

Lead Scoring Case Study

Rohan Sasane

Introduction

- X Education needs help in selecting the most promising leads, i.e. the leads that are most likely to convert into paying customers. The company requires you to build a model wherein you 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, in particular, has given a ballpark of the target lead conversion rate to be around 80%

Steps.

- 1. EDA
- 2. Data pre processing
- 3. Train-Test Data split
- 4. Model Building

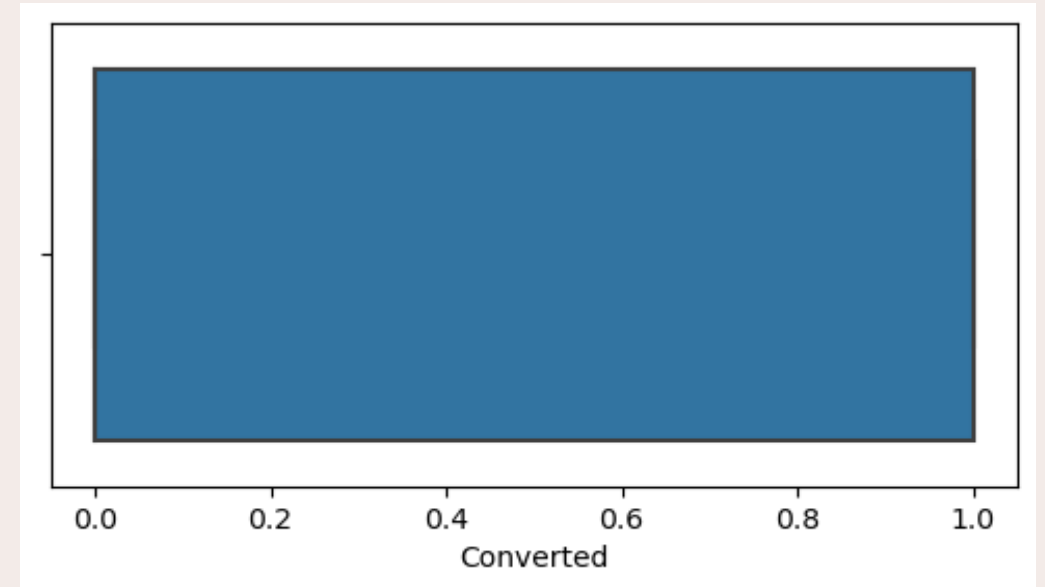
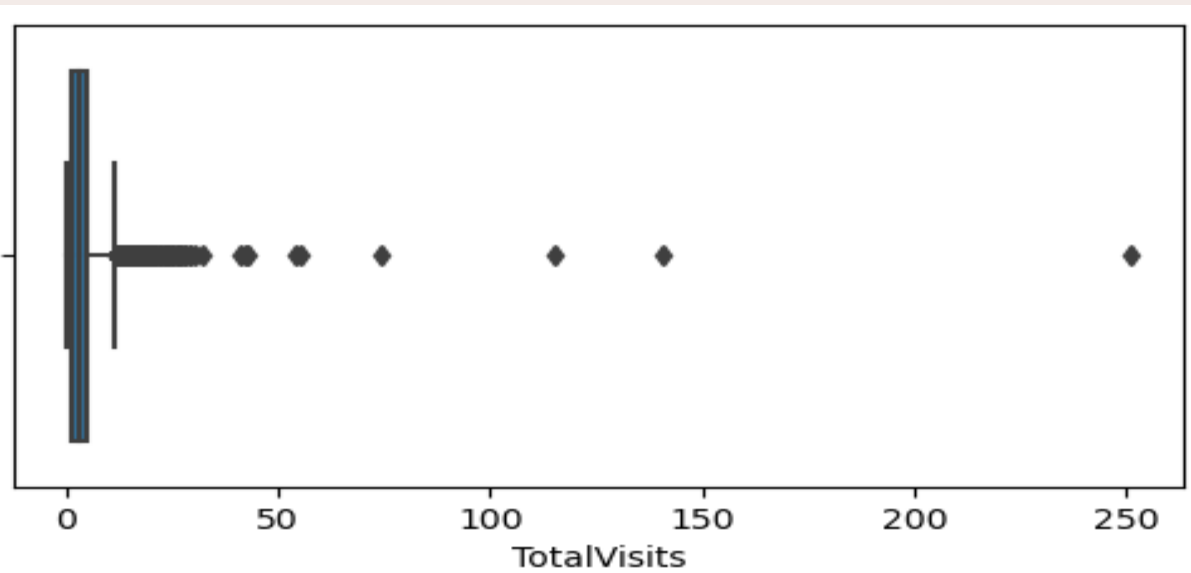
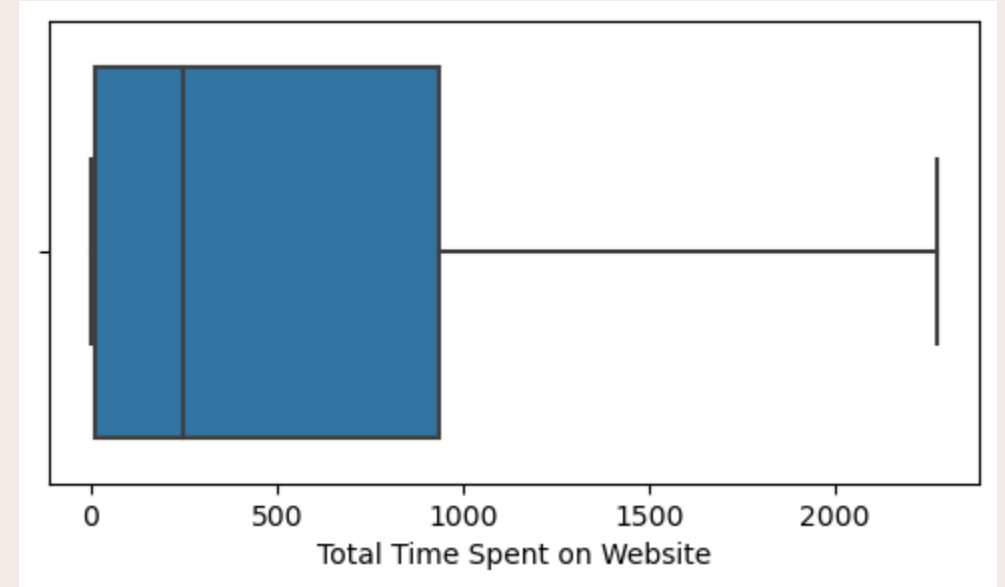
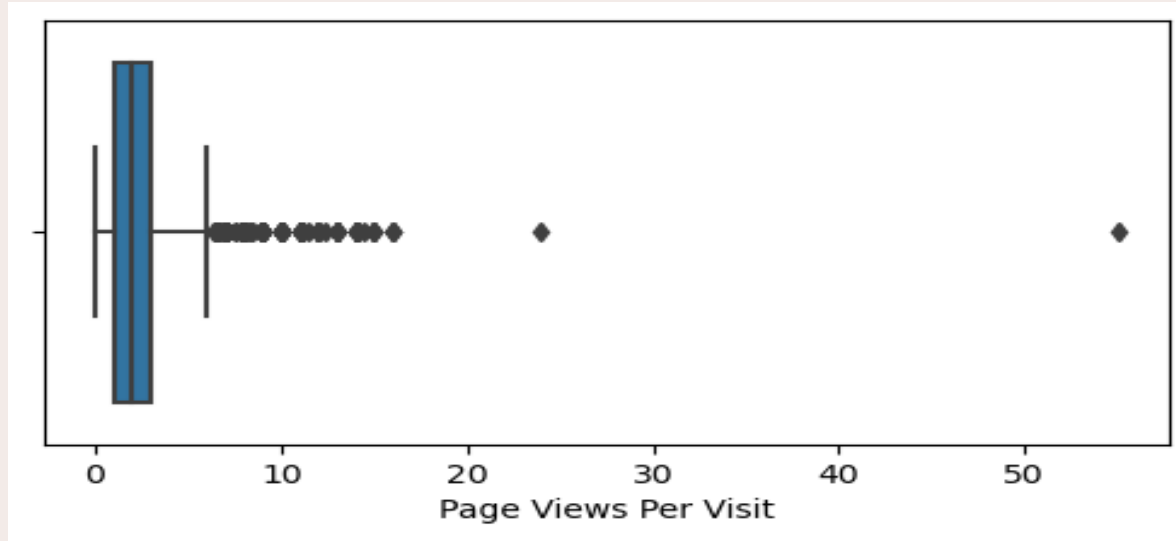
EDA

In the data set 9240 rows and 37 columns, data have missing values which can be handled by dropping the columns which have more than 40% missing data.

