

---

# LEAD SCORE

---

BY  
SAHITHI & HARSHITHA

# Problem Statement

X Education gets a lot of leads, its lead conversion rate is very poor at around 30%. X Education wants to make lead conversion process more efficient by identifying the most potential leads, also known as Hot Leads. Their sales team want to know these potential set of leads, which they will be focusing more on communicating rather than making calls to everyone.

# Objective of the Study:

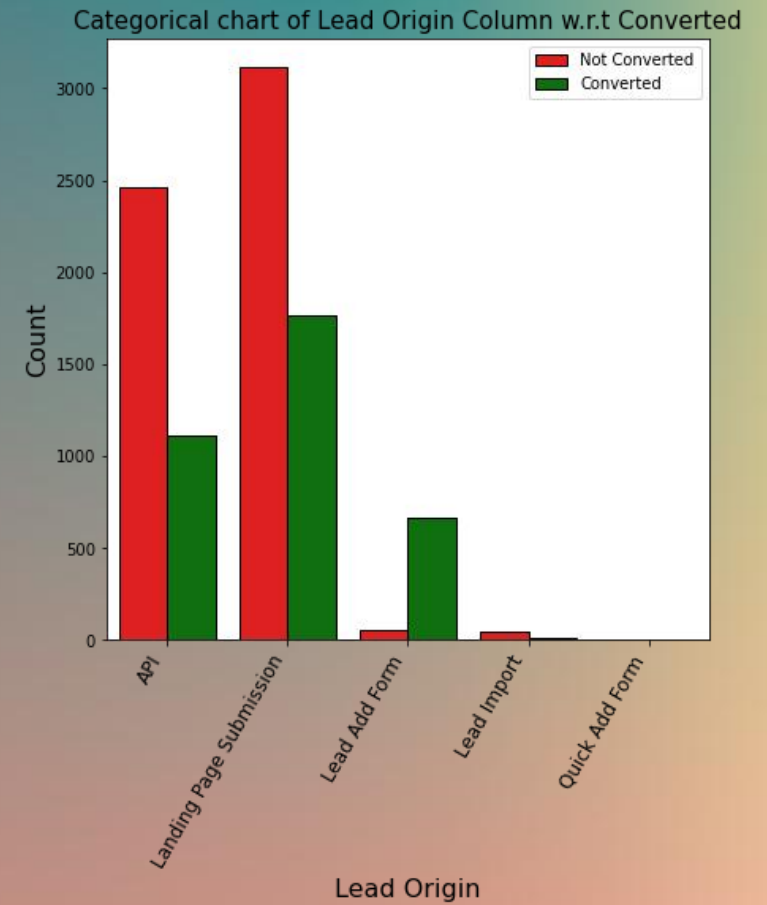
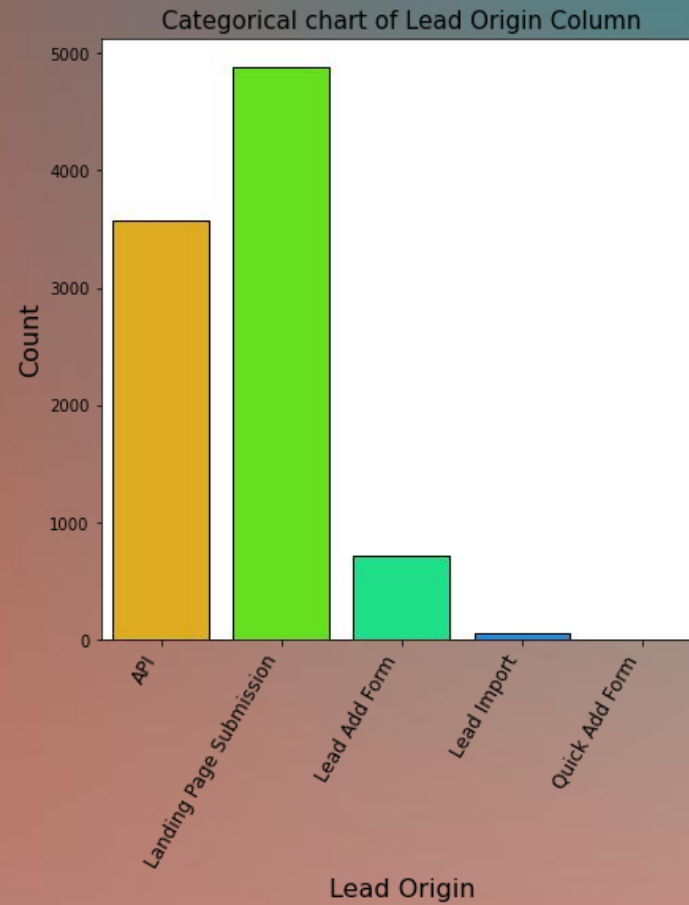
- To help X Education select the most promising leads, i.e., the leads that are most likely to convert into paying customers.
- The company requires us to build a model wherein we need to assign a lead score to each of the leads such that the customers with a higher lead score have a higher conversion chance and the customers with a lower lead score have a lower conversion chance.
- The CEO has given a ballpark of the target lead conversion rate to be around 80%

# DATA CLEANING

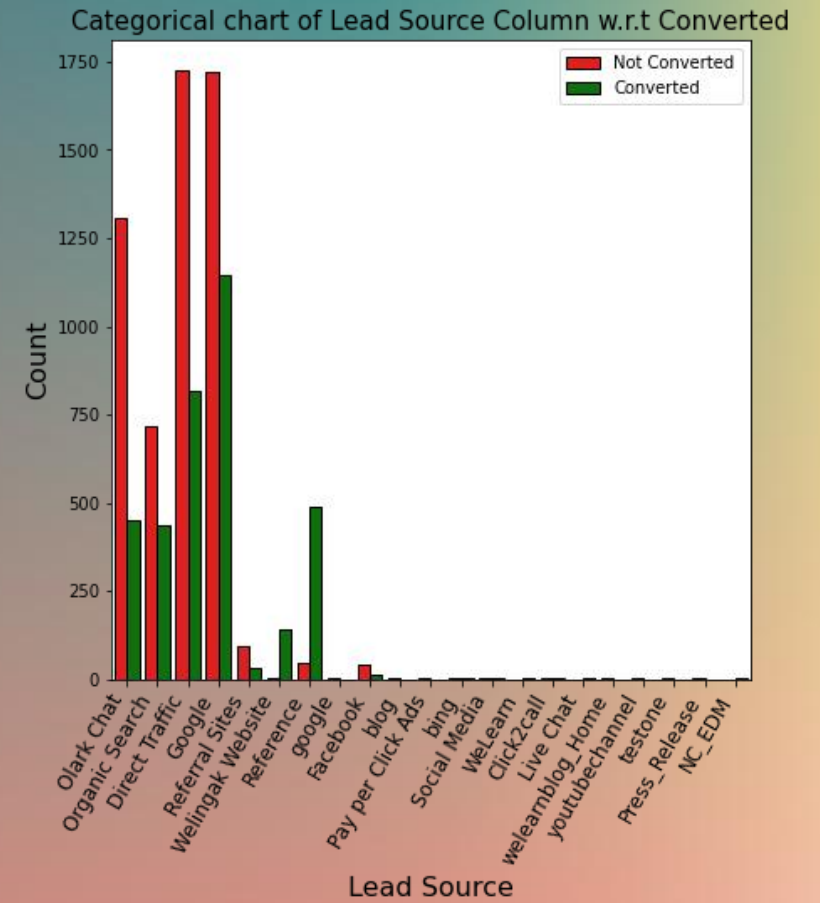
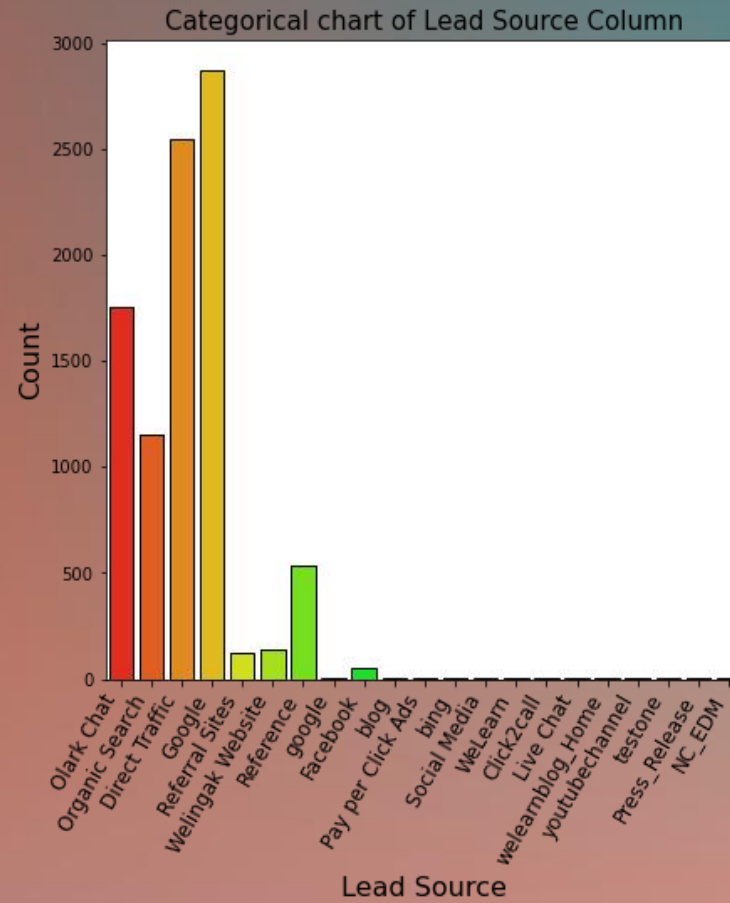
- Dropped the columns with unique values
- The columns with 'Select' value have been nullified since it indicates that people didn't choose any option.
- Dropped the columns having null values greater than 40%.
- Outliers are examined and dealt with missing values.
- Identifying the potential data columns which can factor in for accurate prediction
- Identifying the relationship and distribution of column data using graphs

