

Lead Score Case Study

Harshal Dubey

Problem Statement

- X Education sells online courses to industry professionals.
- X Education gets a lot of leads, its lead conversion rate is very poor.
- 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 Objectives

- Company wants to know most potential leads.
- 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.
- Deployment of model for future use.

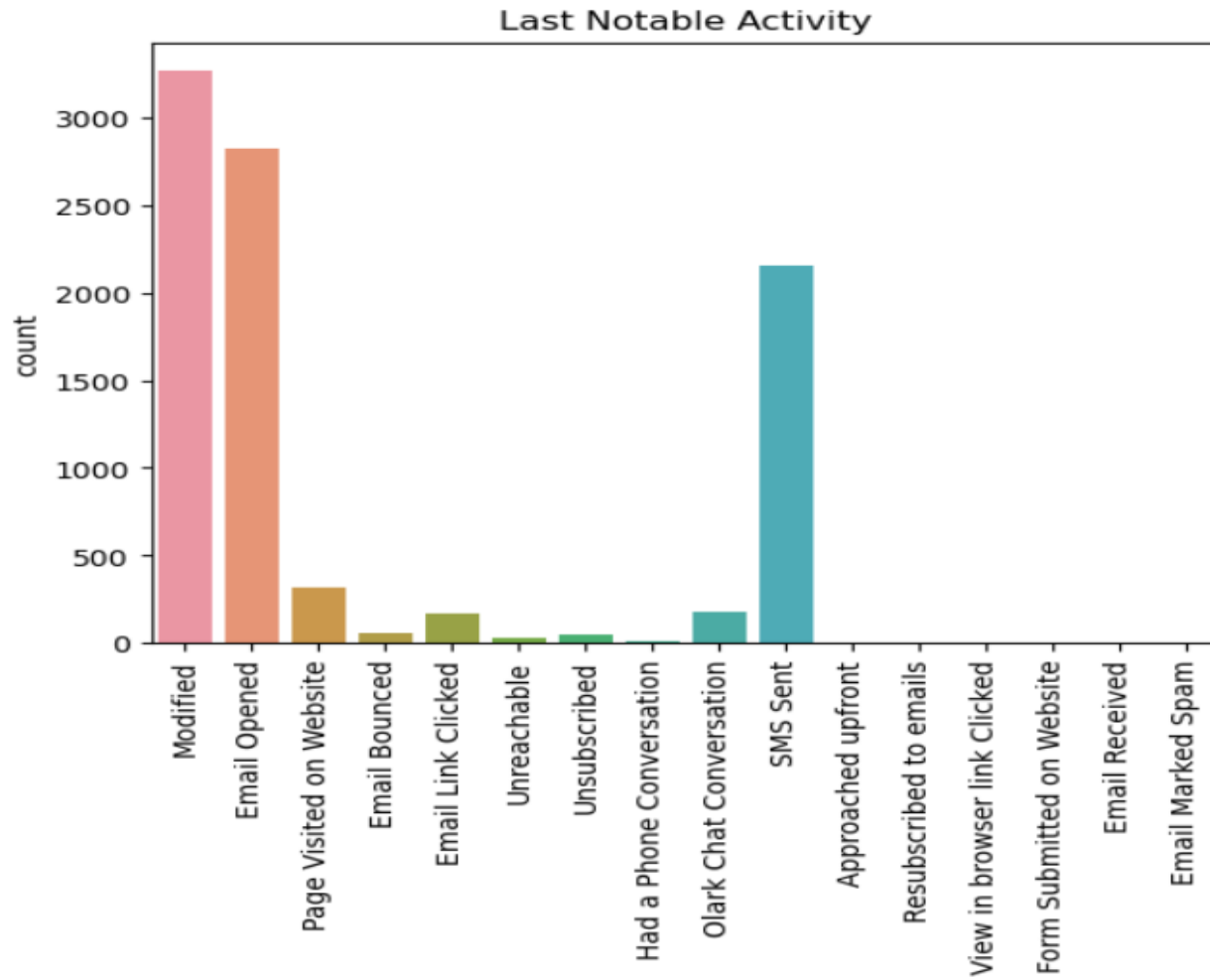