So we drop the column which have missing values more than 40%

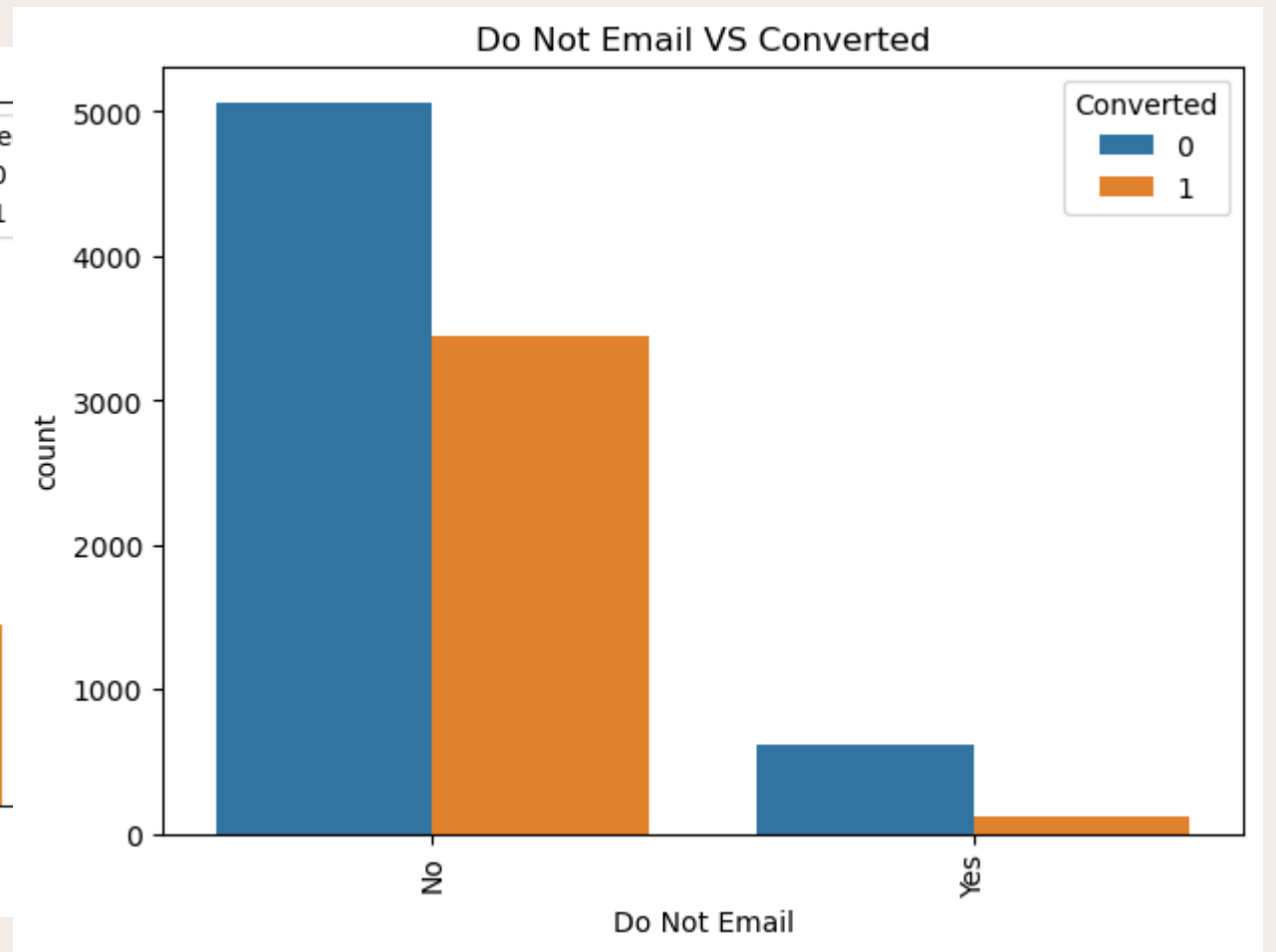