# EDA

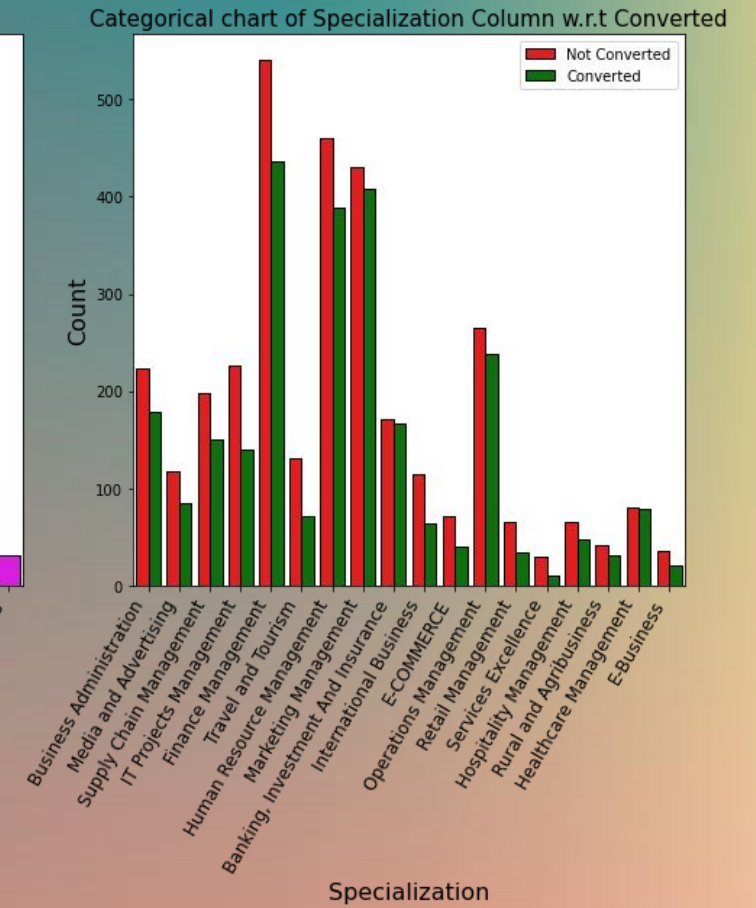
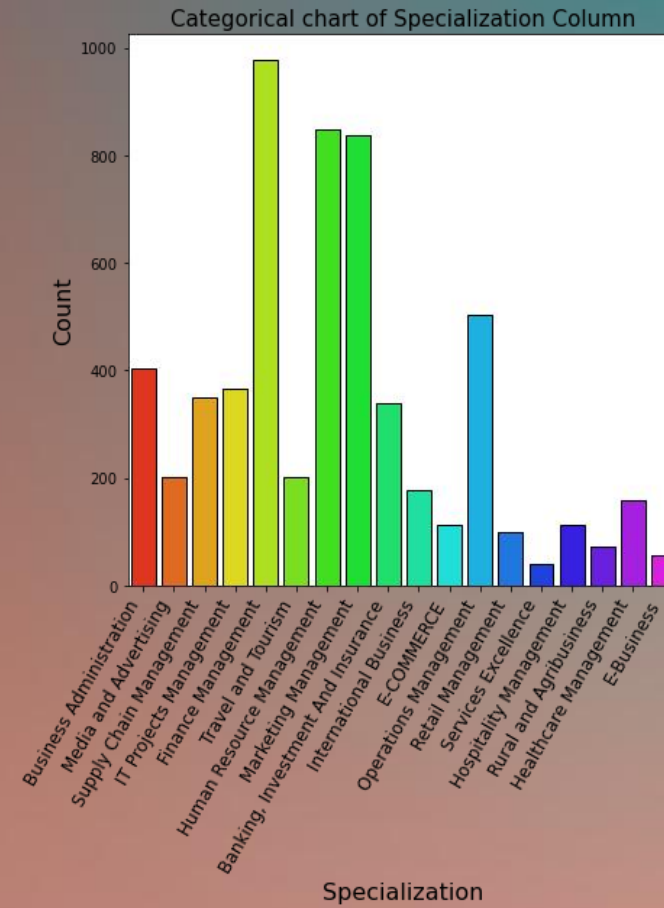
- Categorical chart of lead origin w.r.t conversion



