

LEAD SCORE CASE STUDY

Presented by:

Vundecode Adithya Krishna

Soubhikdas

Ashharulhaque

PROBLEM

- X – education, an Ed-tech company, sells online courses to working professional.
- But the lead conversion rate is very low, about 30%
- The company wants to increase the lead conversion rate to around 80%, and wants to identify the potential leads.

OBJECTIVE

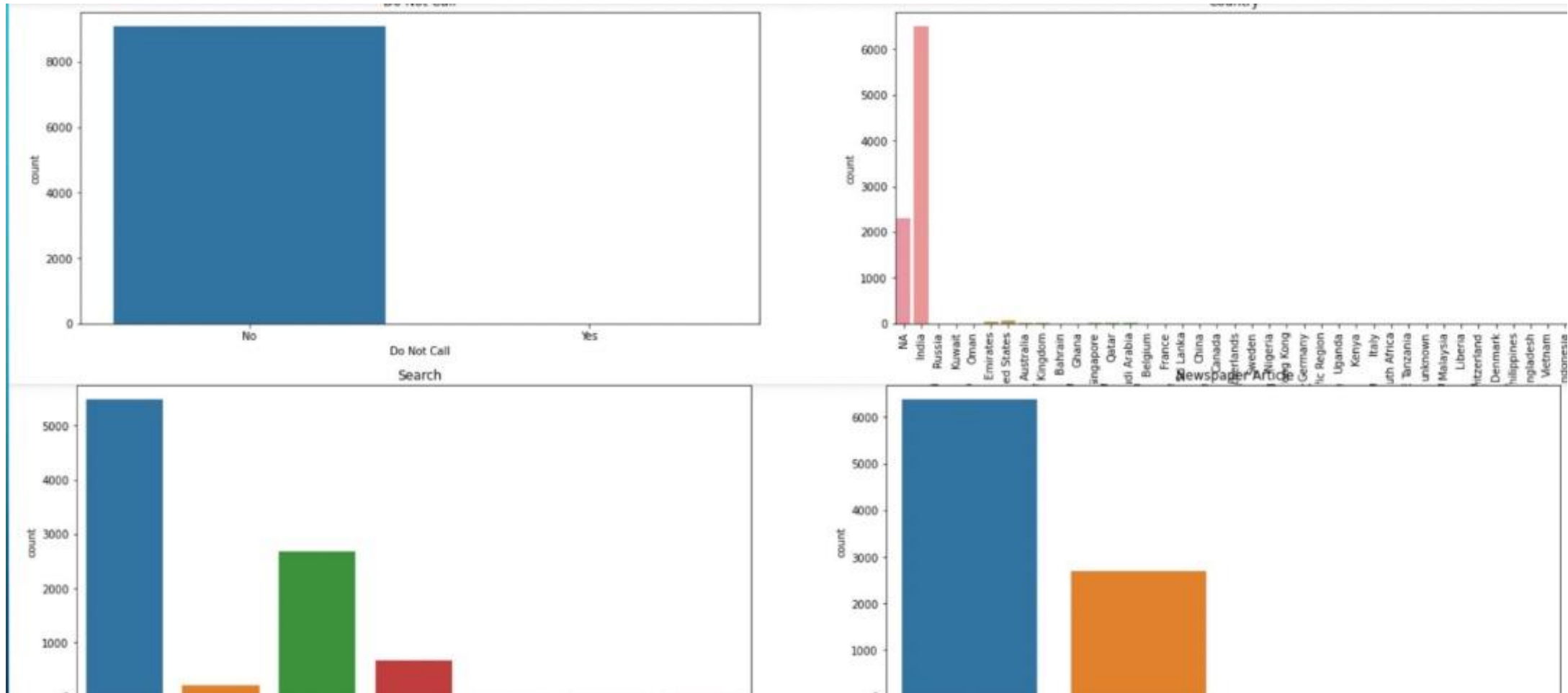
- X-education wants to identify the hot leads.
- Build and delay a model that point out the necessary variables.

