

LEADS SCORING CASE STUDY

SUBMITTED BY :

Shiva Chandra Kante, Krishnakumar V, Kamatchi M



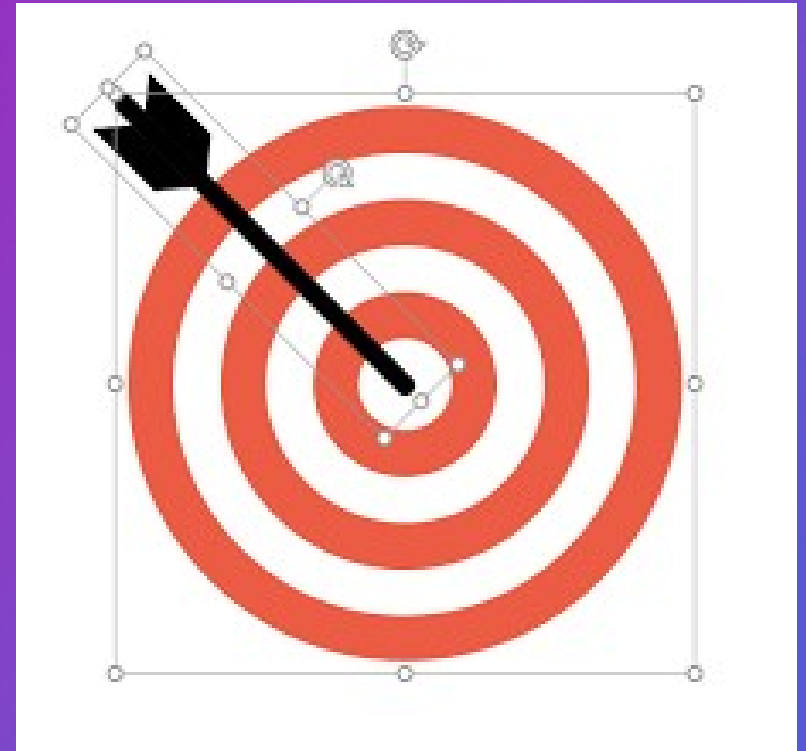


PROBLEM STATEMENT

- X Education sells online courses to industry professionals.
- Leads are acquired through website visits, from submissions, and referrals.
- Sales team follows up via calls and emails to convert leads.
- The typical lead conversion rate is around 30%.
- The company aims to identify high-potential leads (Hot Leads for efficiency.
- Focusing on Hot Leads can improve conversion rate and reduce effort on low-potential leads.

BUSINESS OBJECTIVE

- Develop a logistic Regression model to assign a Lead Score (0-100).
- Higher score – Hot Lead (High conversion probability).
- Lower score – Cold Lead (Low conversion probability).
- Ensure model flexibility to adapt to changing business requirements.
- Incorporate additional business problems as required.



PROBLEM APPROACH

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 a large number of missing values and are not useful for the analysis.
4. Imputation of the values, if necessary.
5. Check and handle outliers in data.