- Categorical chart of lead source w.r.t conversion

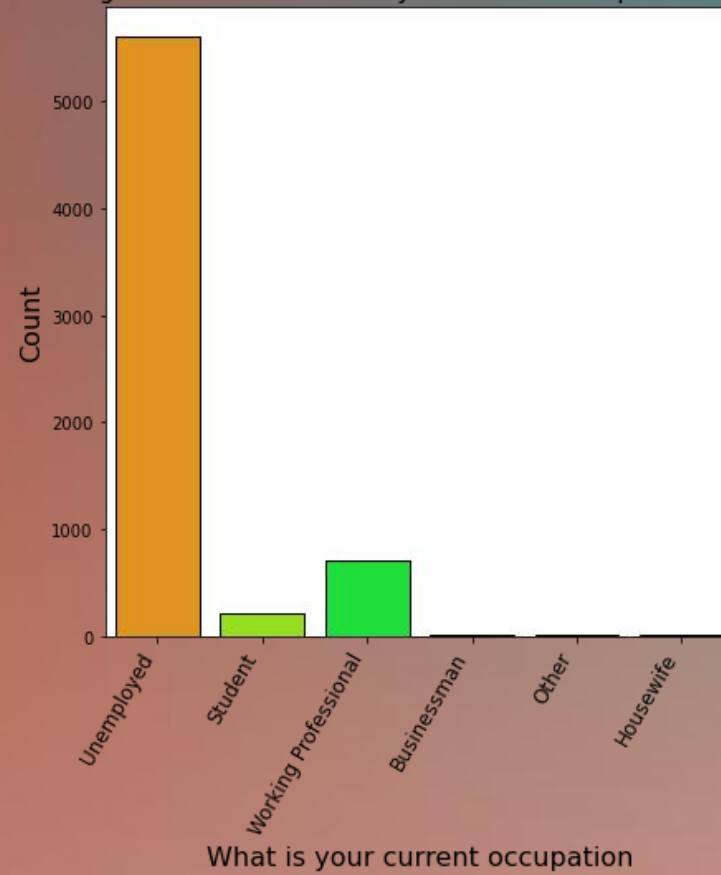


- Categorical chart of specialization w.r.t conversion

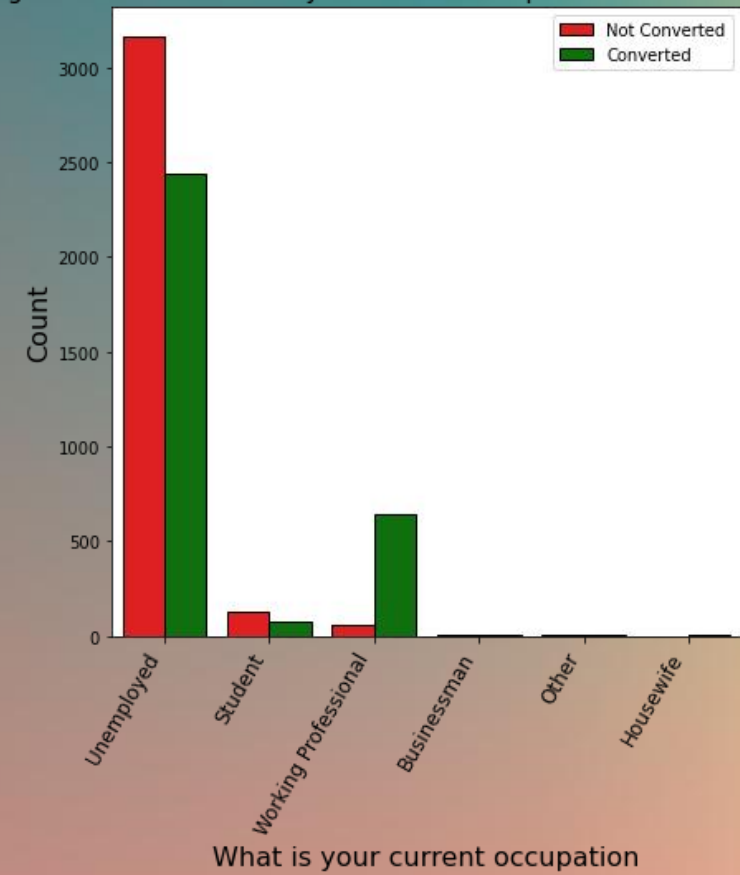


- Categorical chart of specialization w.r.t conversion

Categorical chart of What is your current occupation

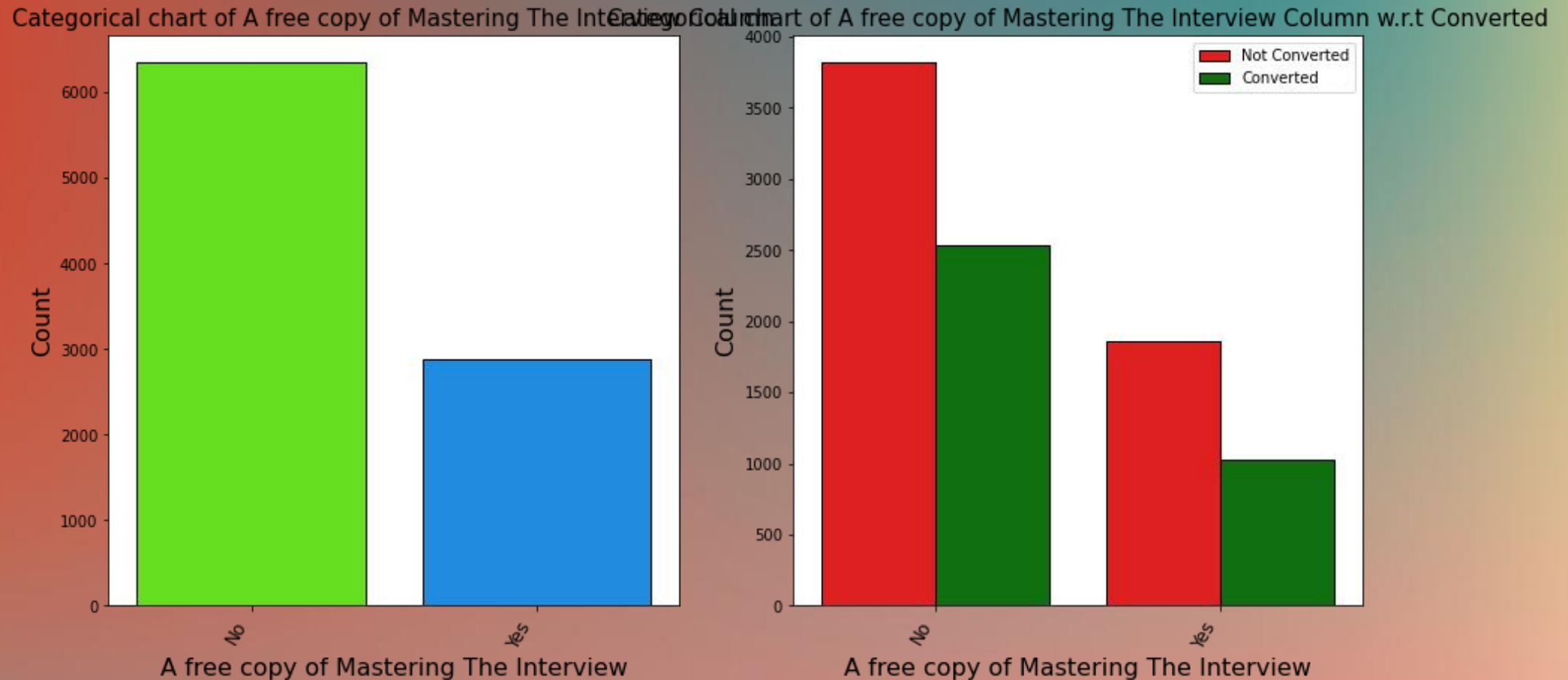


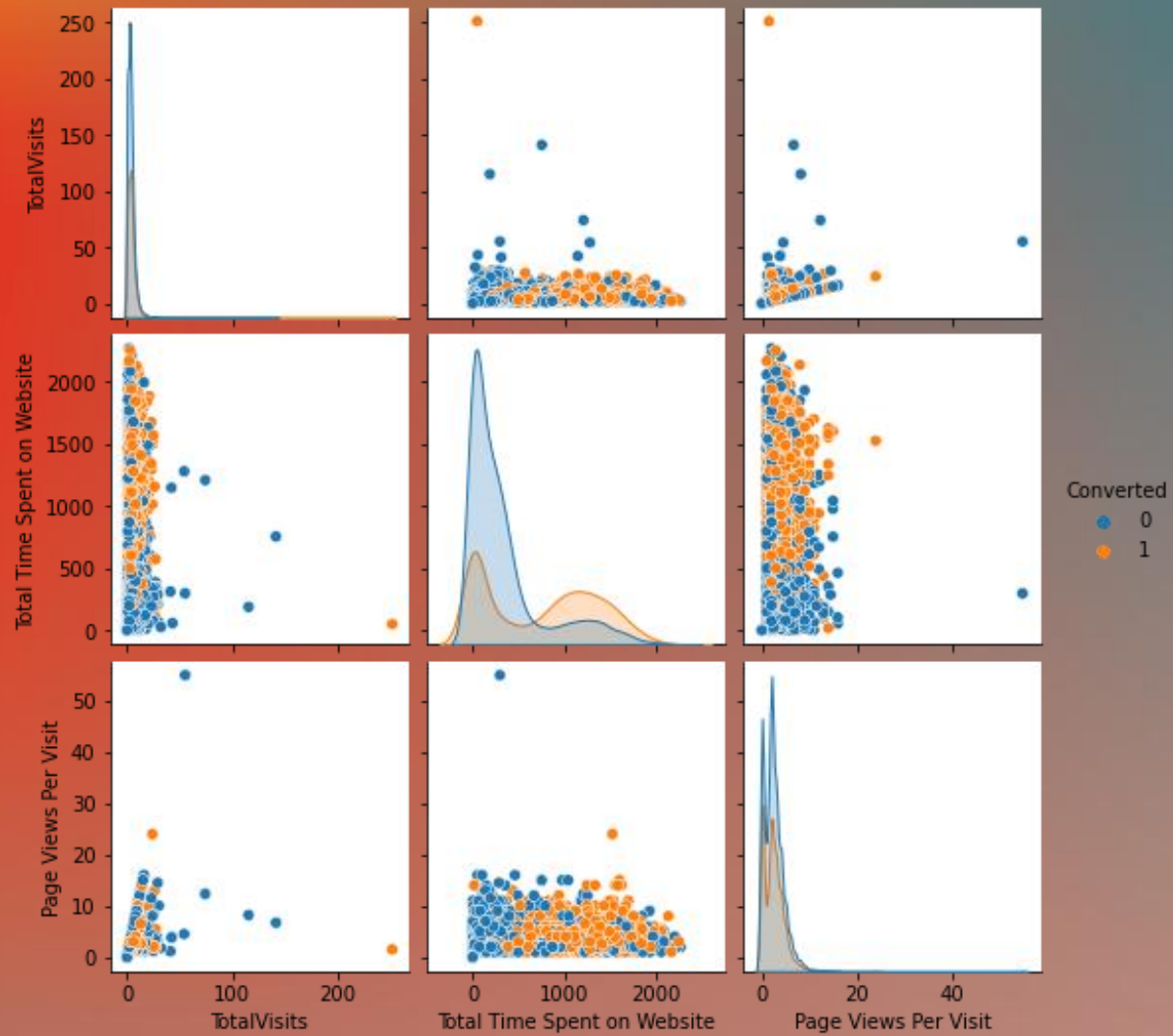
Categorical chart of What is your current occupation Column w.r.t Converted