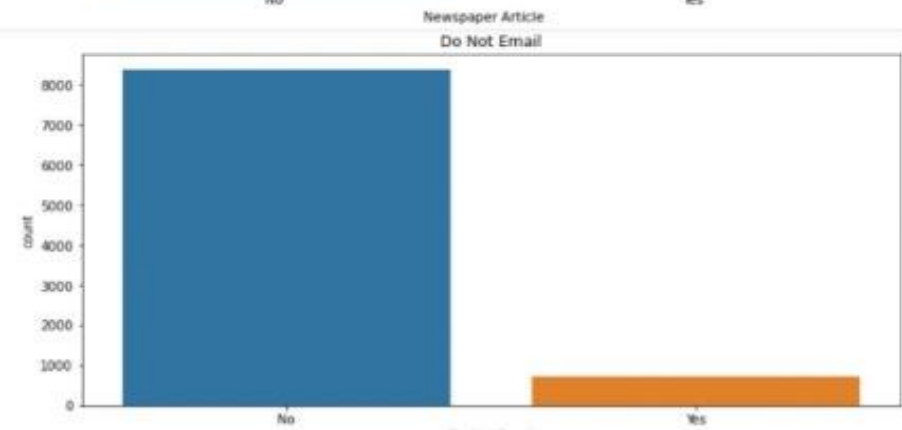
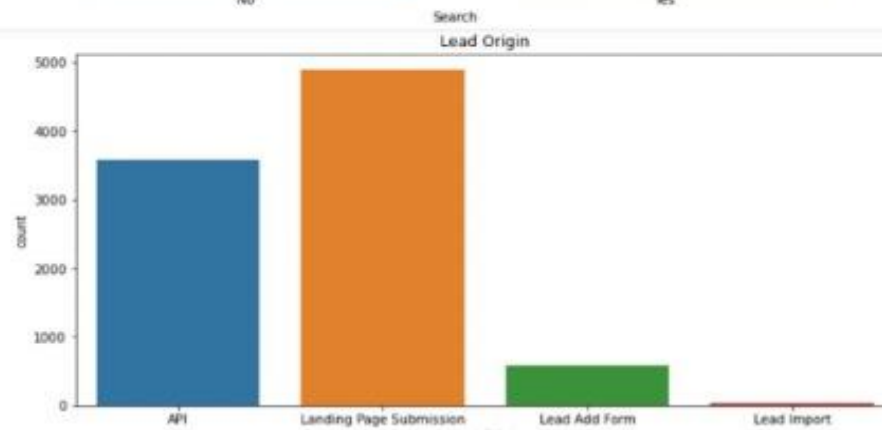
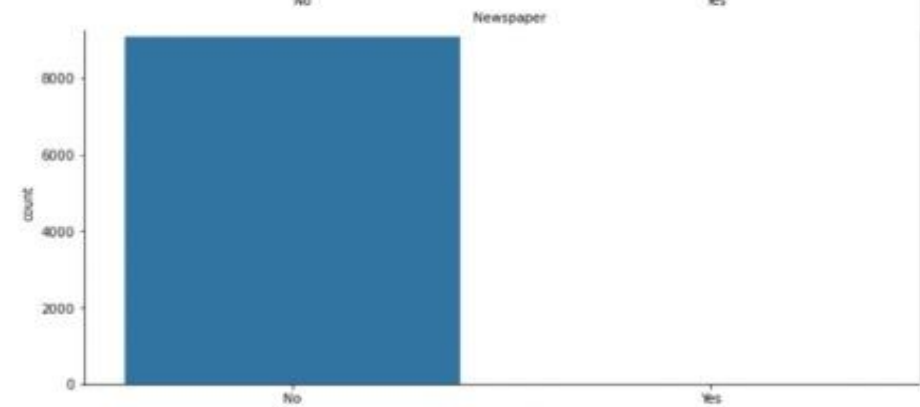
