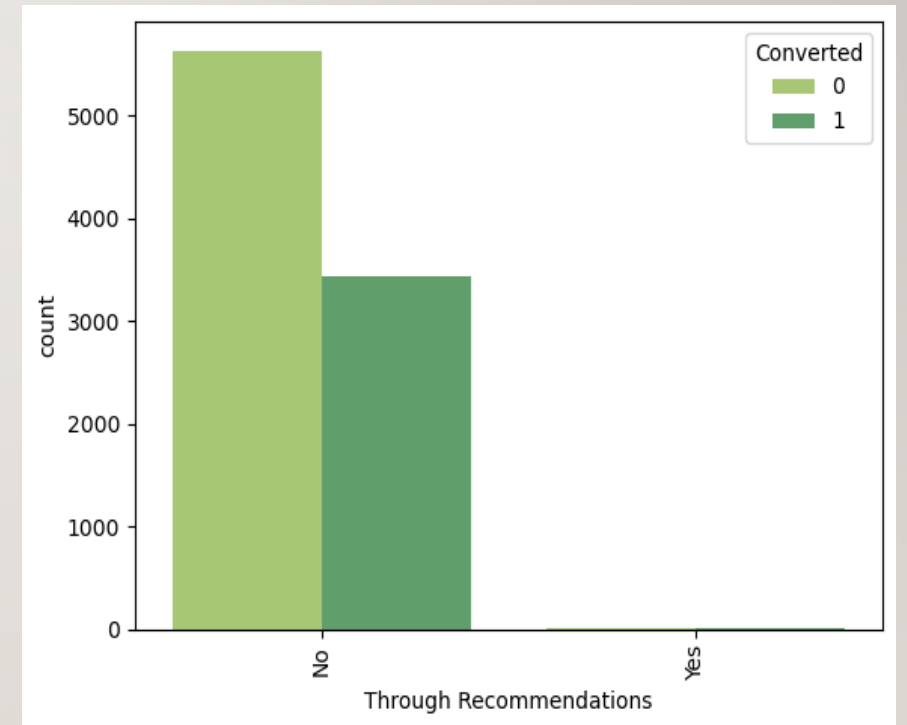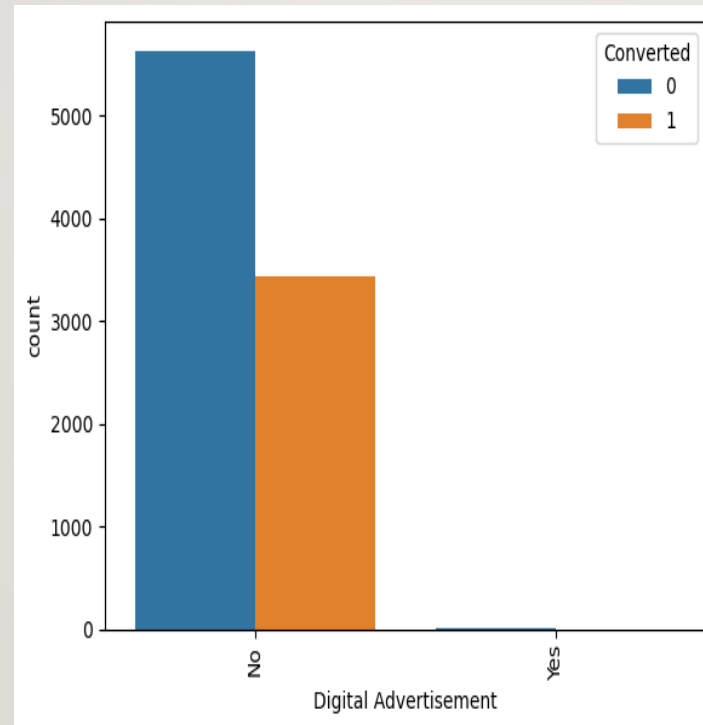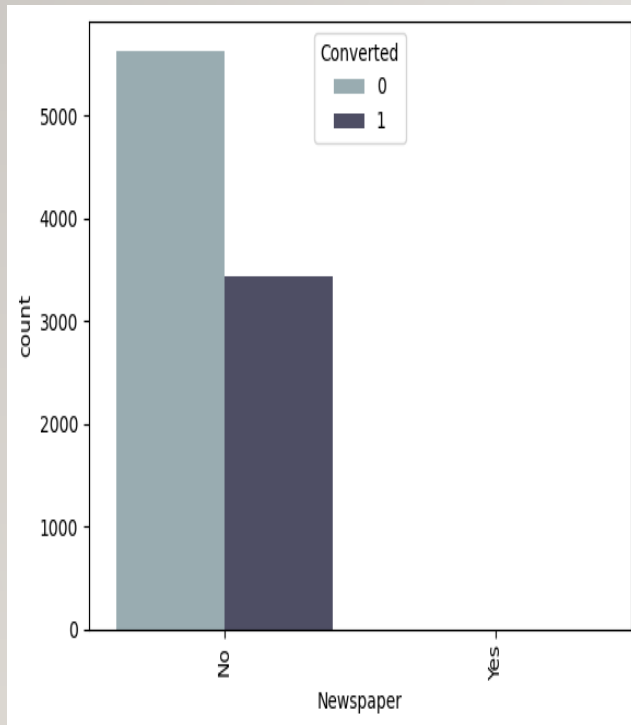