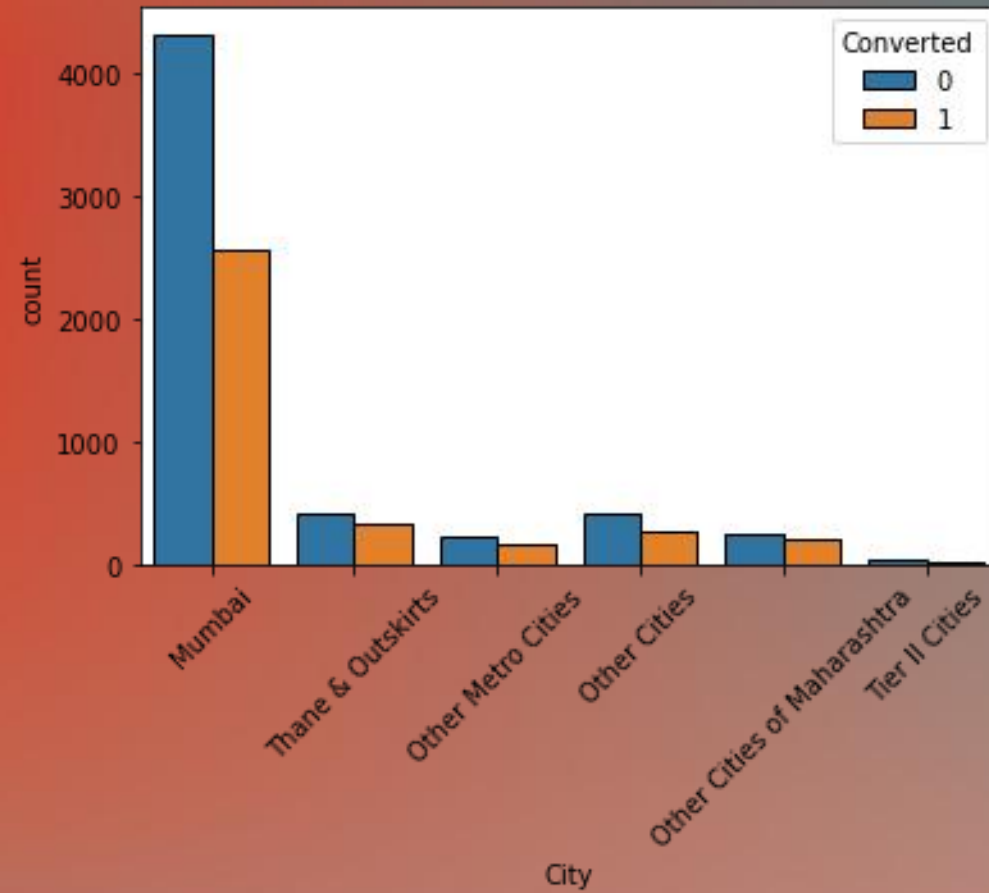


- Categorical chart of A free copy of mastering the interview w.r.t conversion

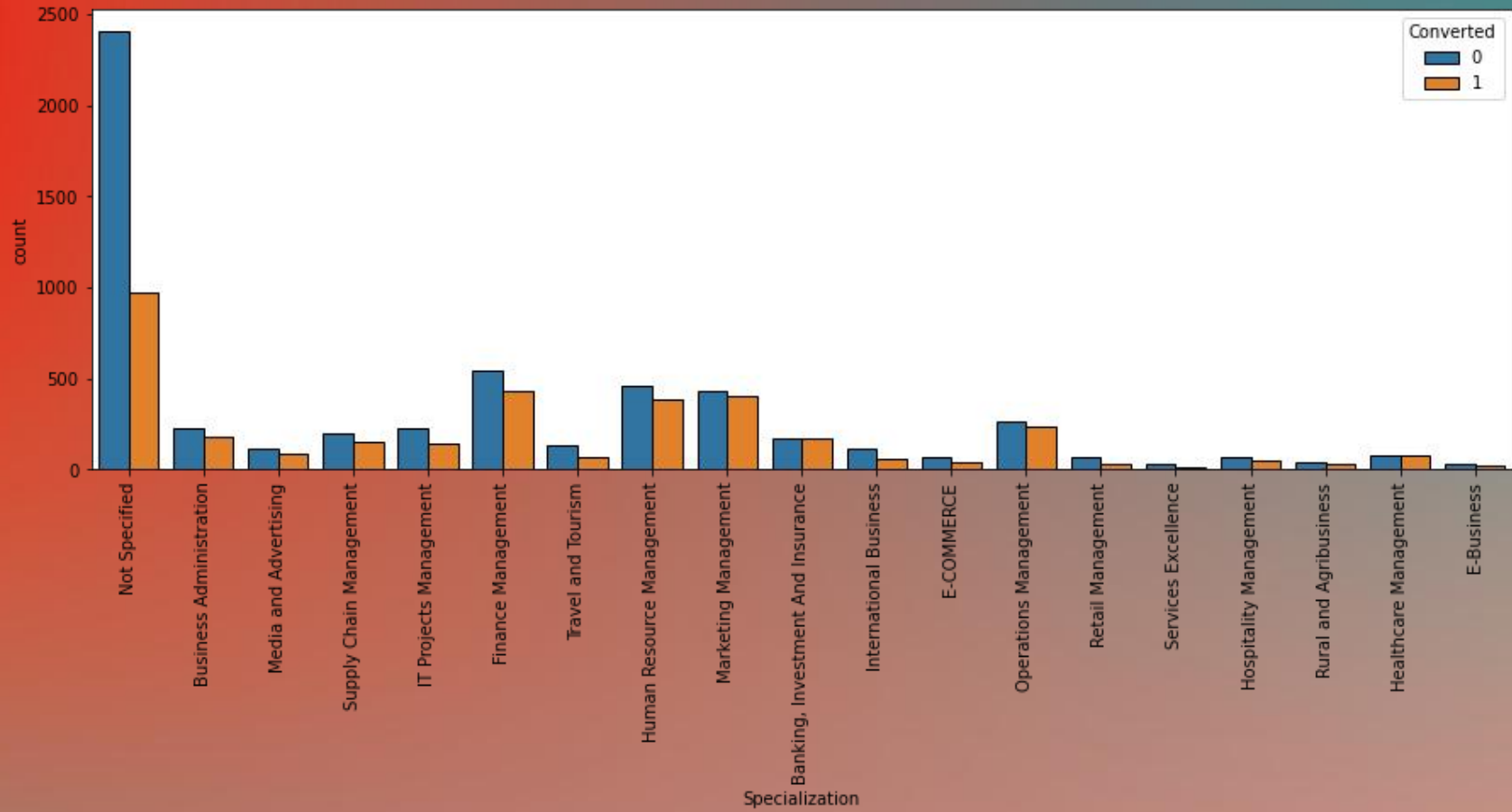




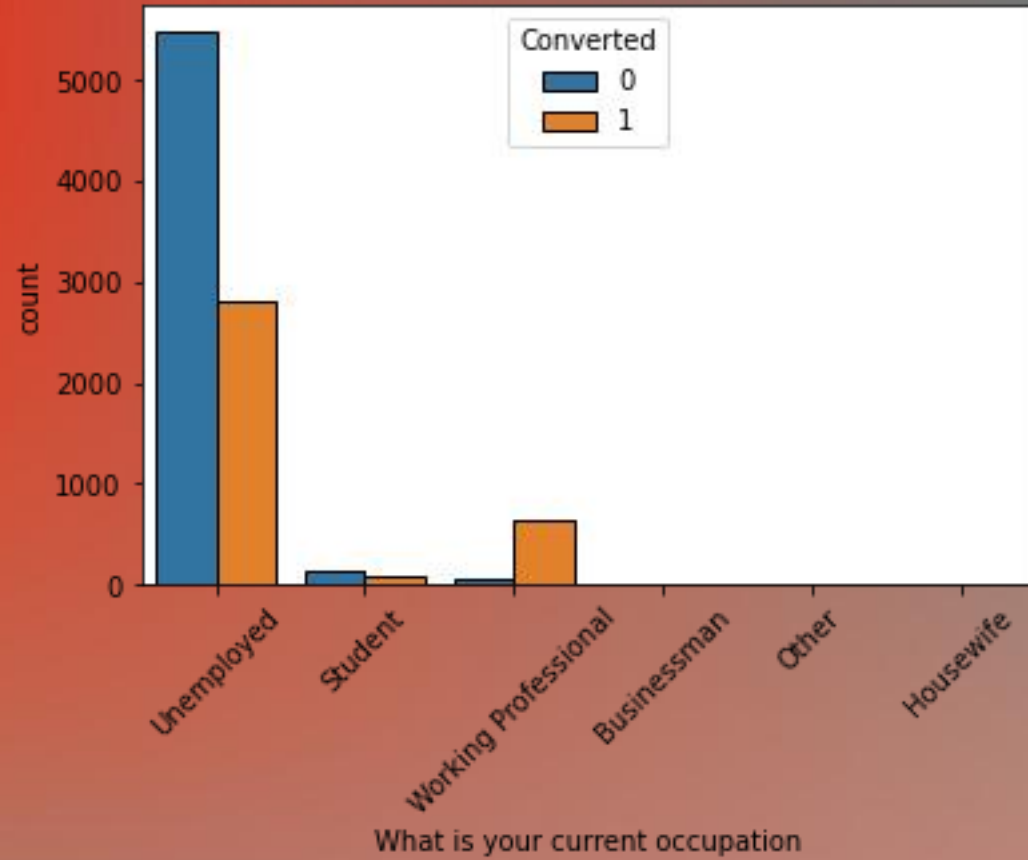