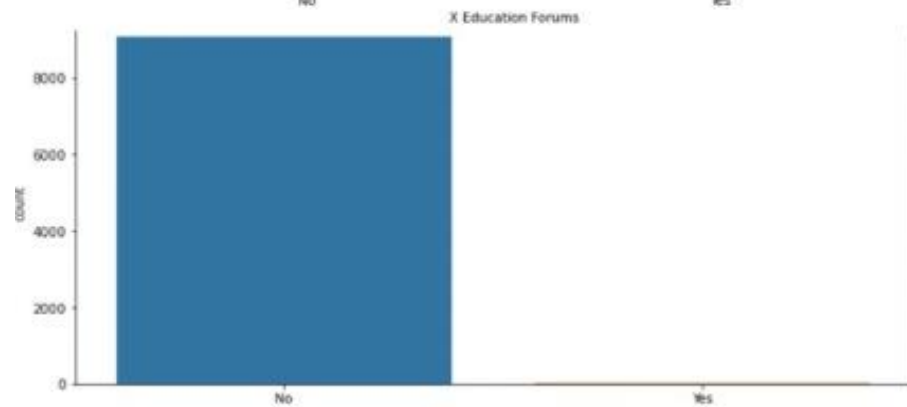
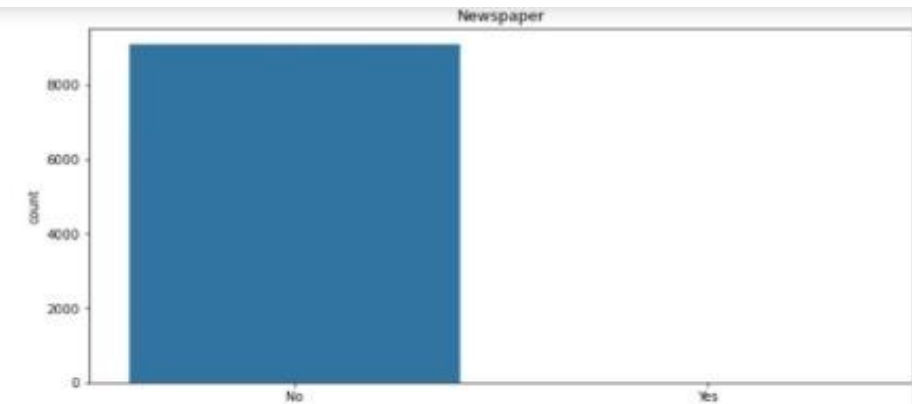
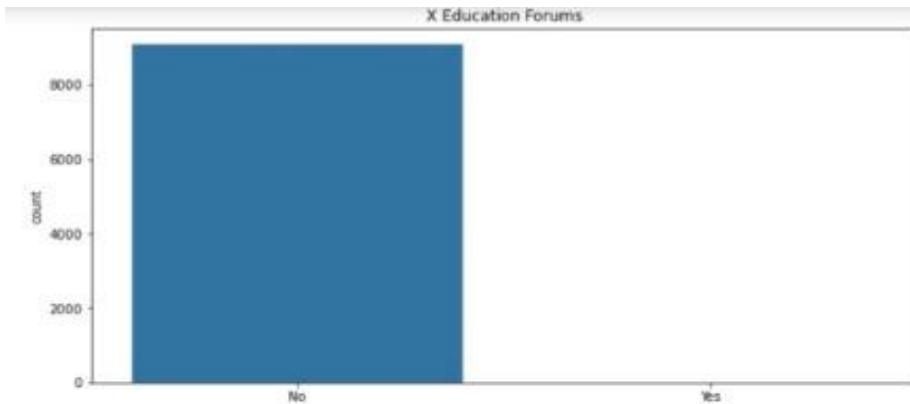
APPROACH