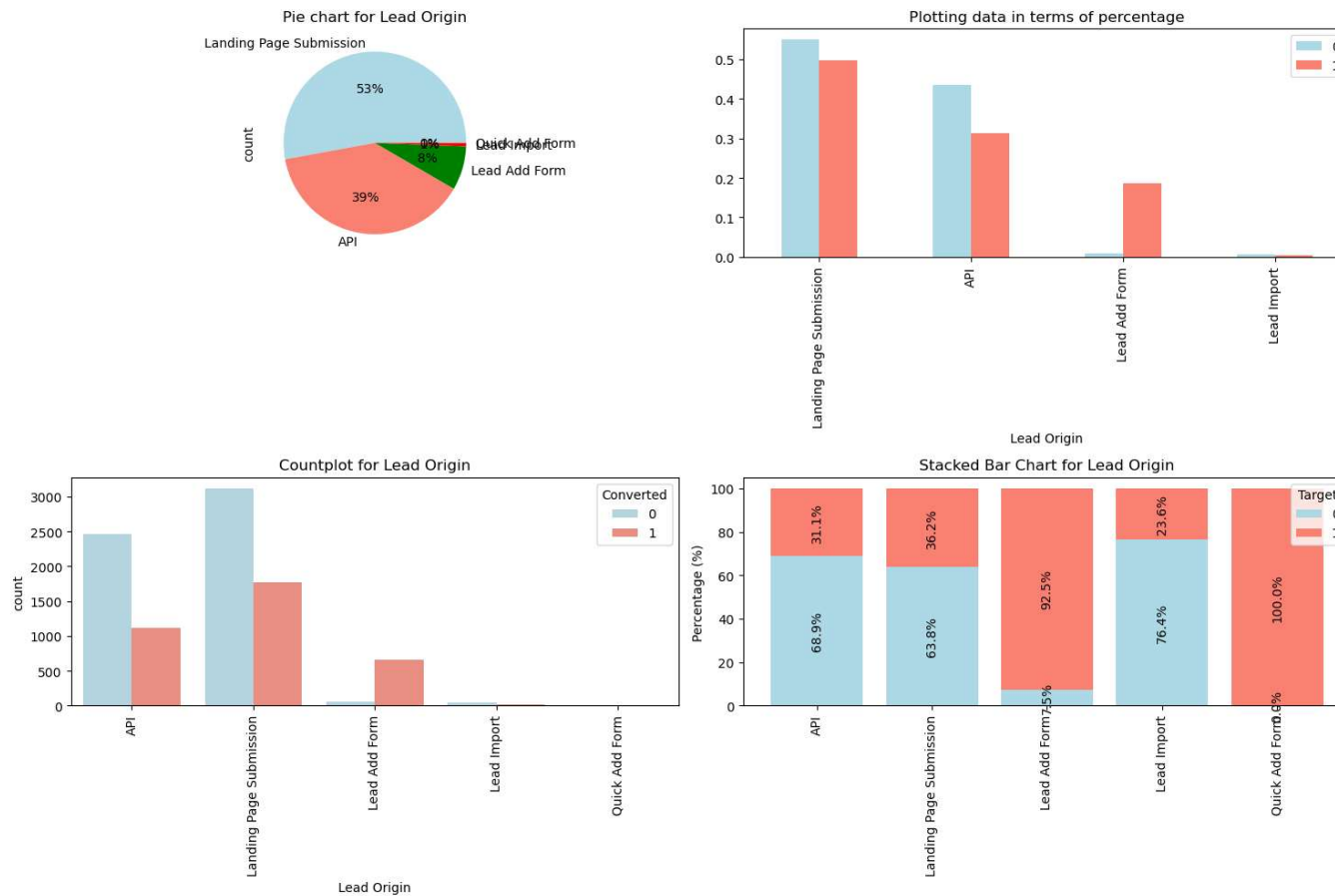
Exploratory Data Analysis (EDA)

1. Univariate data analysis: value count, distribution of variables, etc.
2. Bivariate data analysis: correlation coefficients and pattern between the variables etc.
3. Feature scaling & Dummy variables and encoding of the data.
4. Classification technique: logistic regression is used for model making and prediction.
5. Validation of the model.
6. Model presentation.
7. Conclusions and recommendations.

OBSERVATION FROM UNIVARIATE ANALYSIS

Lead Origin: "Landing Page Submission" identified 53% of customers, while "API" identified 39%.