Solution Methodology

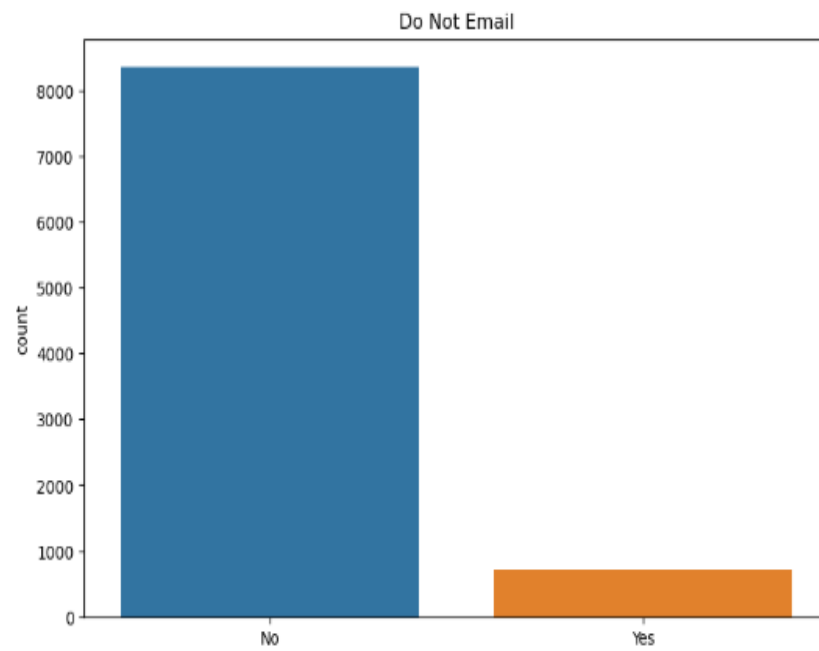
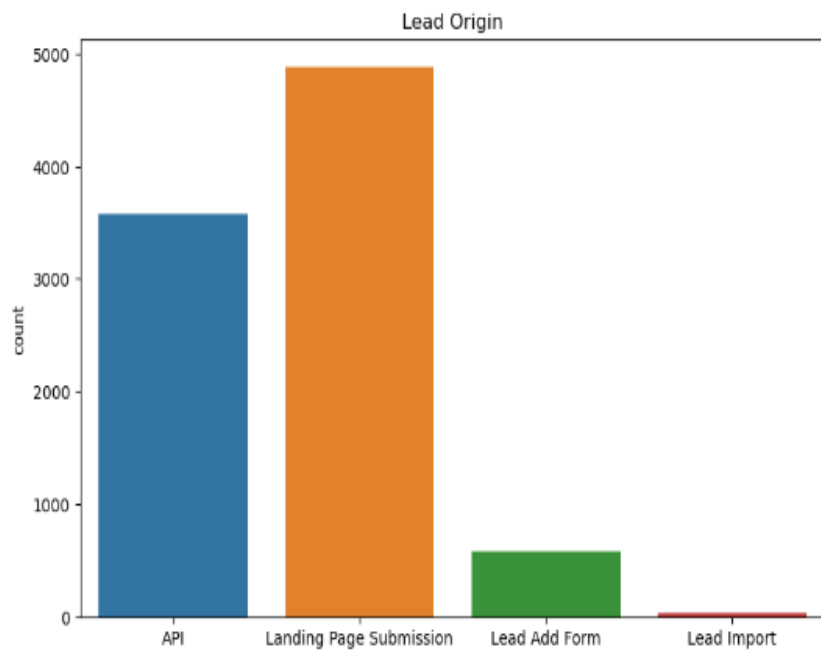
- Data Cleaning and Manipulation
 1. Check and handle duplicate data
 2. Handle missing values.
 3. Drop columns with large number of missing values and not useful for analysis.
 4. check for outliers.
- EDA
 1. Univariate data analysis
 2. Bivivariate data analysis
- Feature scaling and dummy variables and encoding of data.