TOTAL TIME SPENT ON WEBSITE
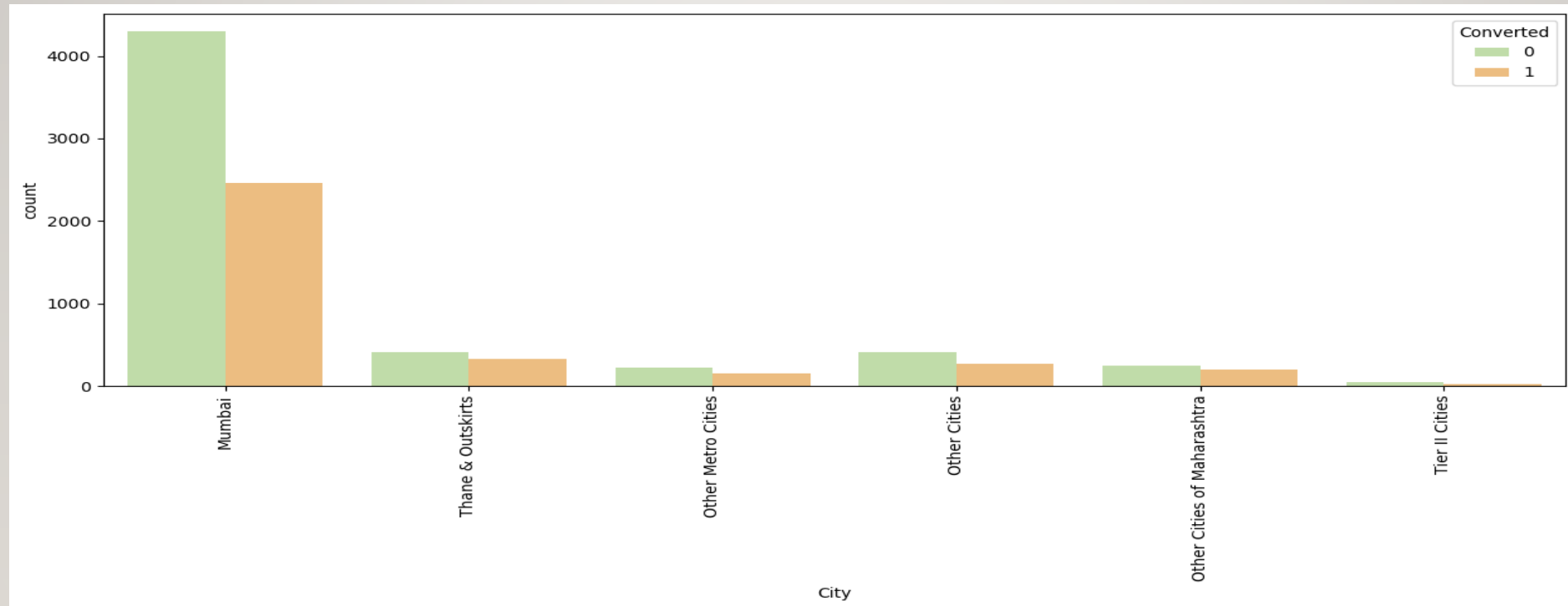
WHAT IS YOUR CURRENT OCCUPATION

