Lead scoring case study

GROUP MEMBERS:
SACHIN
AKSHIT SHAH

PROBLEM STATEMENT:

- An education company named X Education sells online courses to industry professionals. Interested ones lands on website and fill the form. By that they become a lead but only 30% of the leads are converting to actual customers.
- We need to filter out these leads to hot leads using the data.

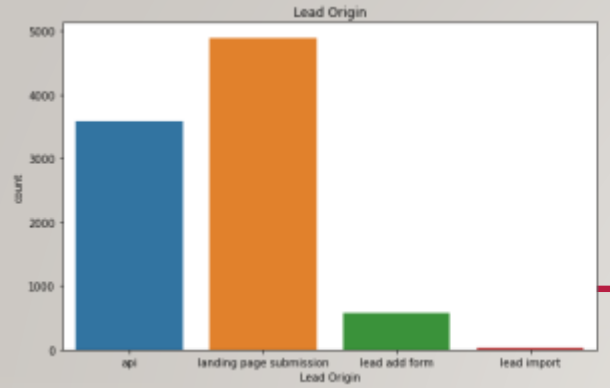
ASSUMPTION:

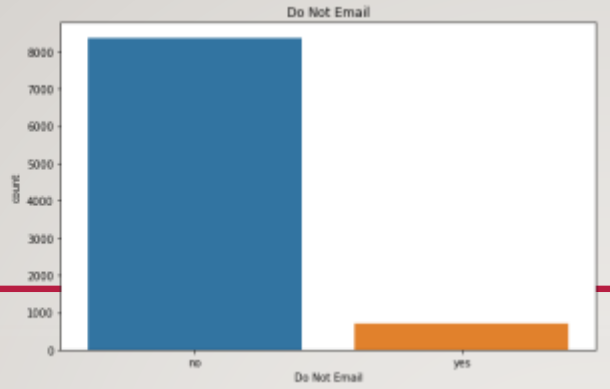
- There are 4 columns that have large number of null values but it looks like they are crucial columns to the data. So we are assuming that the null values are the information which is not provided.
- So we have converted the null values to string "**not provided**"

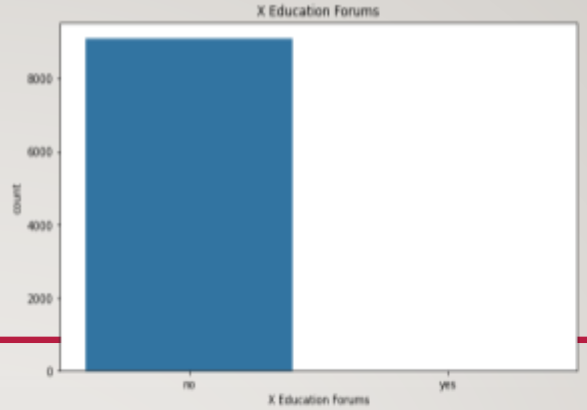
EDA

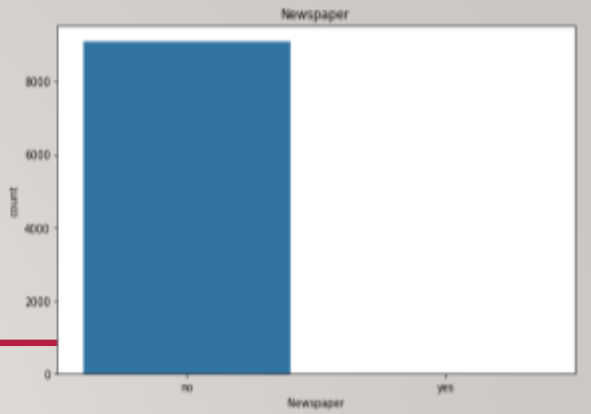
- Dropped column with more than 35% of missing value.
- Converted country column's data into 3 slot and that are India, Out side of India and not provided.