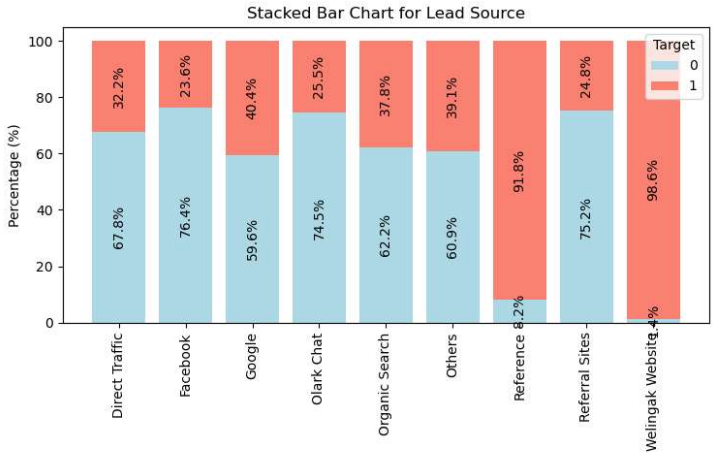
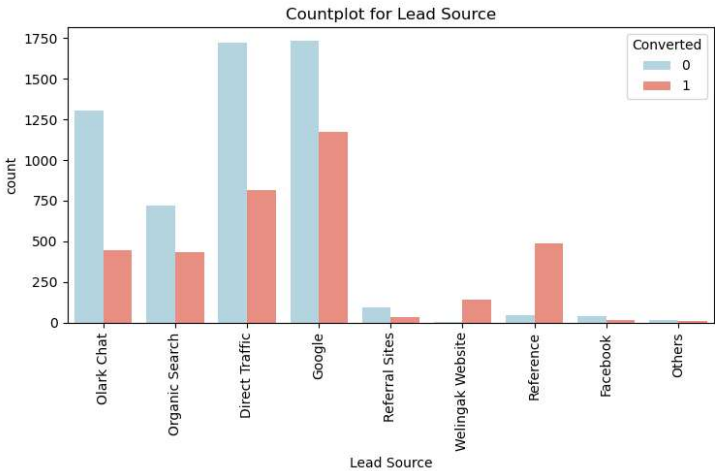
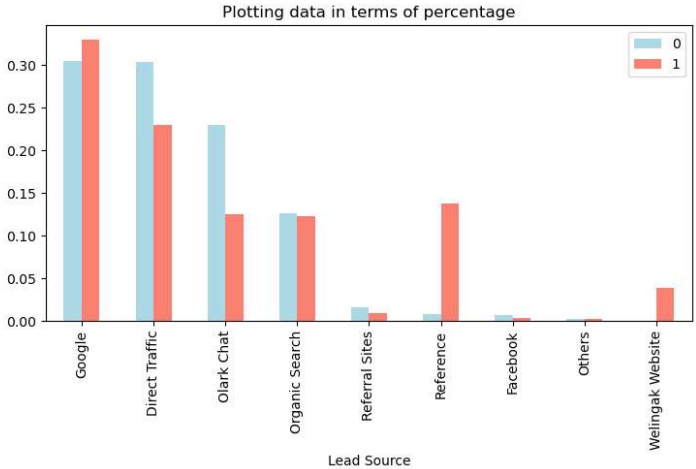
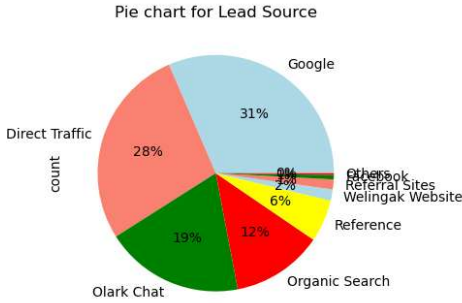
PLOT FOR LEAD ORIGIN



**EXPLORATORY
DATA ANALYSIS
(EDA)
UNIVARIATE
ANALYSIS**

Lead Source: 58% of leads come from Google & Direct Traffic combined.

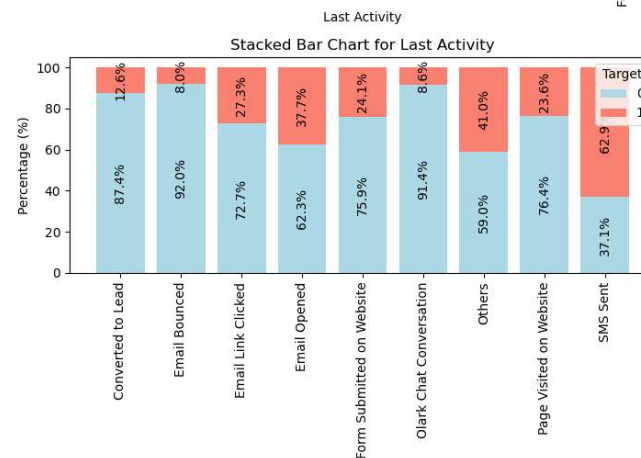
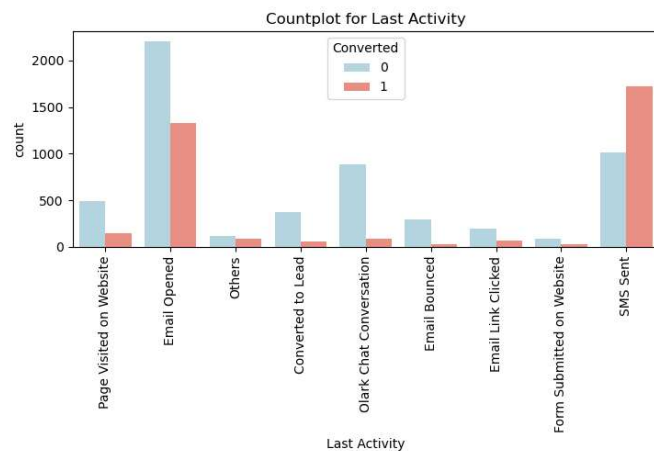
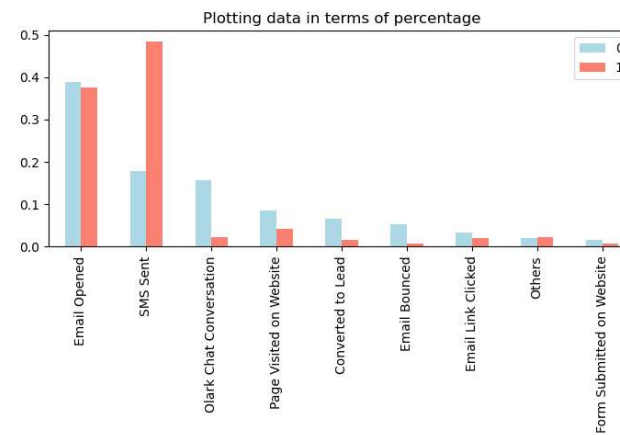
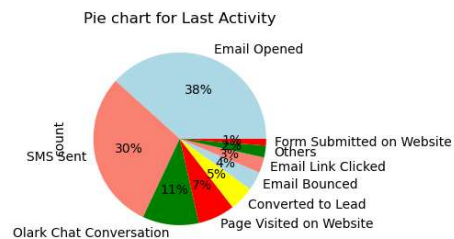
PLOT FOR LEAD SOURCE