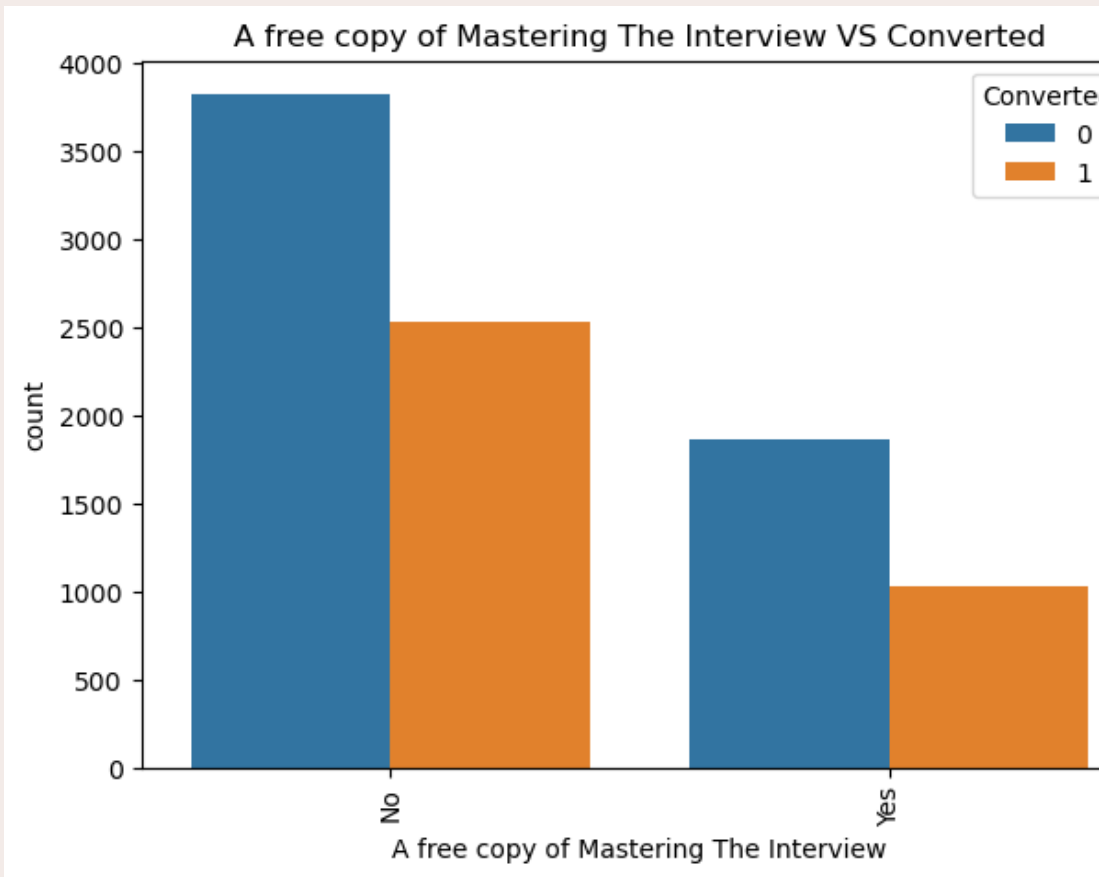
Finally we have 9240 and 16 rows and columns