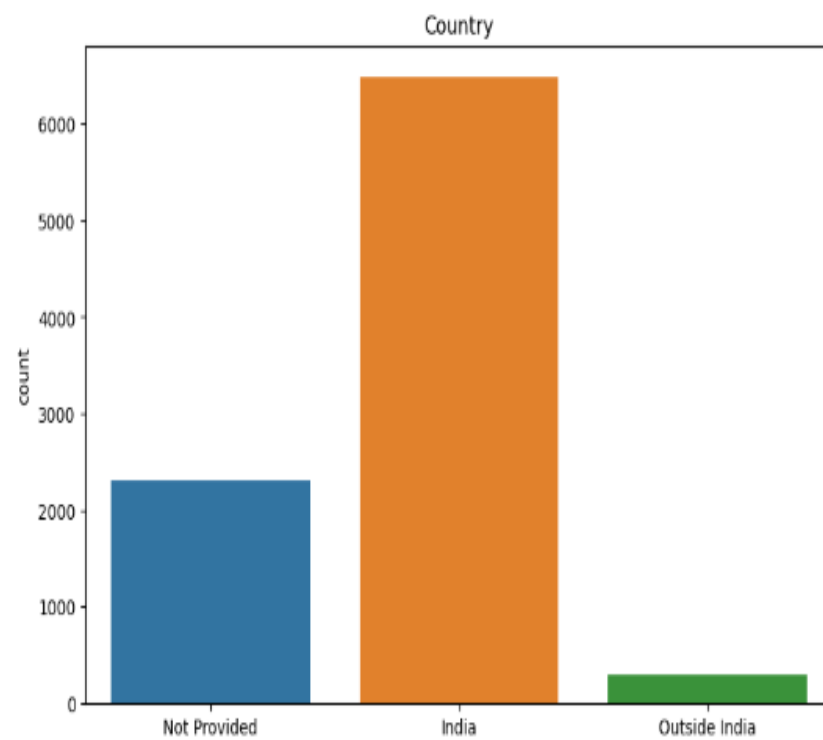
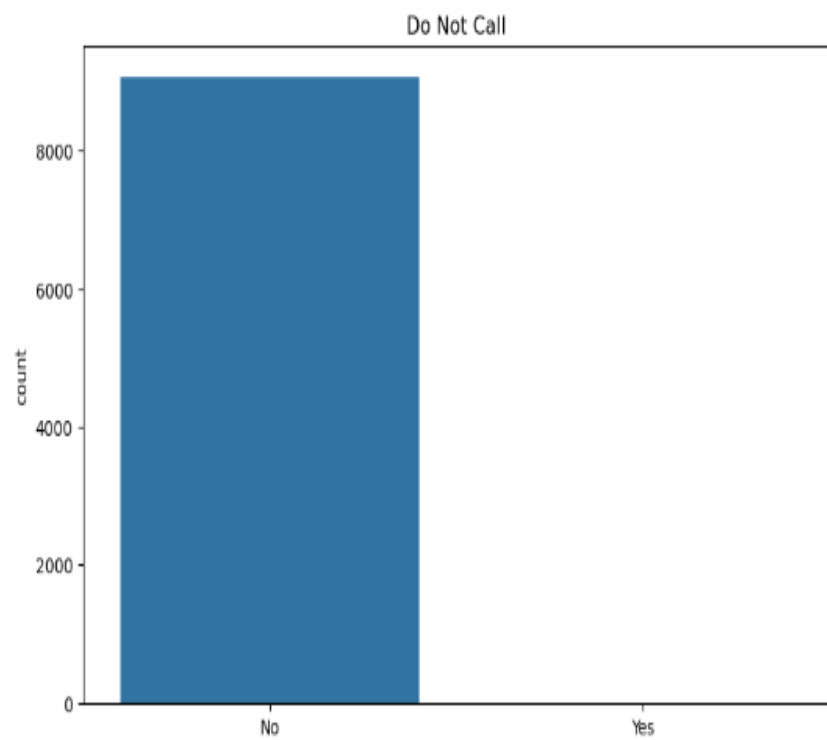
Data Manipulation