EXPLORATORY
DATA ANALYSIS
(EDA)
UNIVARIATE
ANALYSIS

Last Activity: 68% of customer interactions are from SMS Sent & Email Opened activities.

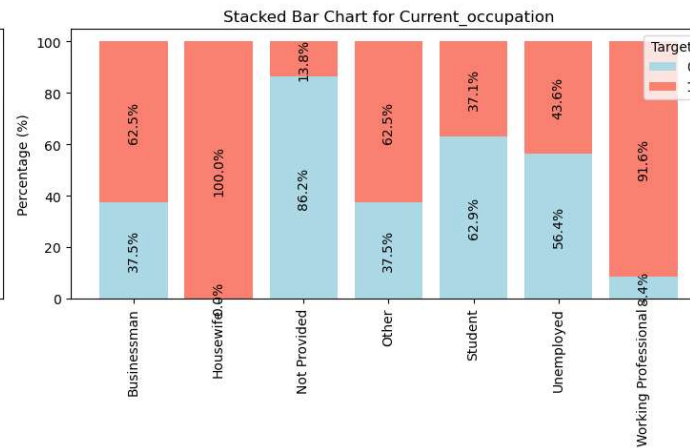
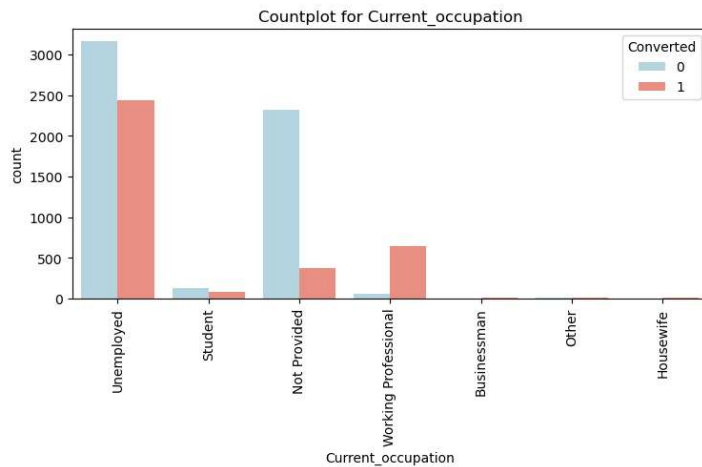
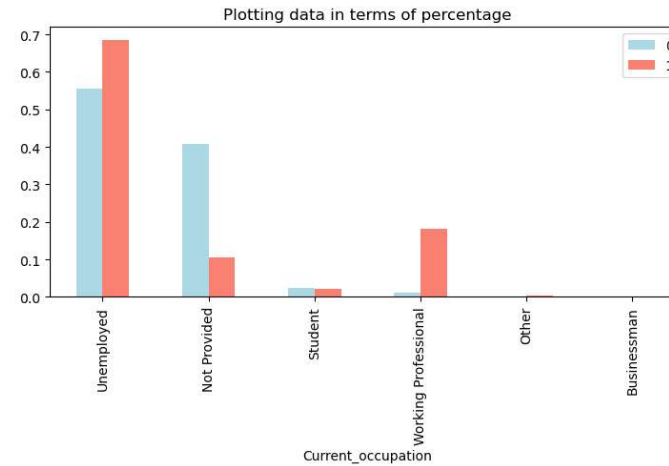
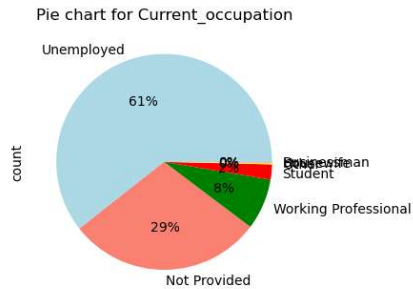
PLOT FOR LEAD ACTIVITY