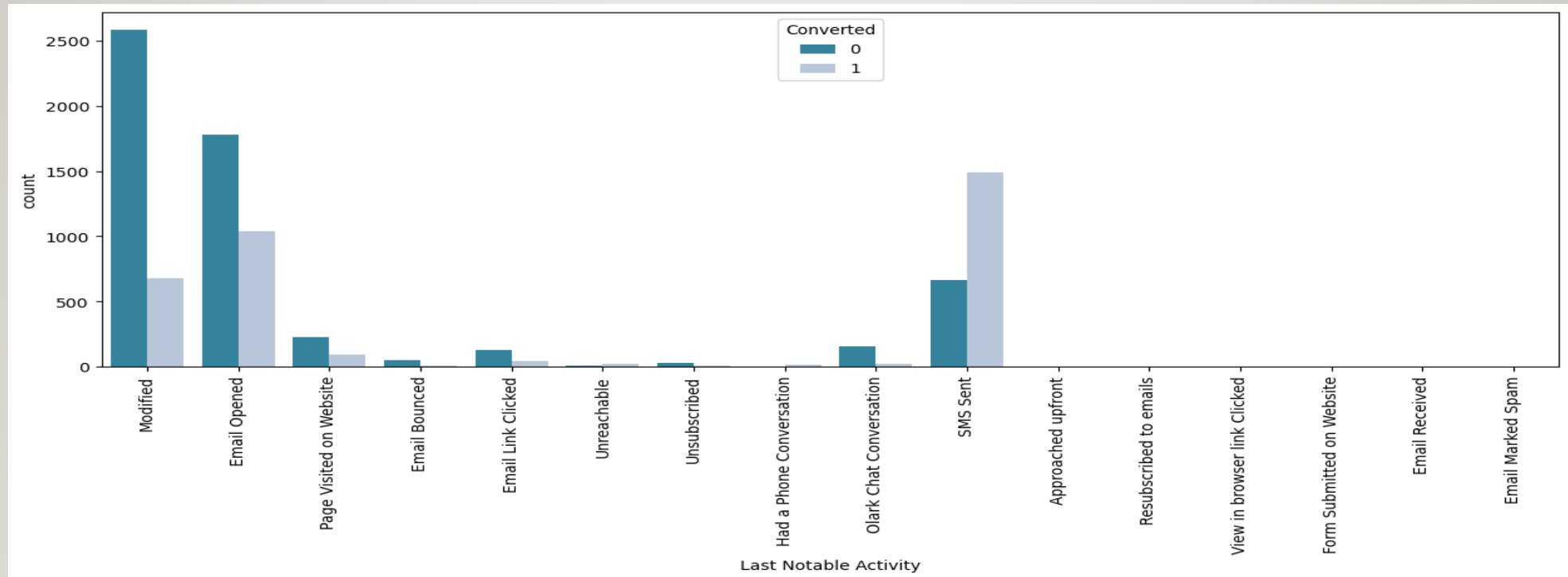# VISUALIZE ON SEARCH, MAGAZINE, NEWSPAPER ARTICLE

VISUALIZE ON NEWSPAPER, DIGITAL ADVERTISEMENT, RECOMMENDATIONS

VISUALIZE CITY DATA

VISUALIZE LAST NOTABLE ACTIVITY

MODEL BUILDING

- Splitting the Data into Training and Testing Sets

- The first basic step for regression is performing a train-test split, we have chosen 70:30 ratio.

- Use RFE for Feature Selection, Running RFE with variables as output

- Building Model by removing the variable

- Predictions on test data set

- Overall accuracy 80.4- 81.7%

ROC CURVE:

FINDING OPTIMAL CUT OFF POINT
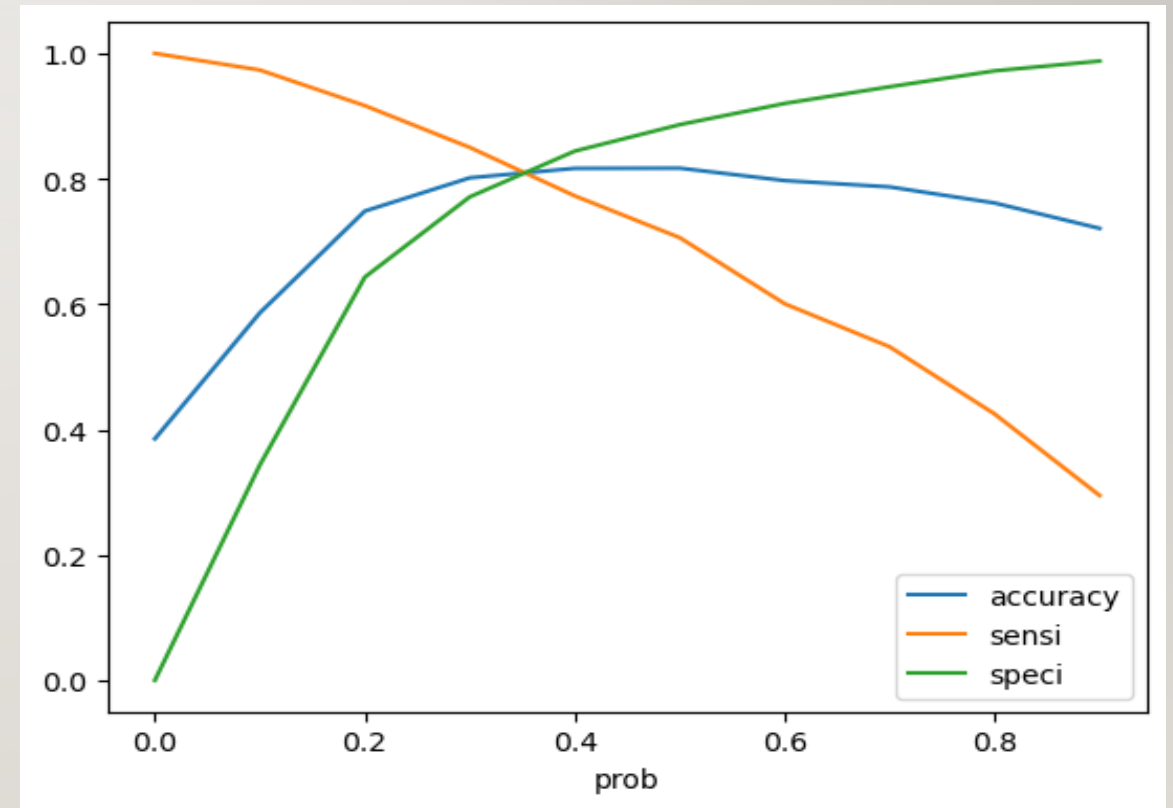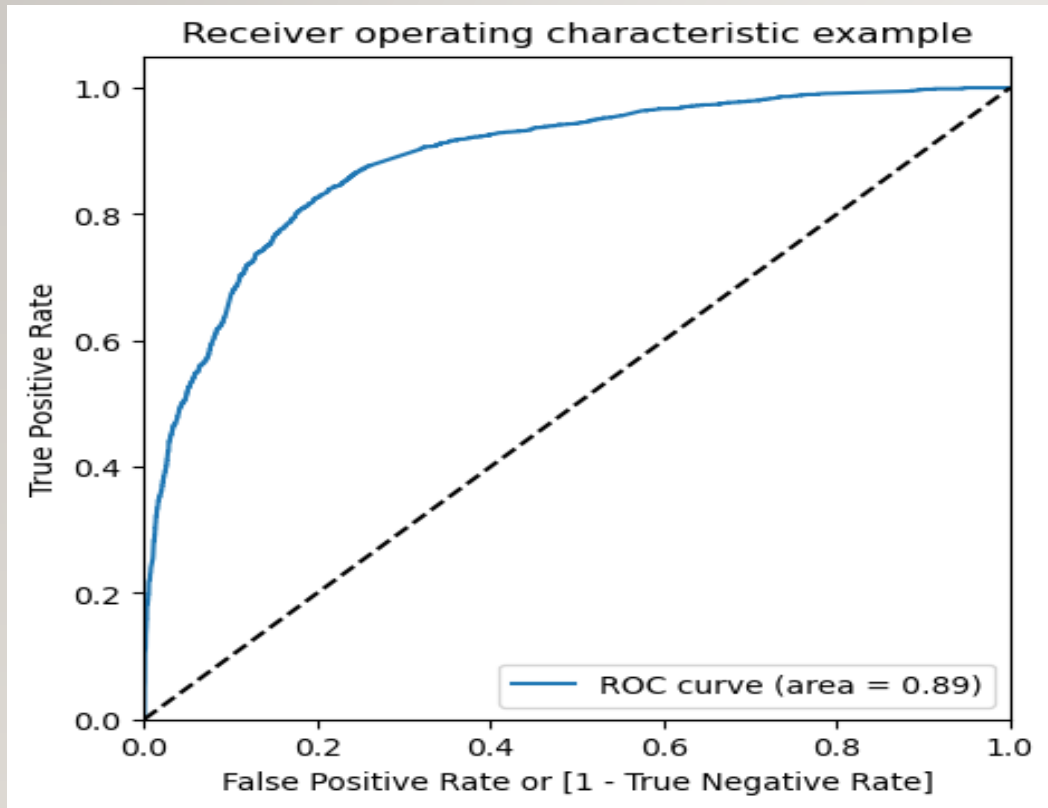
CONCLUSION

- The total time spend on the Website.

- Total number of visits.

- When the lead source was: a. Google , Direct traffic, Organic search, Welingak website

- When the last activity was: SMS , Olark chat conversation

- When their current occupation is as a working professional. Keeping these in mind the X Education can flourish as they have a very high chance to get almost all the potential buyers to change their mind and buy their courses.