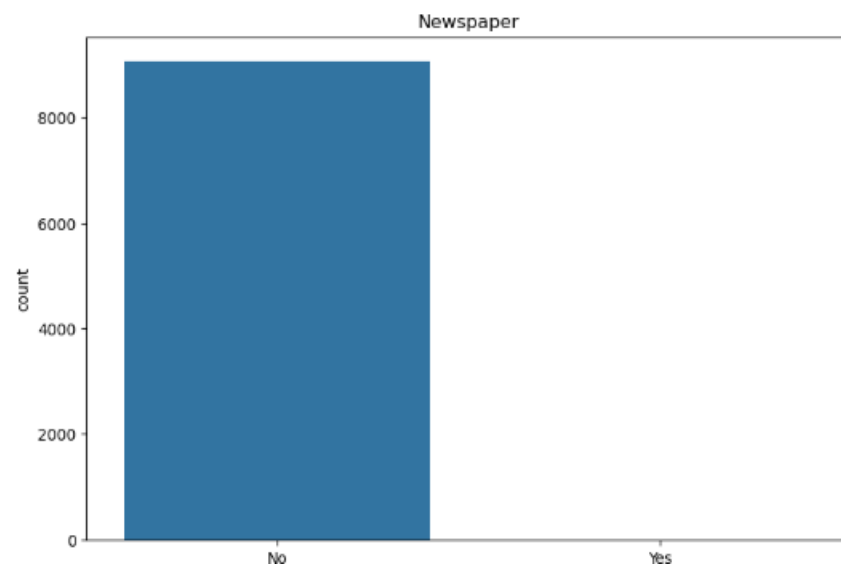
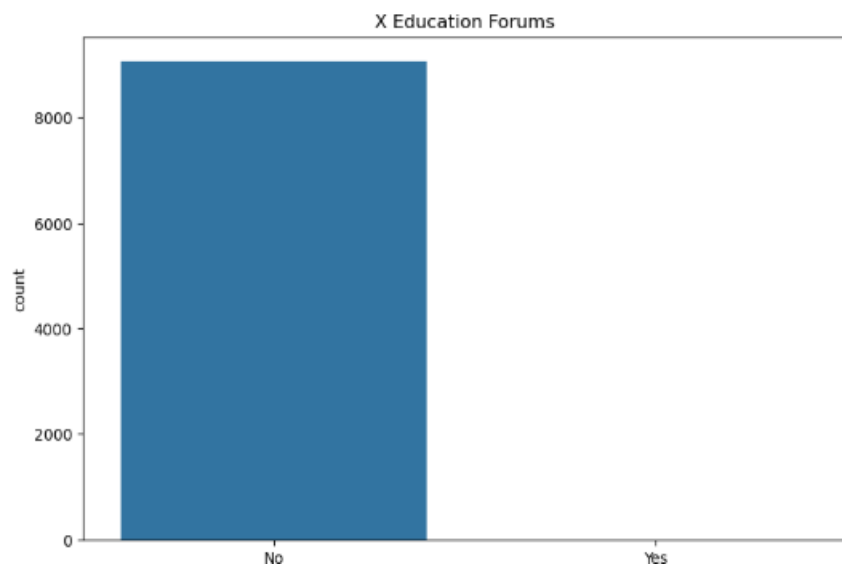
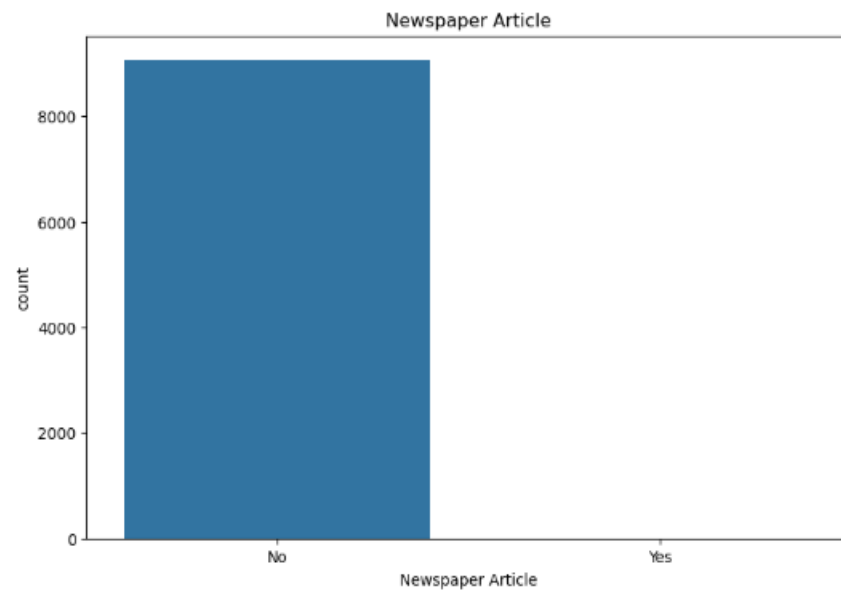
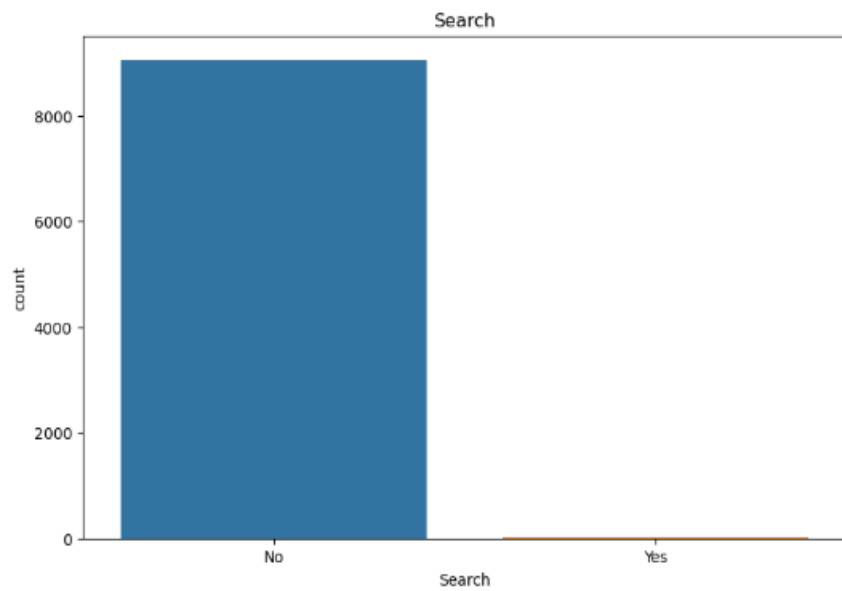
- Total number of rows – 9240 and columns – 37
- Columns with single values are dropped like 'Magazine', 'Receive More Updates About Our Courses', 'I agree to pay the amount through cheque', 'Get updates on DM Content', 'Update me on Supply Chain Content'.
- Dropped the columns having more than 40% of missing values.
- Categorized values in Country column, as Not provided count is more and single values of other countries in 3 categories as India/outside India/Not provided.