**EXPLORATORY
DATA ANALYSIS
(EDA)
UNIVARIATE
ANALYSIS**

Current Occupation: 90% of customers are Unemployed.

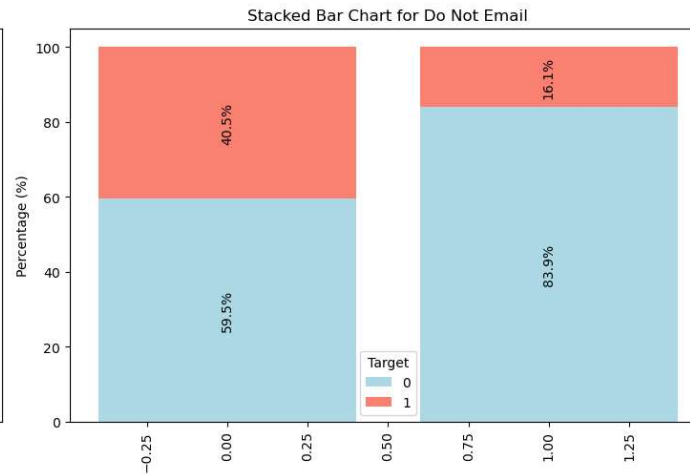
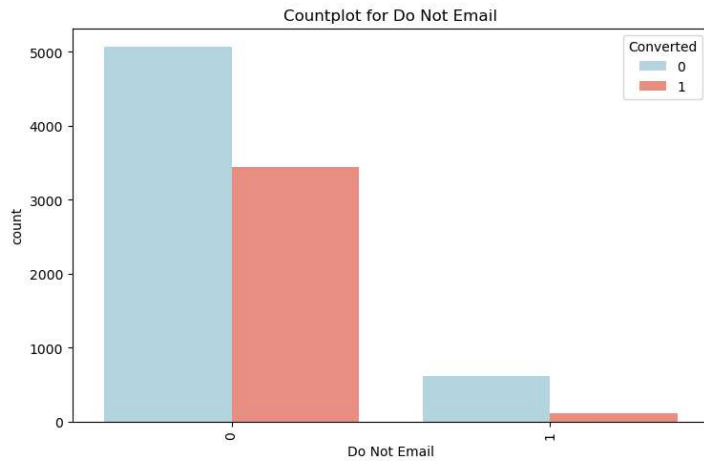
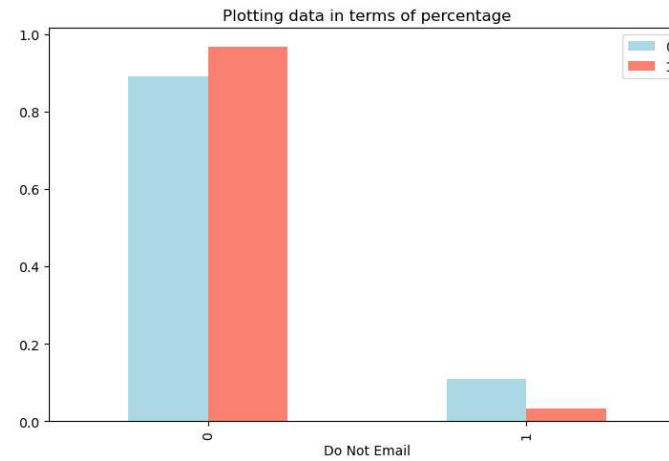
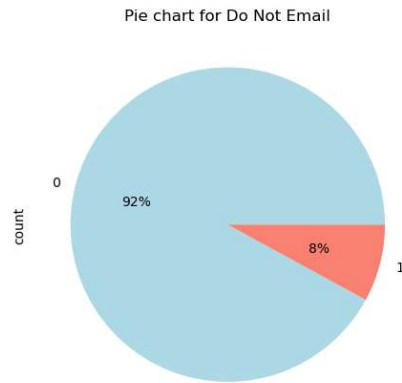
PLOT FOR CURRENT OCCUPATION



EXPLORATORY DATA ANALYSIS (EDA) UNIVARIATE ANALYSIS

Do Not Email: 92% of people opted not to receive emails about the course.