Outlier Check

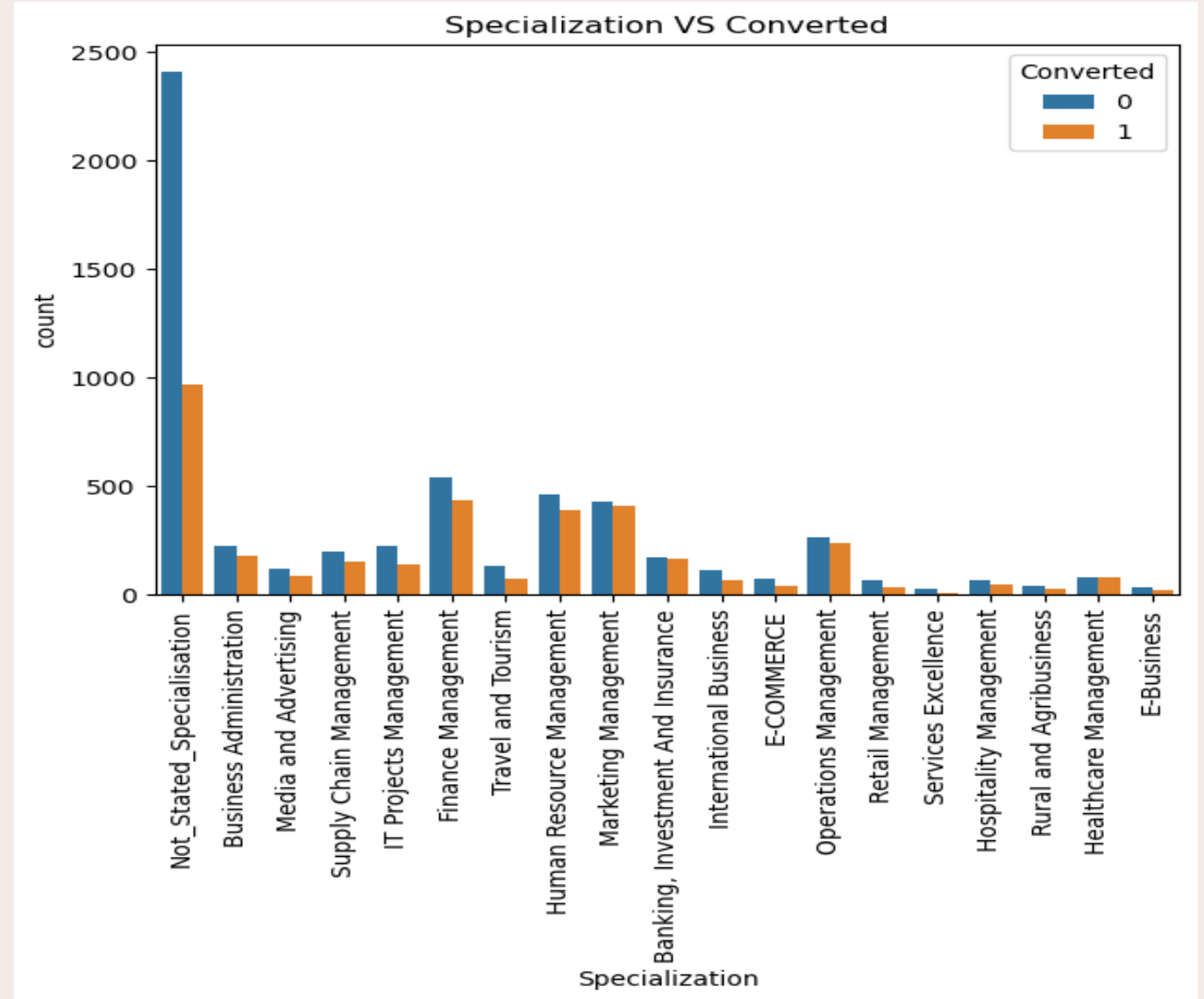