- Pair plot of conversions with different categories



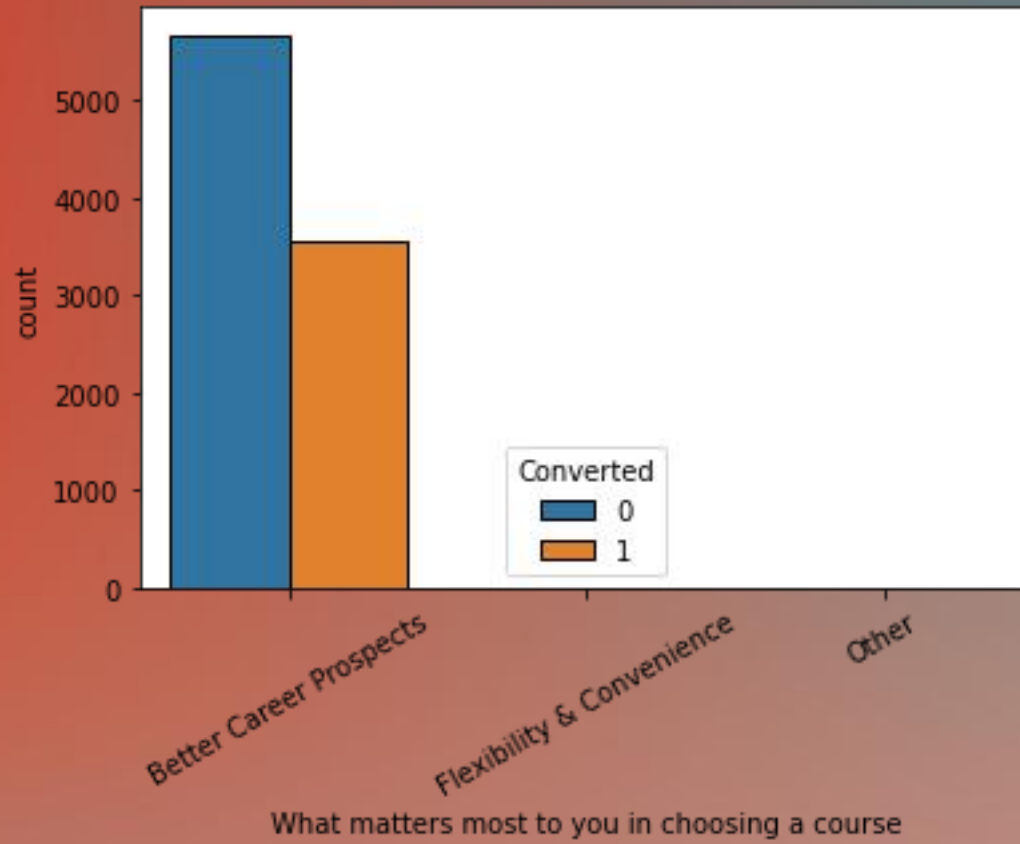
- Categorical chart of conversion from different cities



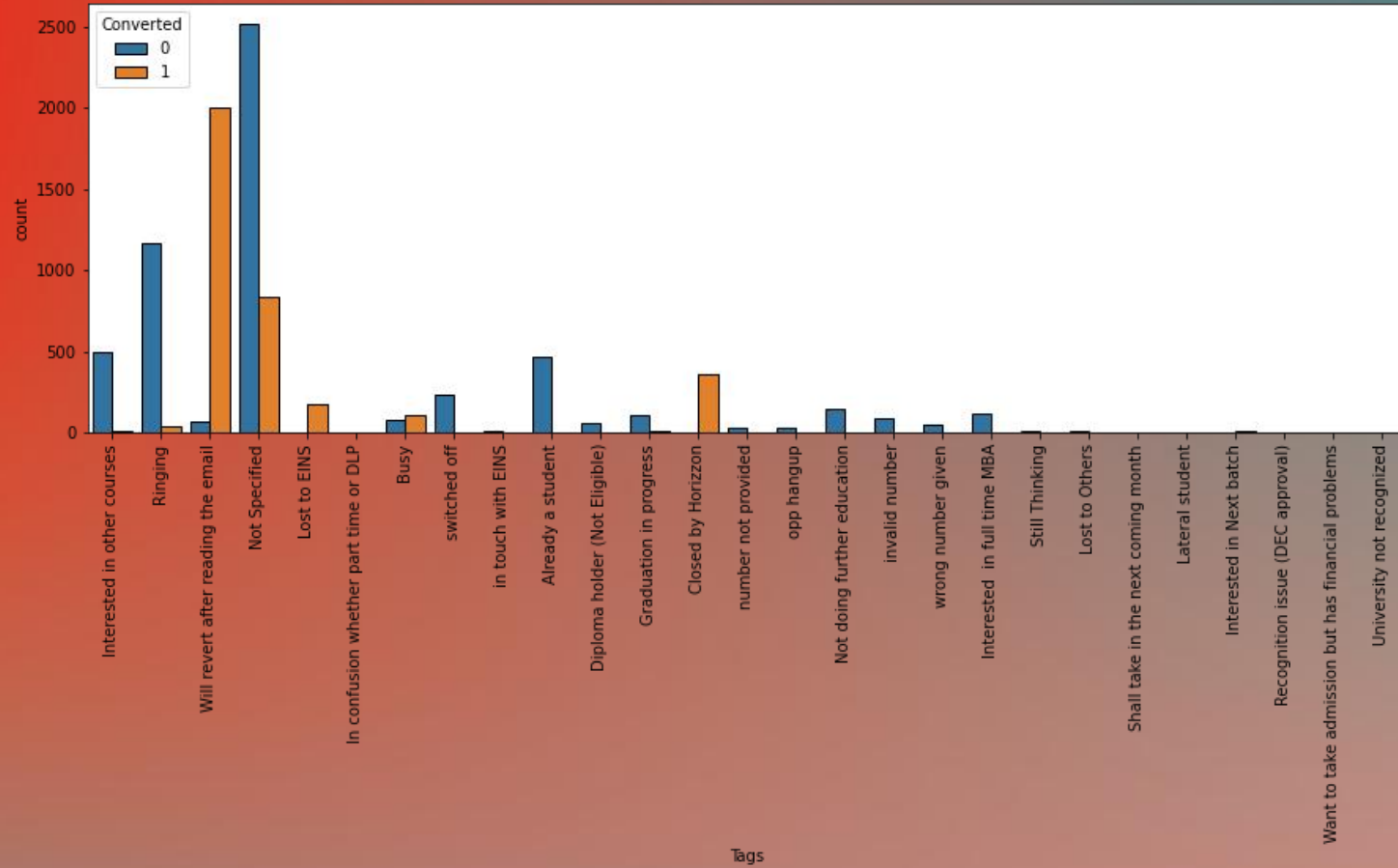
- Categorical chart of specializations converted or not.