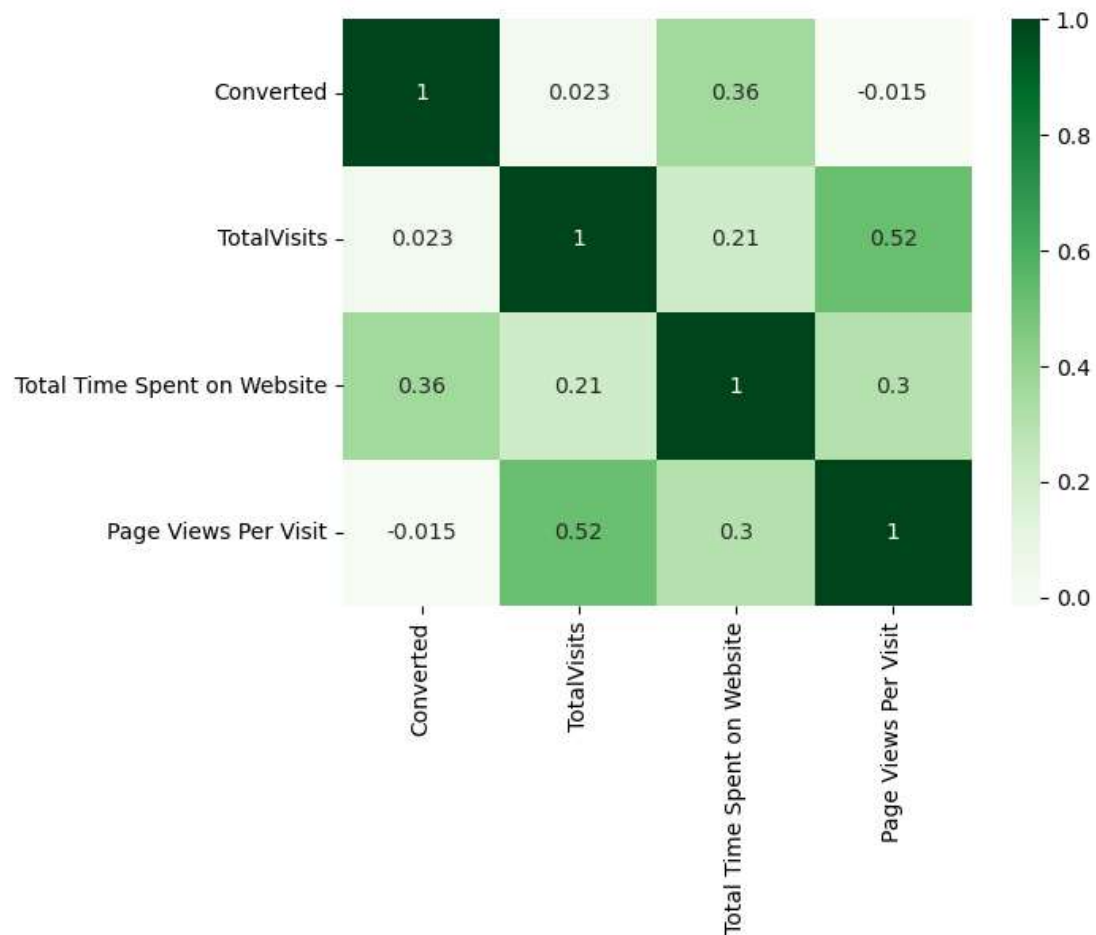
PLOT FOR DO NOT EMAIL



**EXPLORATORY
DATA ANALYSIS
(EDA)**

EXPLORATORY DATA ANALYSIS (EDA) BIVARIATE ANALYSIS

HEATMAP TO SHOW CORRELATION BETWEEN NUMERICAL VALUES



MODEL BUILDING

- Splitting the Data into Training and Testing Sets
- The first basic step for regression is train-test split, we have chosen 70:30 ratio.
- Use RFE for Feature selection
- Running RFE with 15 variables as output
- Building Model by removing the variable whose p-value is greater than 0.05 and
vi value is greater than 5
- Predictions on test data set
- Overall accuracy 82%