EDA

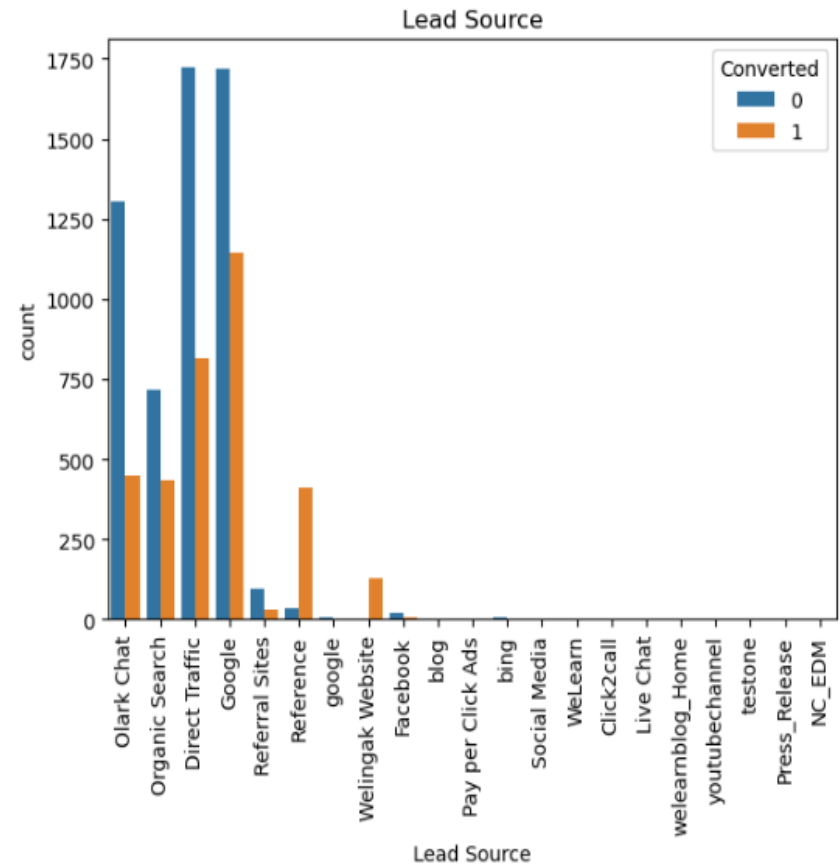
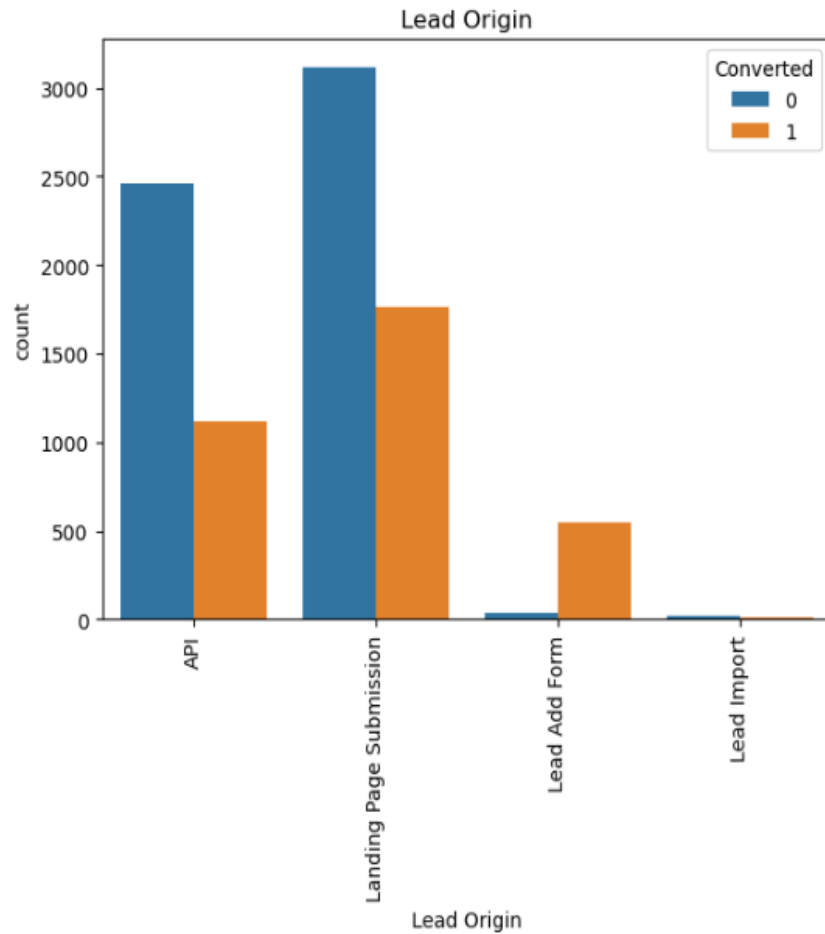