Categorical chart of occupation of people if converted or not.

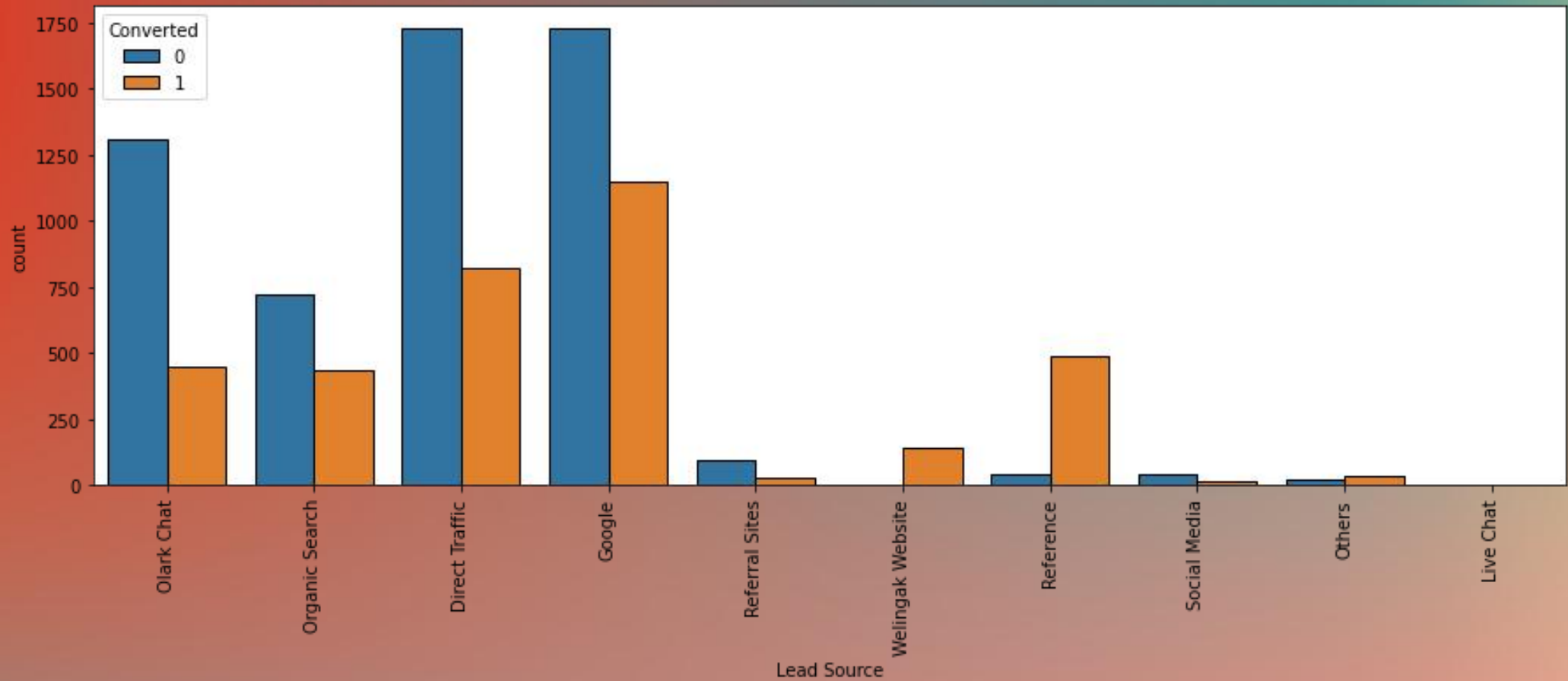


- Categorical chart of purpose of choosing a course



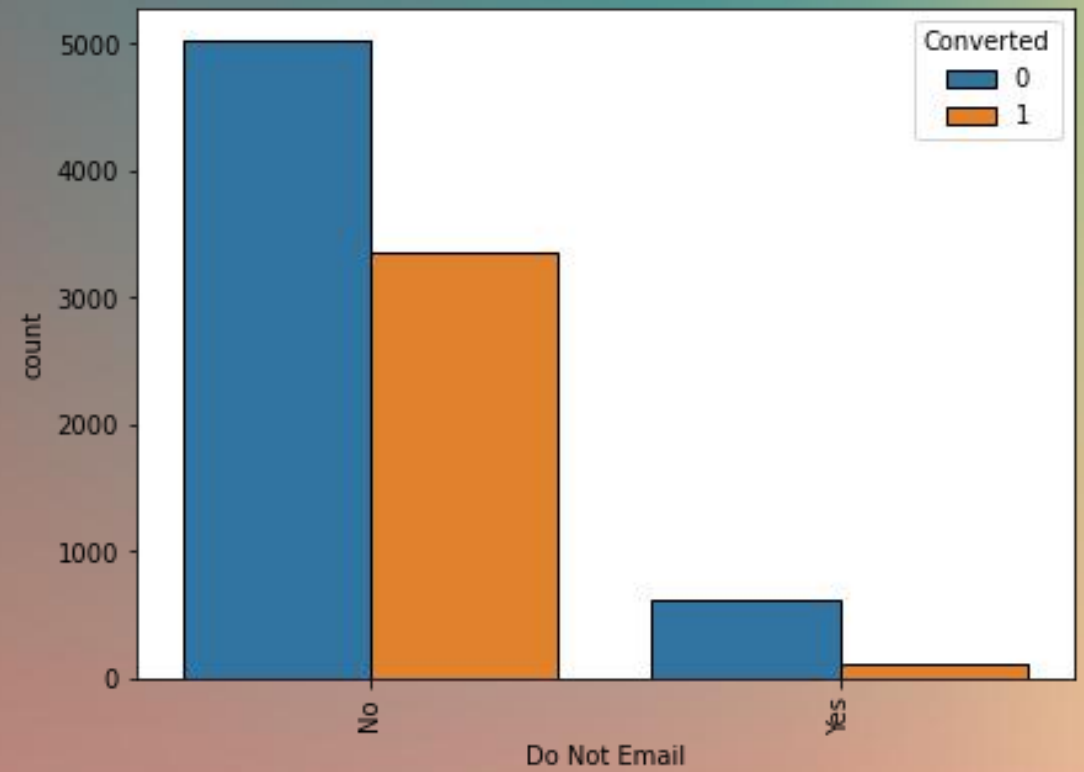
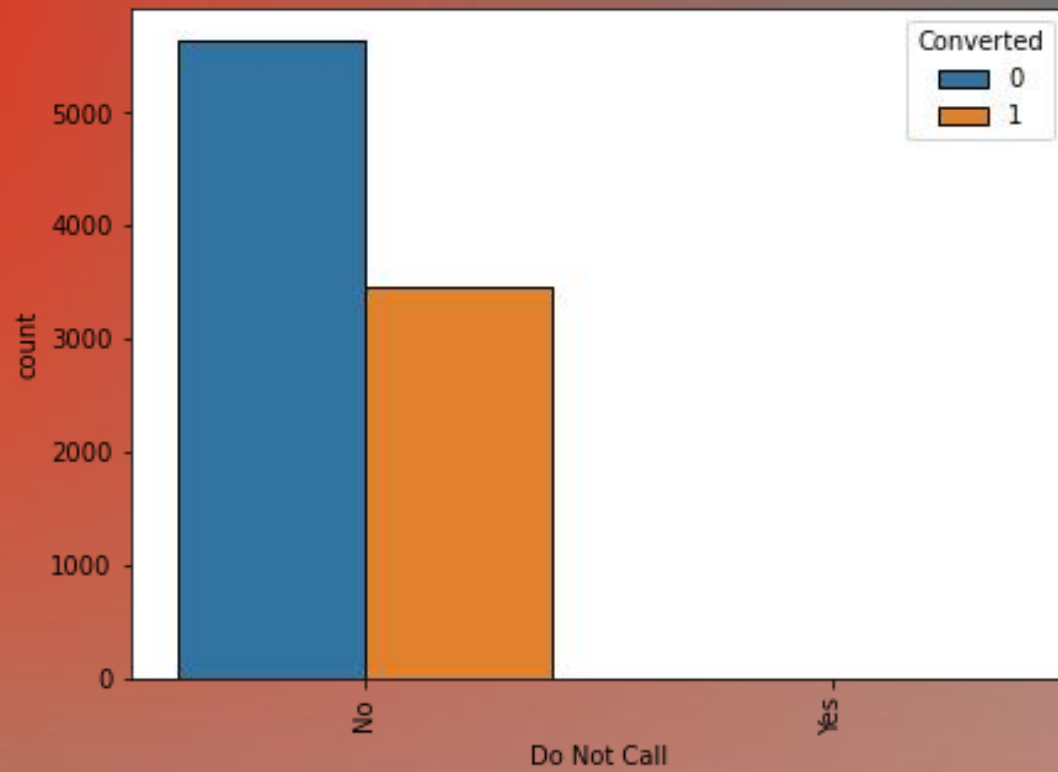
- Categorical chart of tags getting converted or not