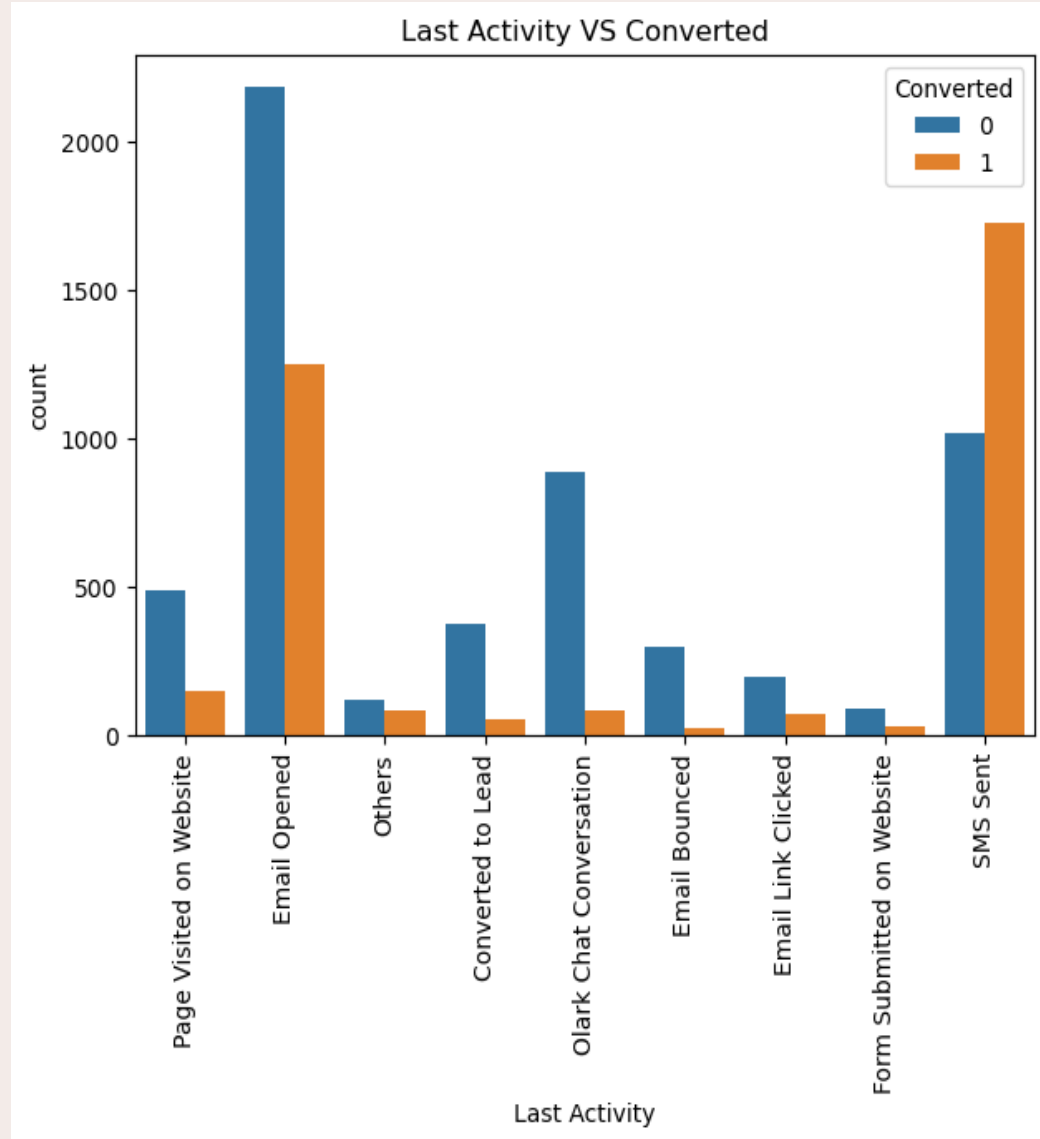