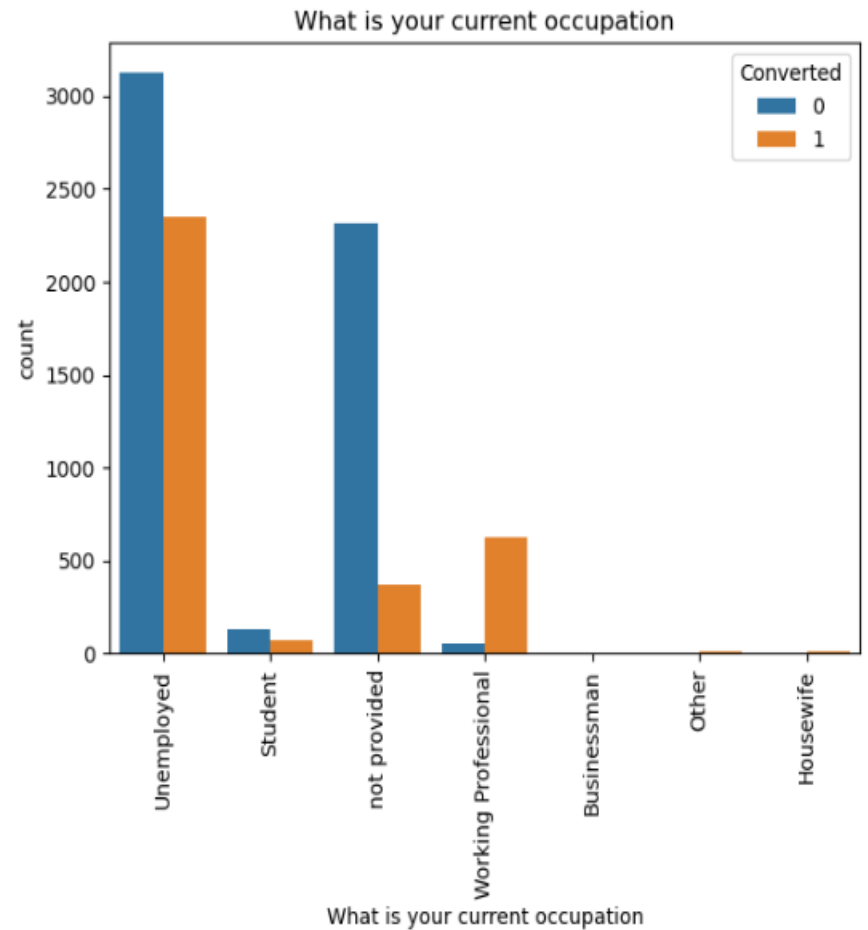
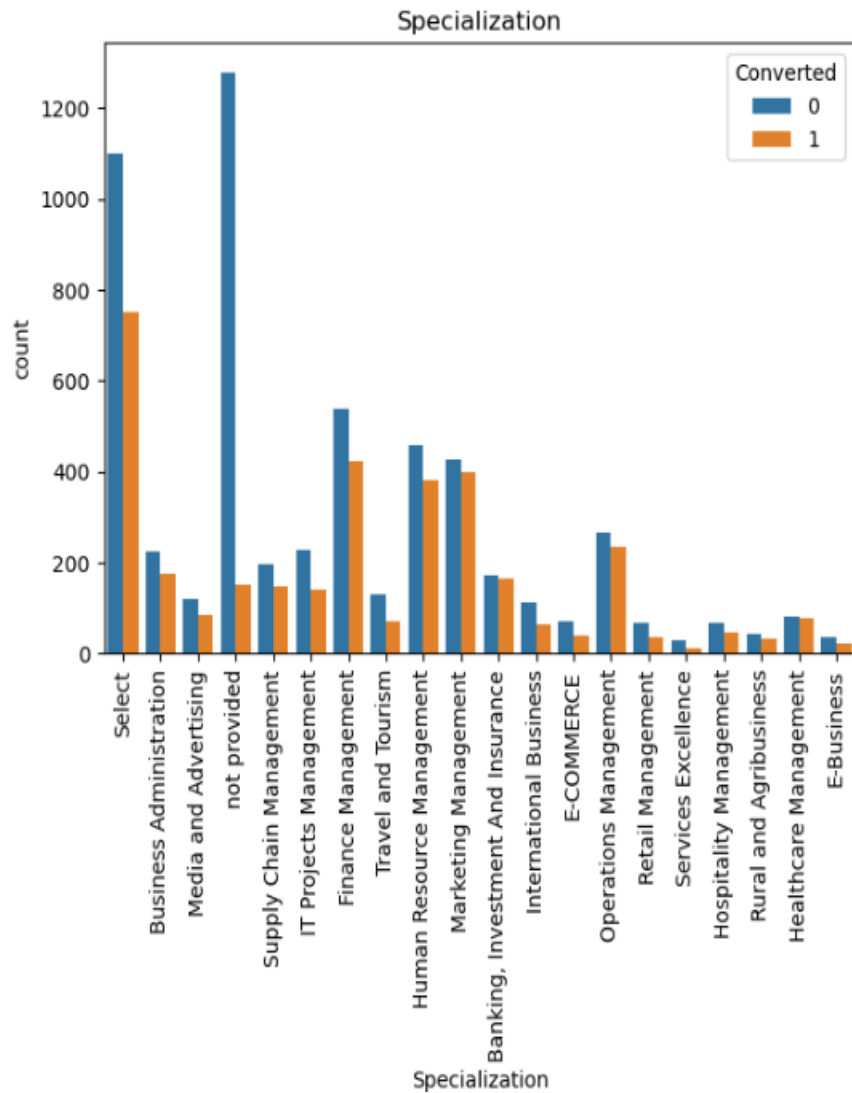