- Categorical chart of lead sources getting converted or not

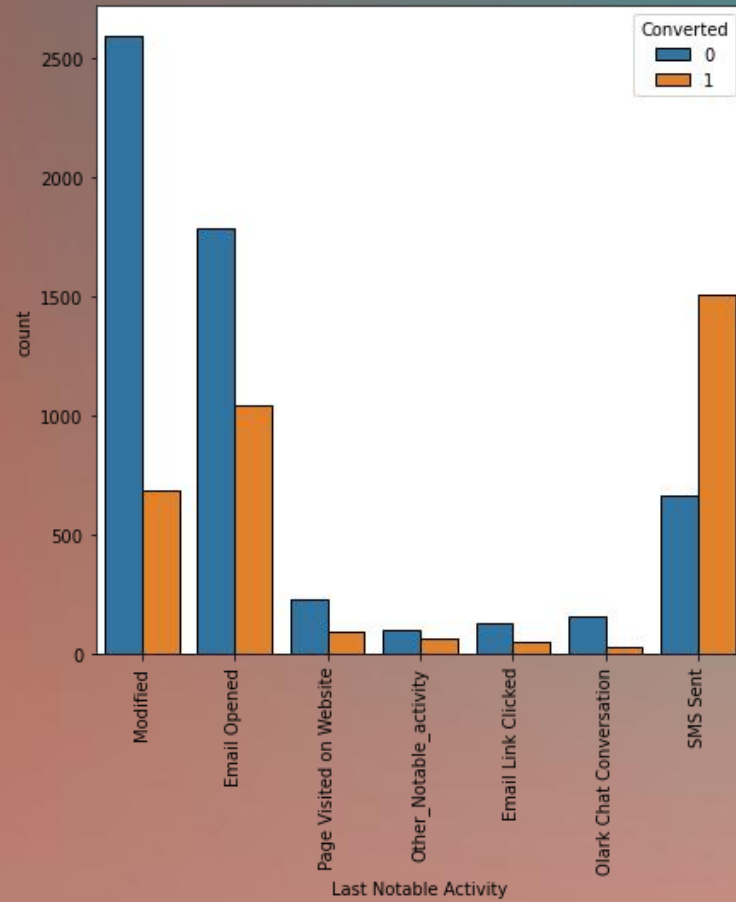




- Categorical chart of do not call and do not email tags getting converted



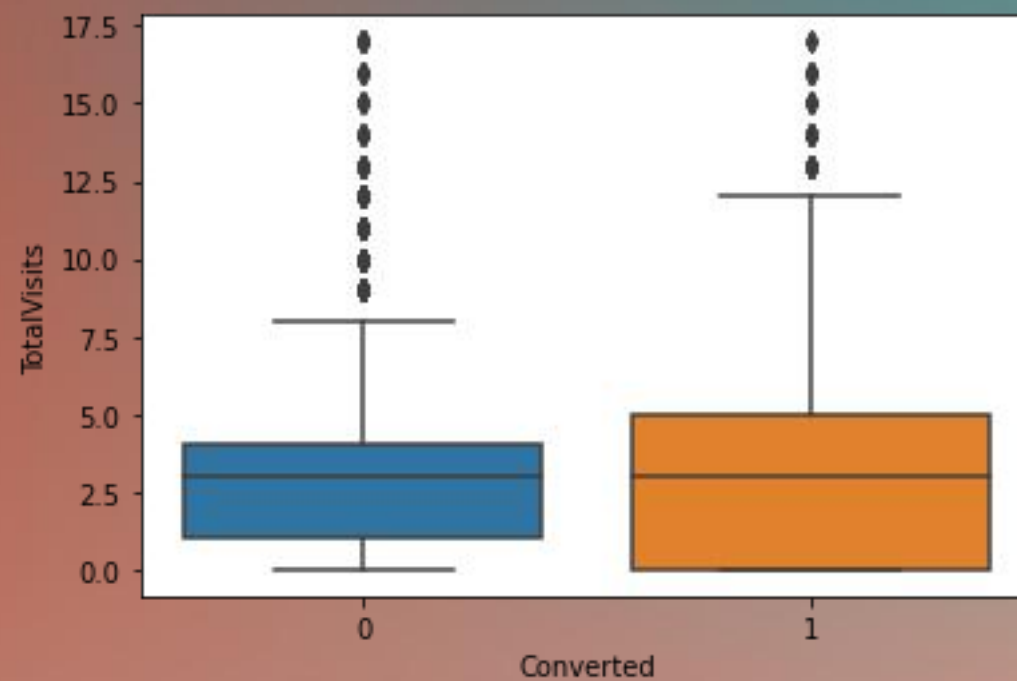
- Categorical chart of last notable activity getting converted