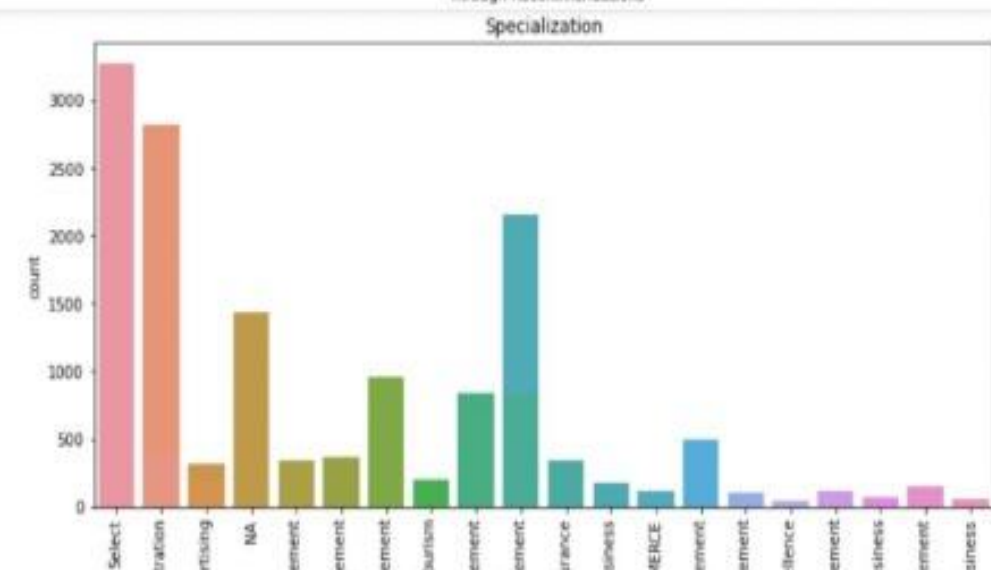
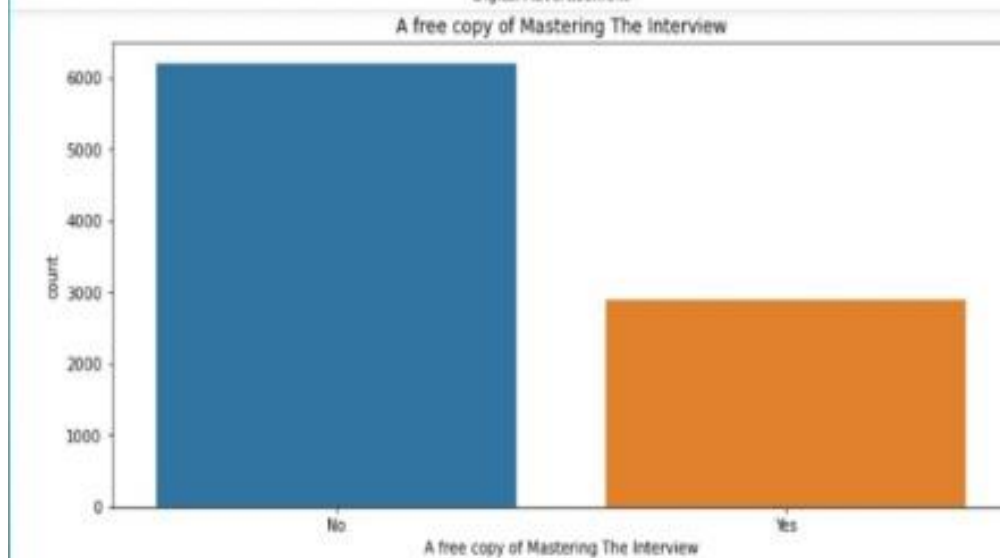
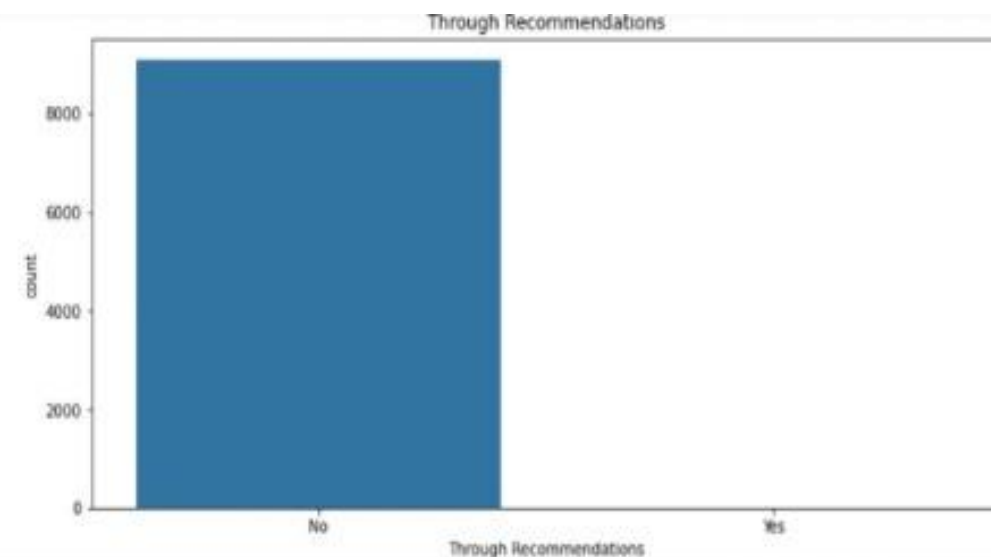
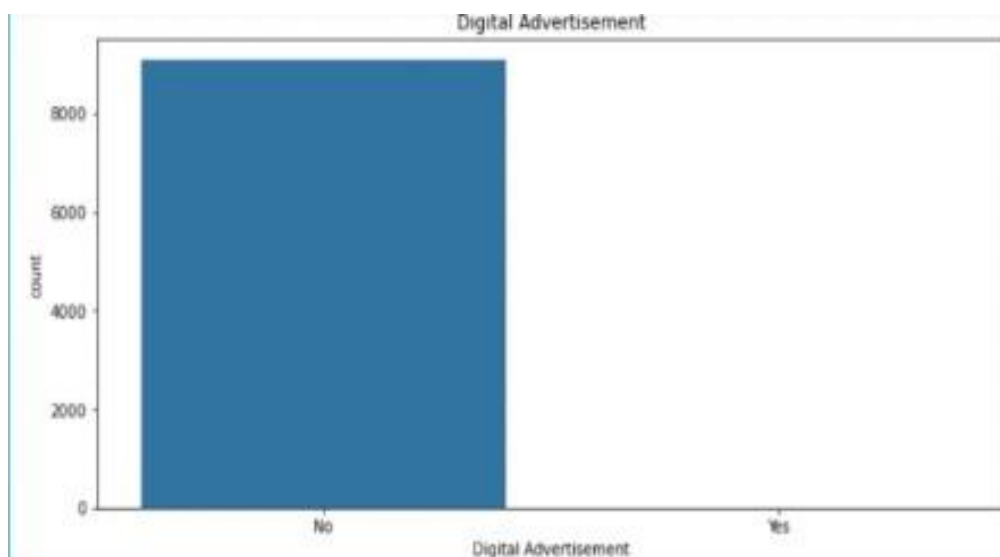
- The first step is to clean the data, to set the context, to identify missing values and make the data fit enough for EDA.
- Next comes the Exploratory Data Analysis- doing analysis to find patterns in the data and some statistical values, correlation etc.
- Perform Feature Scaling and Dummy Variables for building variables for the model.
- Splitting the dataset, fitting a linear regression model and check accuracy.
- Use the testing set to see the accuracy of the model.