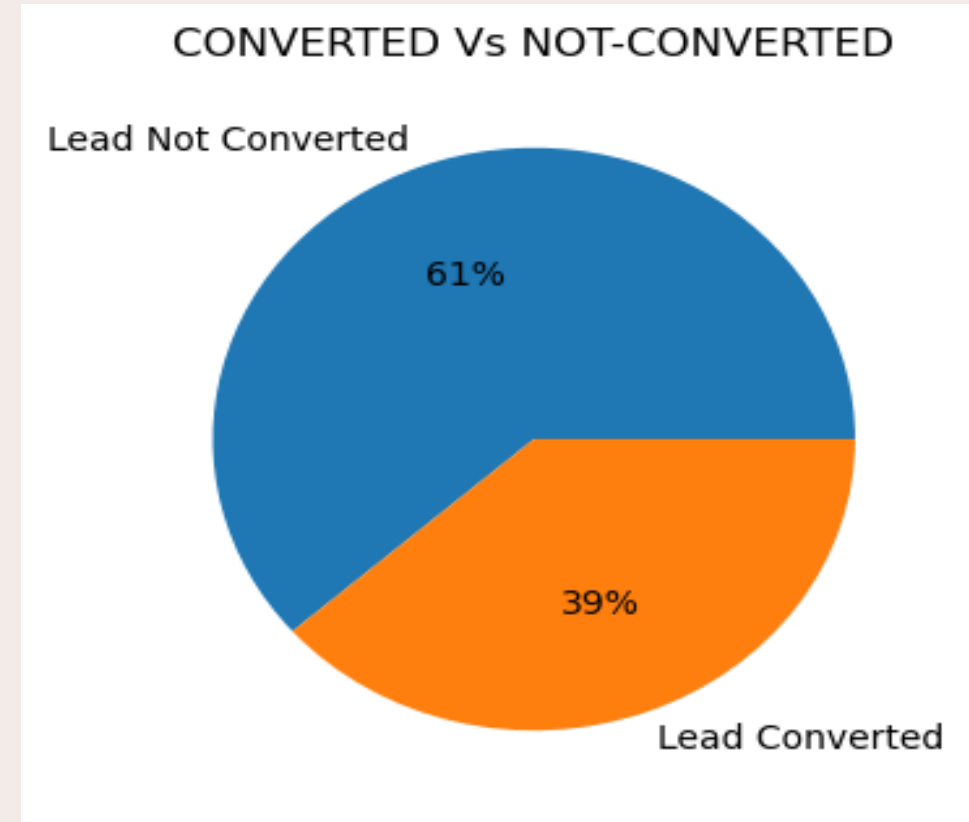
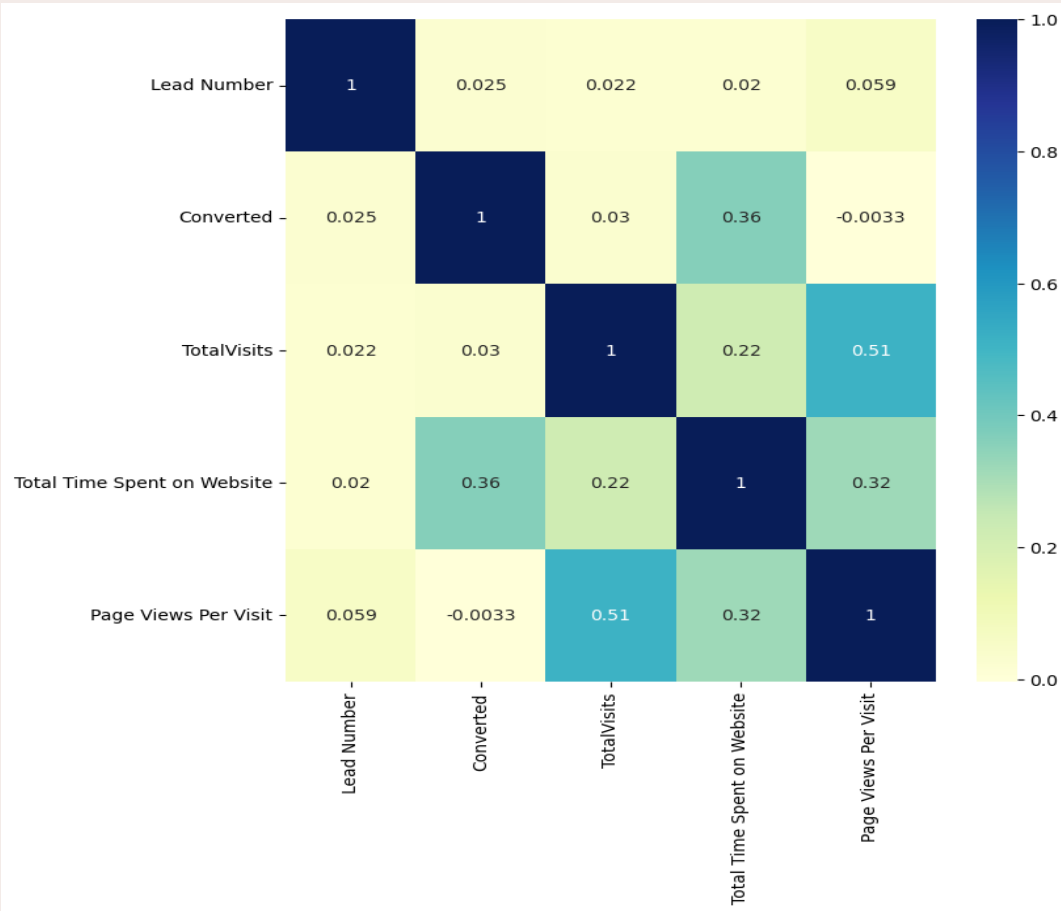
EDA

- did univariate analysis and then outlier treatment these were some potential outliers which I did capping.
- did bivariate analysis
- Lead origins categories “Lead Add Form” category have highest conversion ratio
- “SMS Sent” category in “Last Activity” column has highest conversion ratio
- Following that “Email Opened” has the second highest
- In “Lead Source” category “Reference” is performing better in conversion following with “Google”
- Should target “Working Professional” more than “Unemployed”.



EDA





“

Top 3 factors which can help generate more leads:

1. Lead Source_Wellingak website

2. Lead Origin_Lead Add form

3. Tags_Will revert after reading email.

”

Thank you

Rohan Sasane.