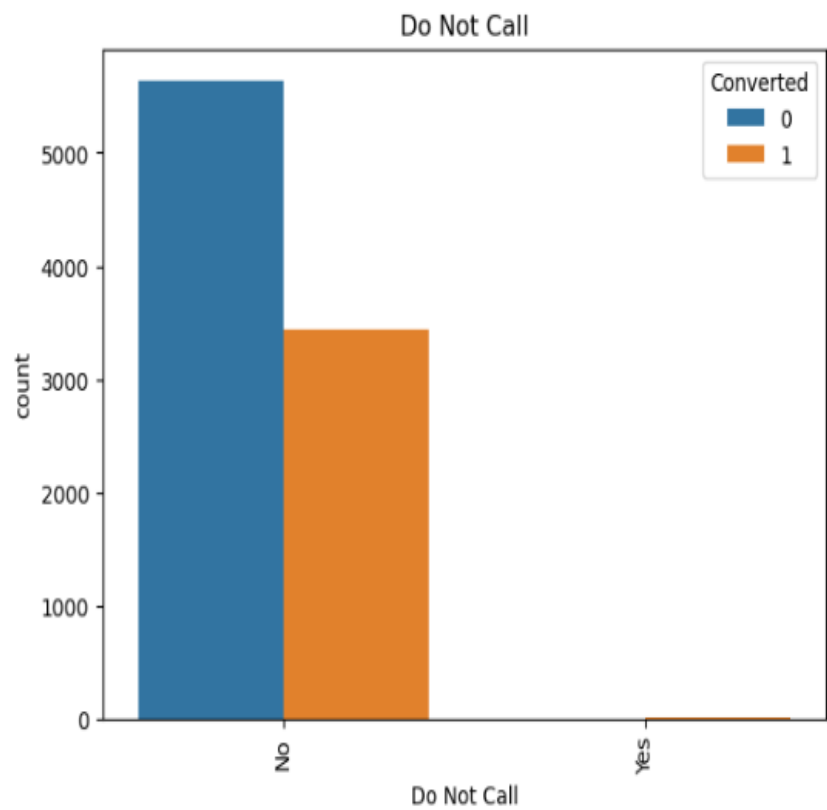
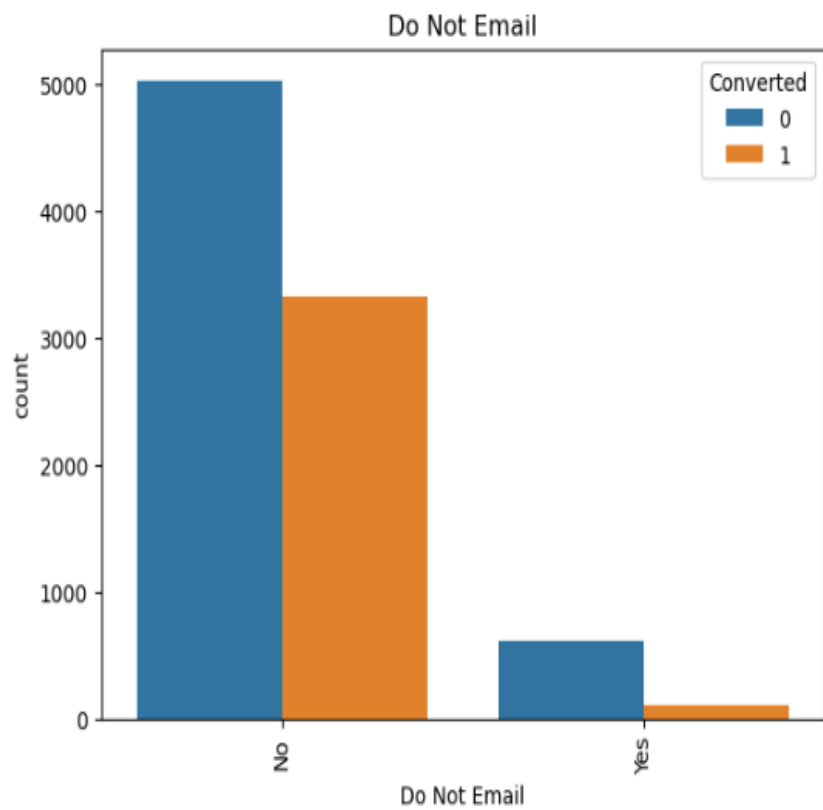