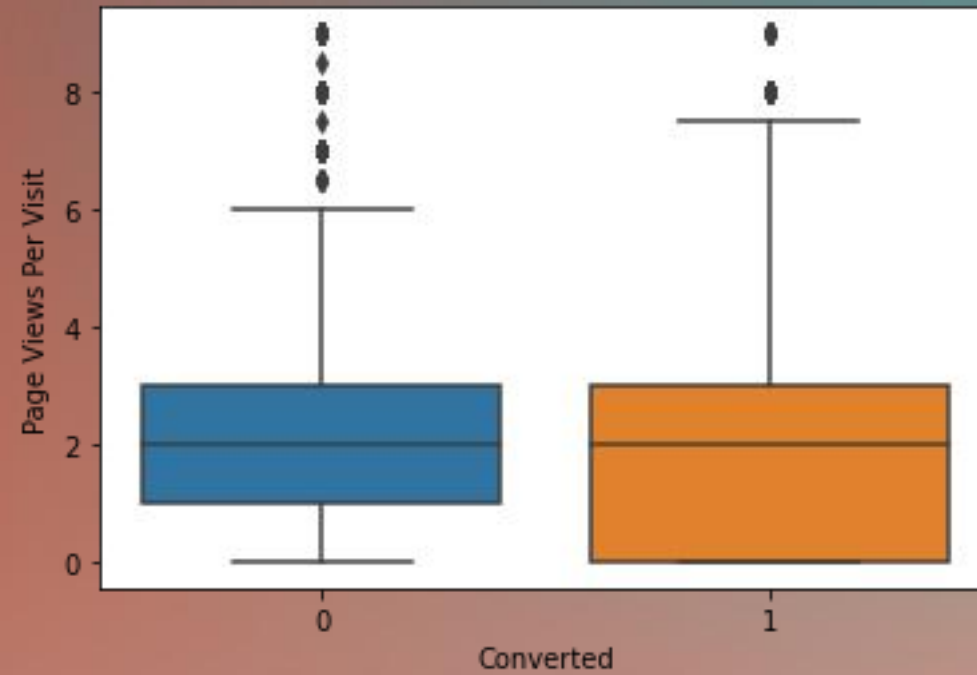
- Correlation between features



- box plot of total visit and conversions

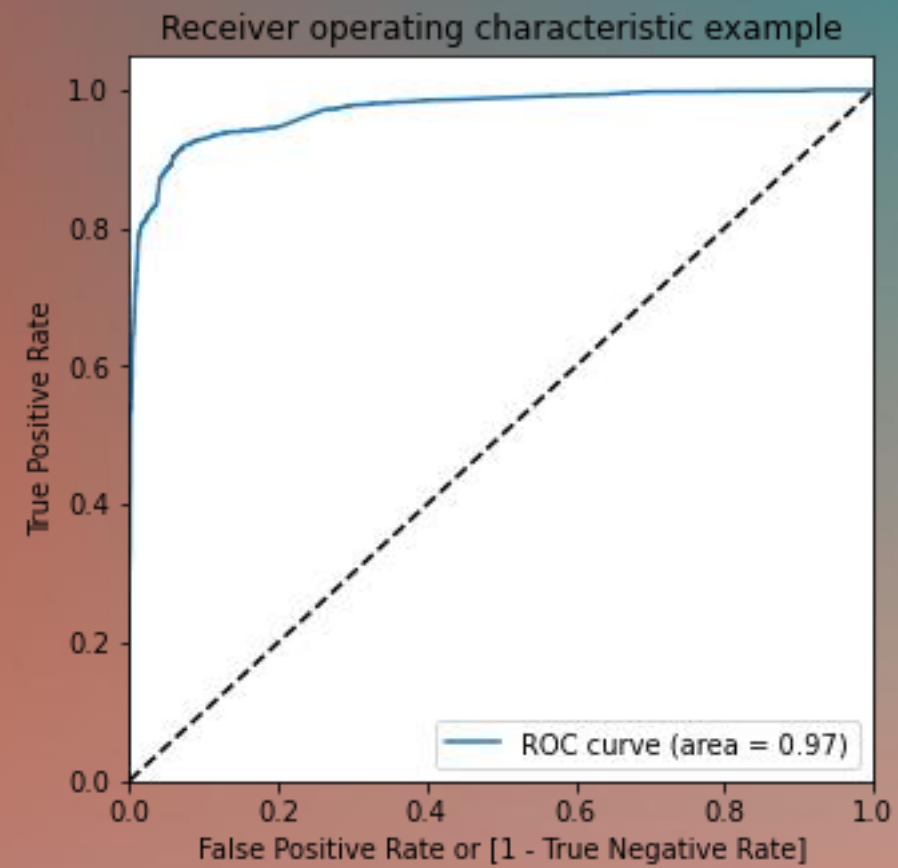


- box plot of page views per visit and conversions

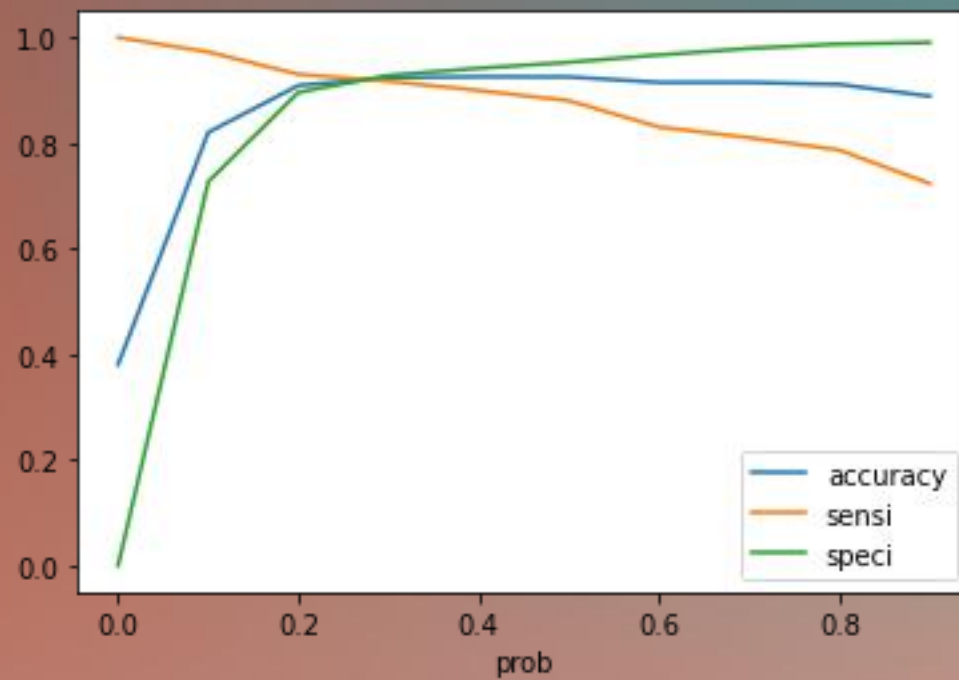




- ROC curve

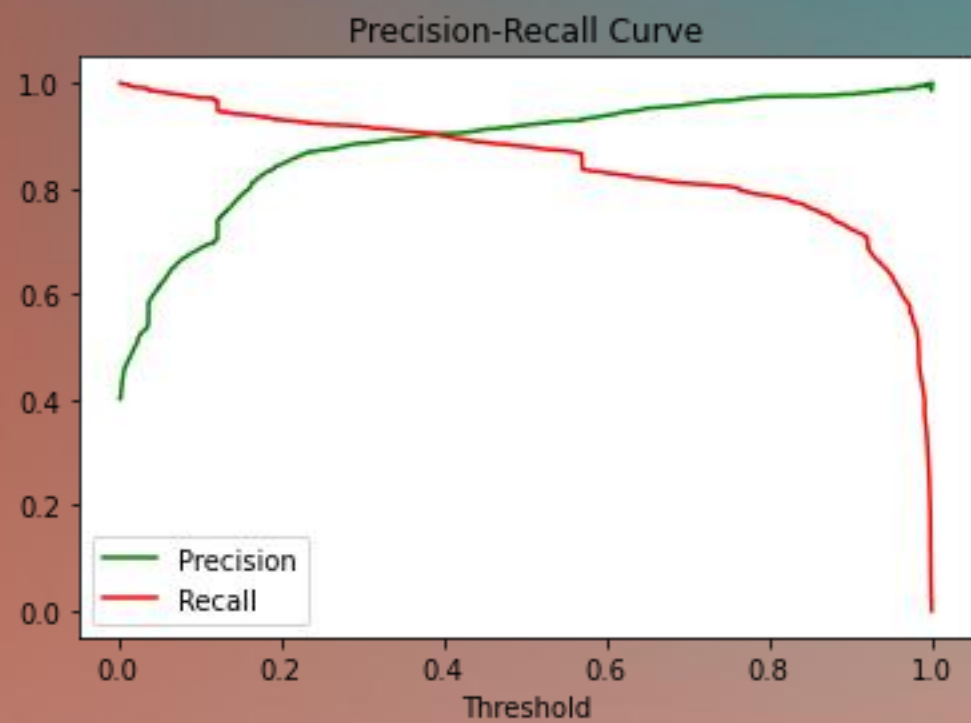


- Finding optimal cutoff point





- Precision and Recall Curve



# CONCLUSION

- More conversions result from lead add forms and landing page submissions.
- More leads from SMS and email marketing convert.
- Increased consumer engagement and conversions can be achieved by lowering website bounce rates.
- Use of Lead Add Form in important places will provide qualified leads.
- For improved conversion rates, the sales team should concentrate on working professionals.
- Lead scores  $>0.35$  have a higher conversion rate and a 91% model accuracy score.