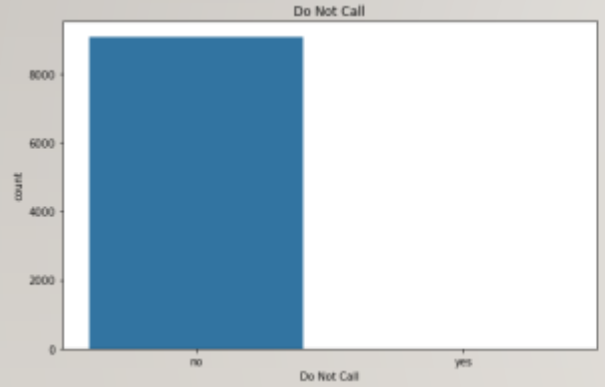
EDA: UNIVARIATE ANALYSIS

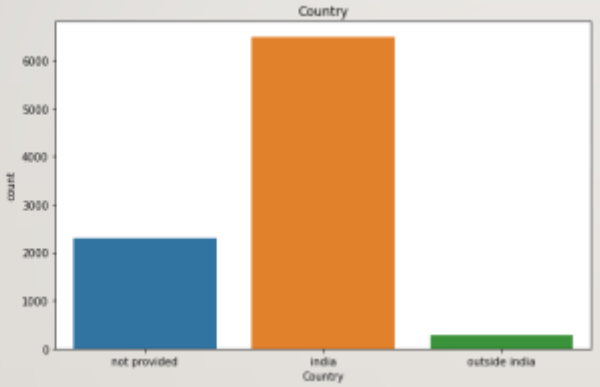


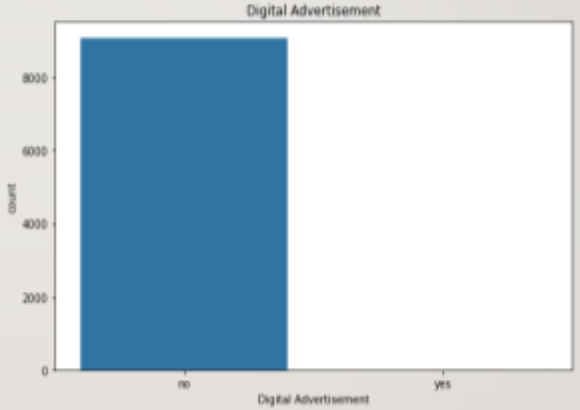


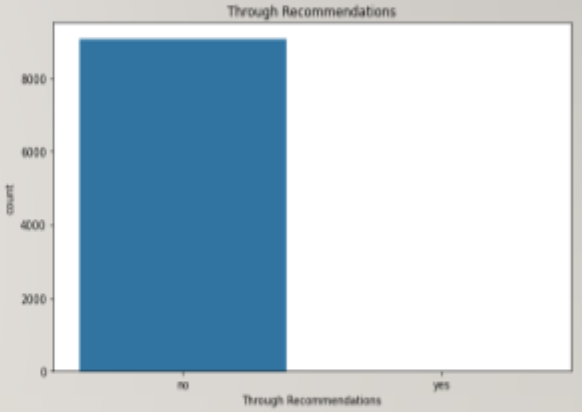


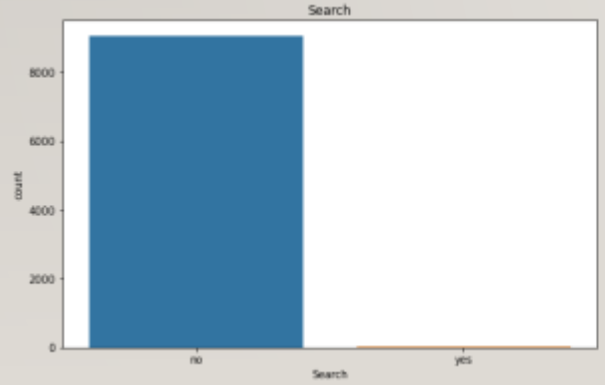


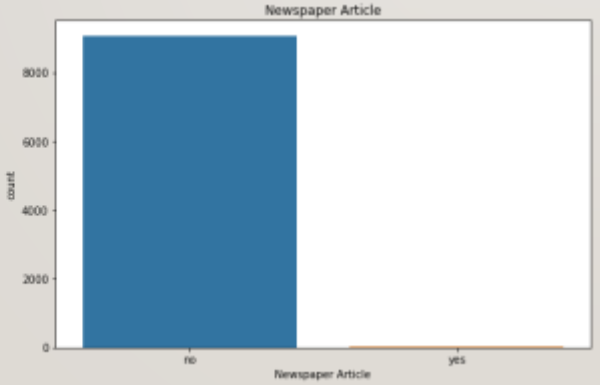


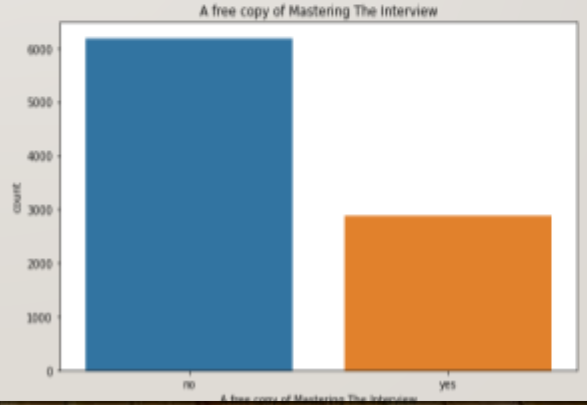


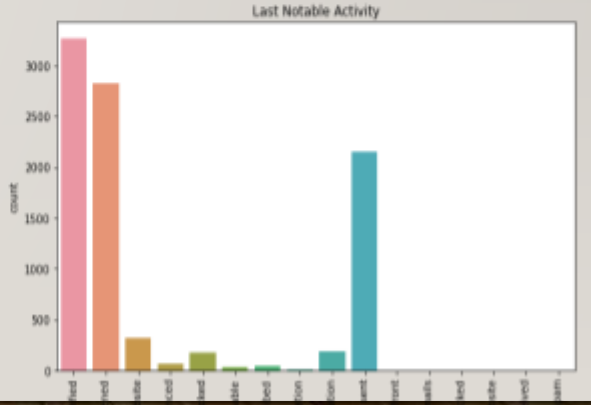












EDA: CORELATION BETWEEN VARIABLES

It is clear from the above EDA that many elements contain relatively little data and will therefore have a lower importance for our study.

APPROACH

- Created dummy variables for required columns.
- Split the data into 70% train and 30% train set.
- Fit transfer data using MinMaxScaler for certain columns.
- Use RFE to chose which column should be use for training.
- Dropped more column using P-value and VIF readings.
- Created prediction and evaluate model using various methods.
- Make prediction on test set and find precision and recall value.

CONCLUSION AND RECOMMENDATION

- According to research, the following factors affected potential purchasers the most (in descending order):

1. The total time spend on the Website.
2. Total number of visits.
3. When the lead source was: a. Google b. Direct traffic c. Organic search d. Welingak website
4. When the last activity was: a. SMS b. Olark chat conversation
5. When the lead origin is Lead add format.
6. When their current occupation is as a working professional. With these in mind, X Education can succeed since they have a very good probability of persuading nearly all prospective customers to change their minds and purchase their courses.