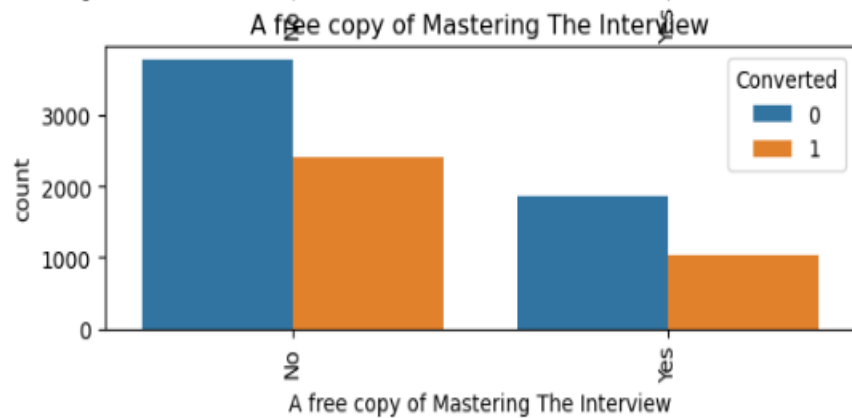
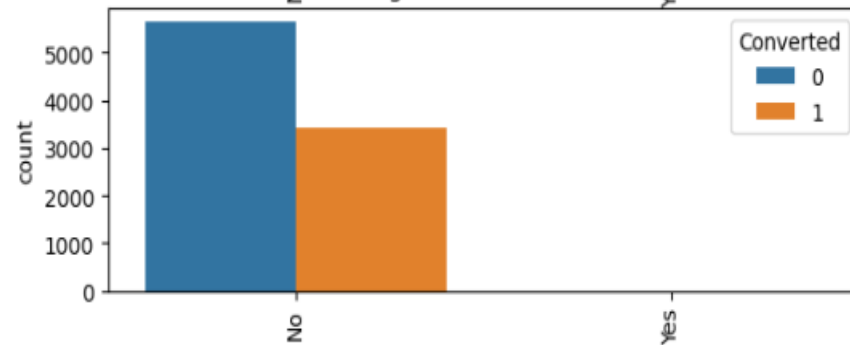
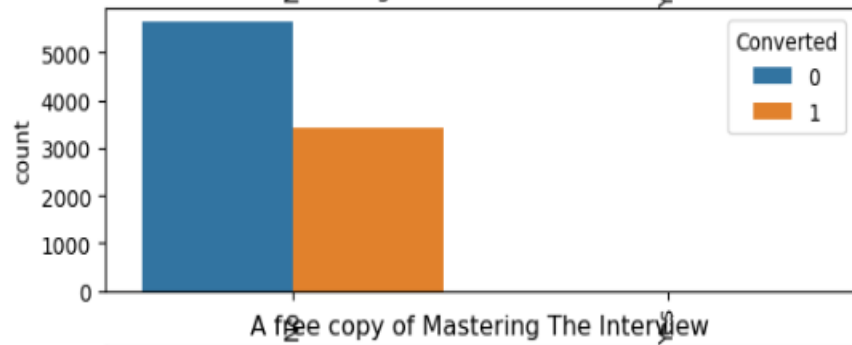
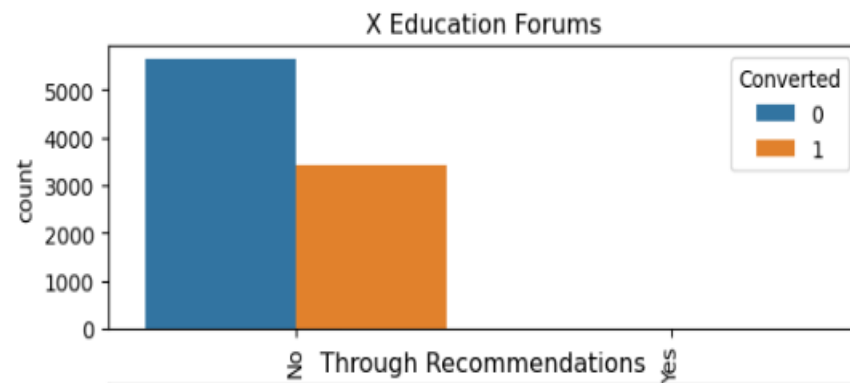
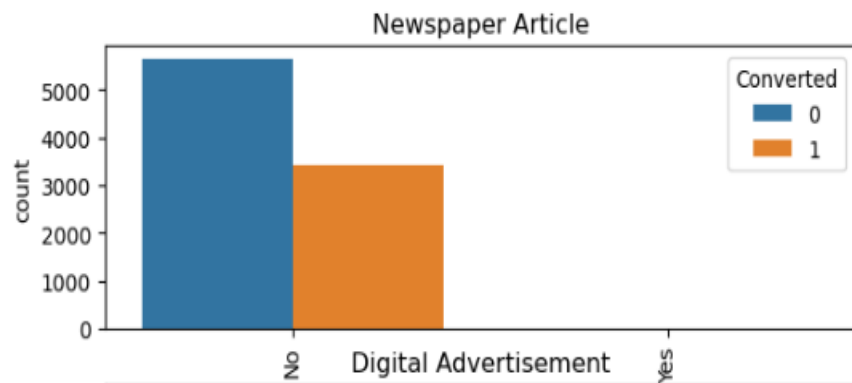


Categorical Variable Relation









Data Conversion

- Dummy variables are created for object type variables.

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 15 variables as output.