EDA

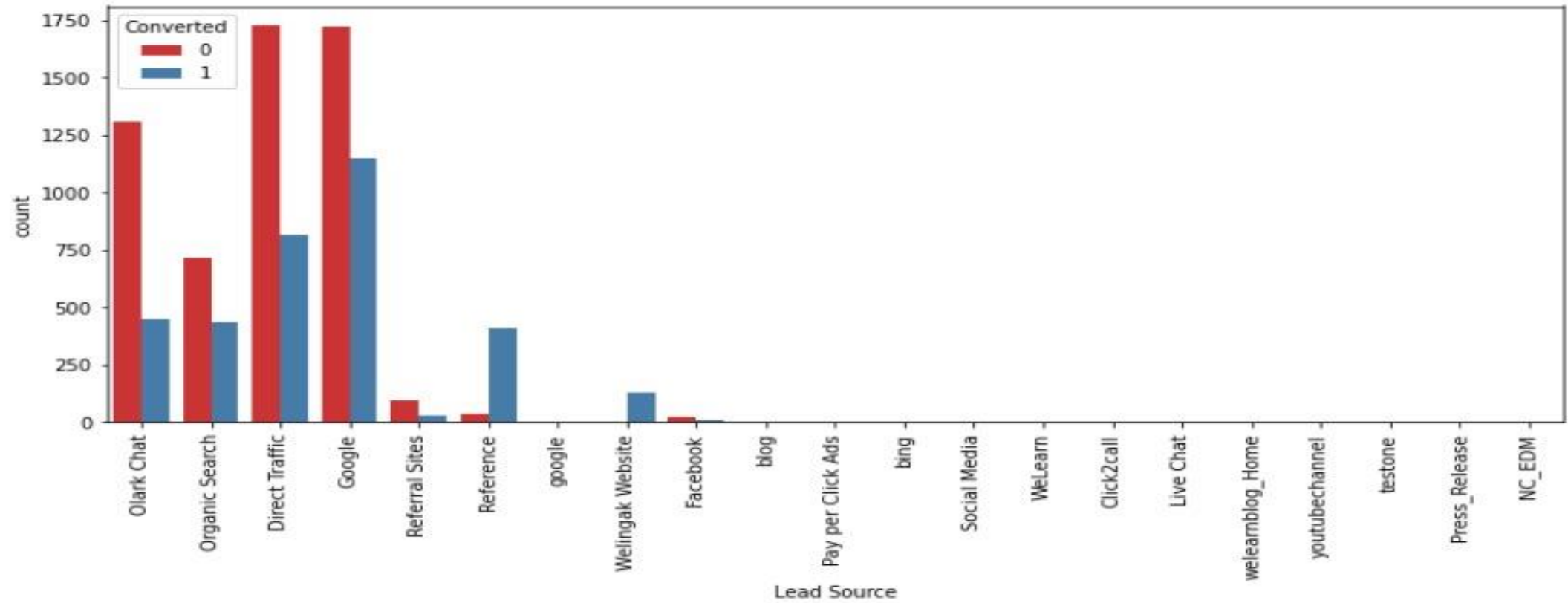
Categorical Variables



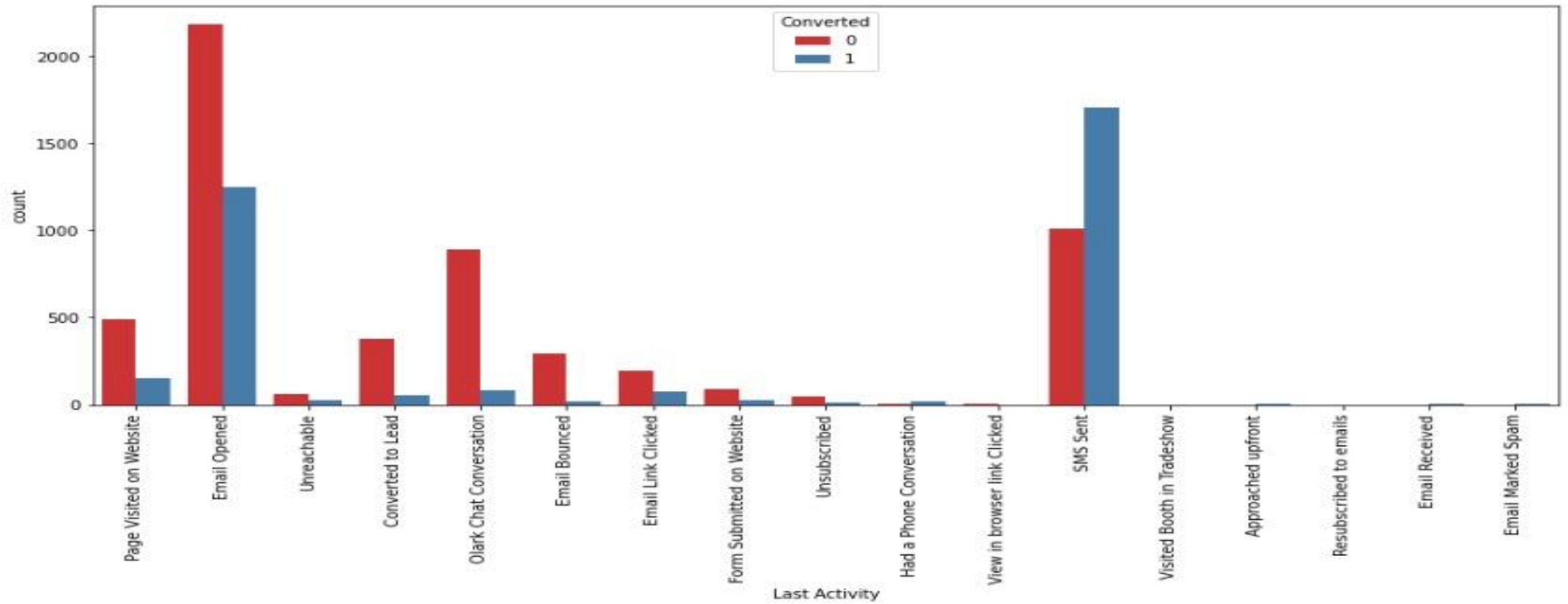




Lead scores



Last activity



DATA CONVERSION

- The numerical variables are normalized
- Only the variables with object as datatype were converted into dummy variables.
- Dimension of dummy variable table: (9074x137)

BUILDING THE MODEL

- Split the data 75% (training set) and 25% (testing set)
- Feature selection by Recursive Feature Elimination (RFE)
- Used 15 variables as initial RFE output.
- Build the model keeping in mind the $p\text{-value} < 0.05$ and variance inflation factor < 5
- Repeat the process of model building till we get all the proper p-values and vifs. Here, we had 2-iterations, the later one being the final model.
- Predict using the testing set.
- Prediction accuracy achieved $\sim 80\%$

CONCLUSION

The variables that directly impact the lead conversion (Ordered in descending order) :

>The total time spent on website.

>Total number of visits.

- When the lead source was: Direct traffic, Welingak website, Organic search.
- When the last activity was: Olark chat conversation.
- When the lead origin is Lead add format.

Prioritizing these factors would lead to a high lead conversion rate.