

LEAD SCORING CASE STUDY



CASE STUDY DONE BY :

Shrinaath V

Shrikant Chavan

Shubham Srivastava

PROBLEM STATEMENT

2

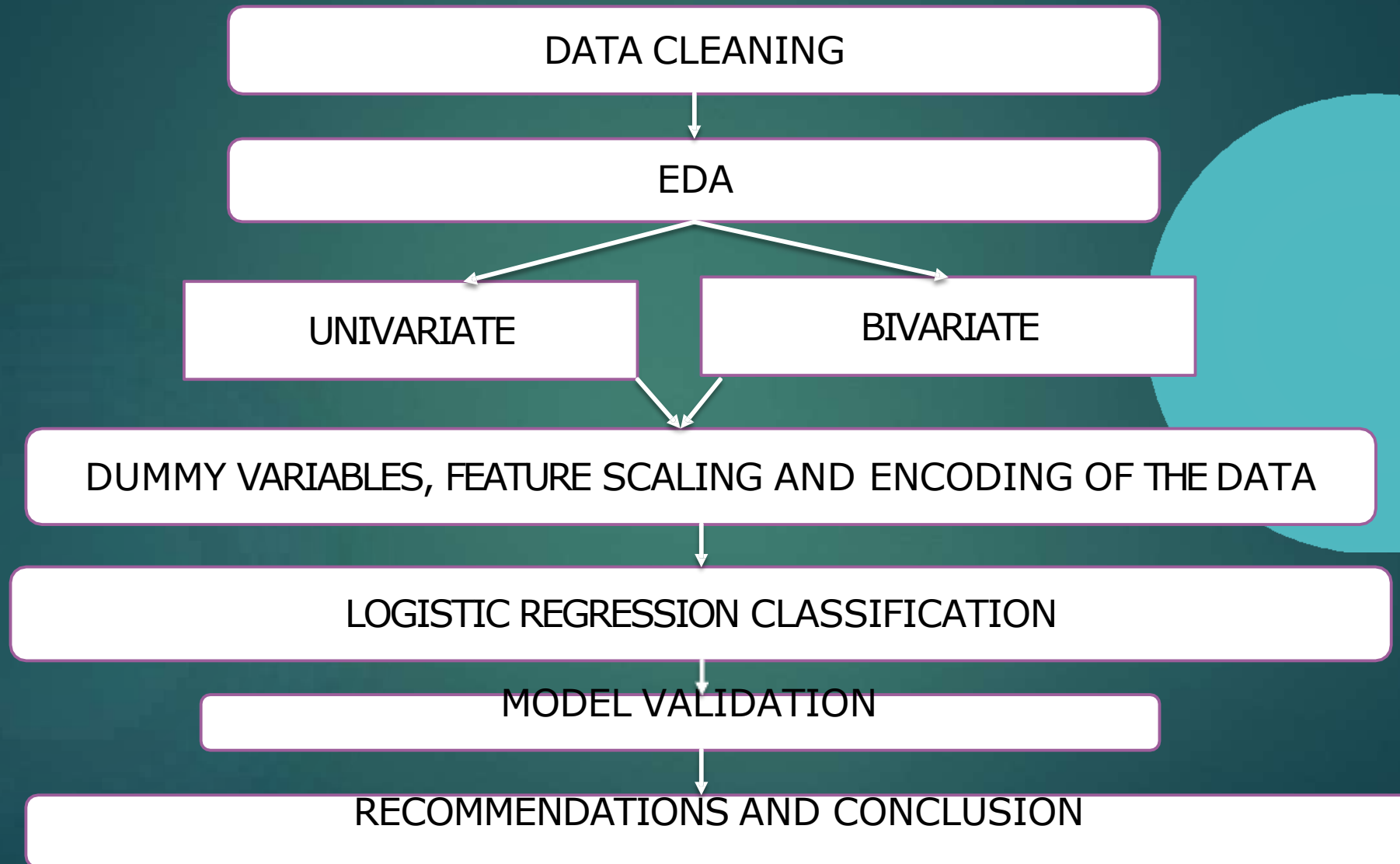
- ▶ An education company named X Education sells online courses to industry professionals.
- ▶ X Education gets a lot of leads, its lead conversion rate is very poor. For example, if, say, they acquire 100 leads in a day, only about 30 of them are converted.
- ▶ To make this process more efficient, the 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.

OBJECTIVE

- ▶ The education company (X) wants to know the most promising leads.
- ▶ They need leads that are most likely to convert into paying customers
- ▶ To make the above objective to into effect they want us to build a model which identifies the hot leads.
- ▶ The CEO has given a ballpark of the target lead conversion rate to be around 80%.
- ▶ The model should be used in deployment for future use.

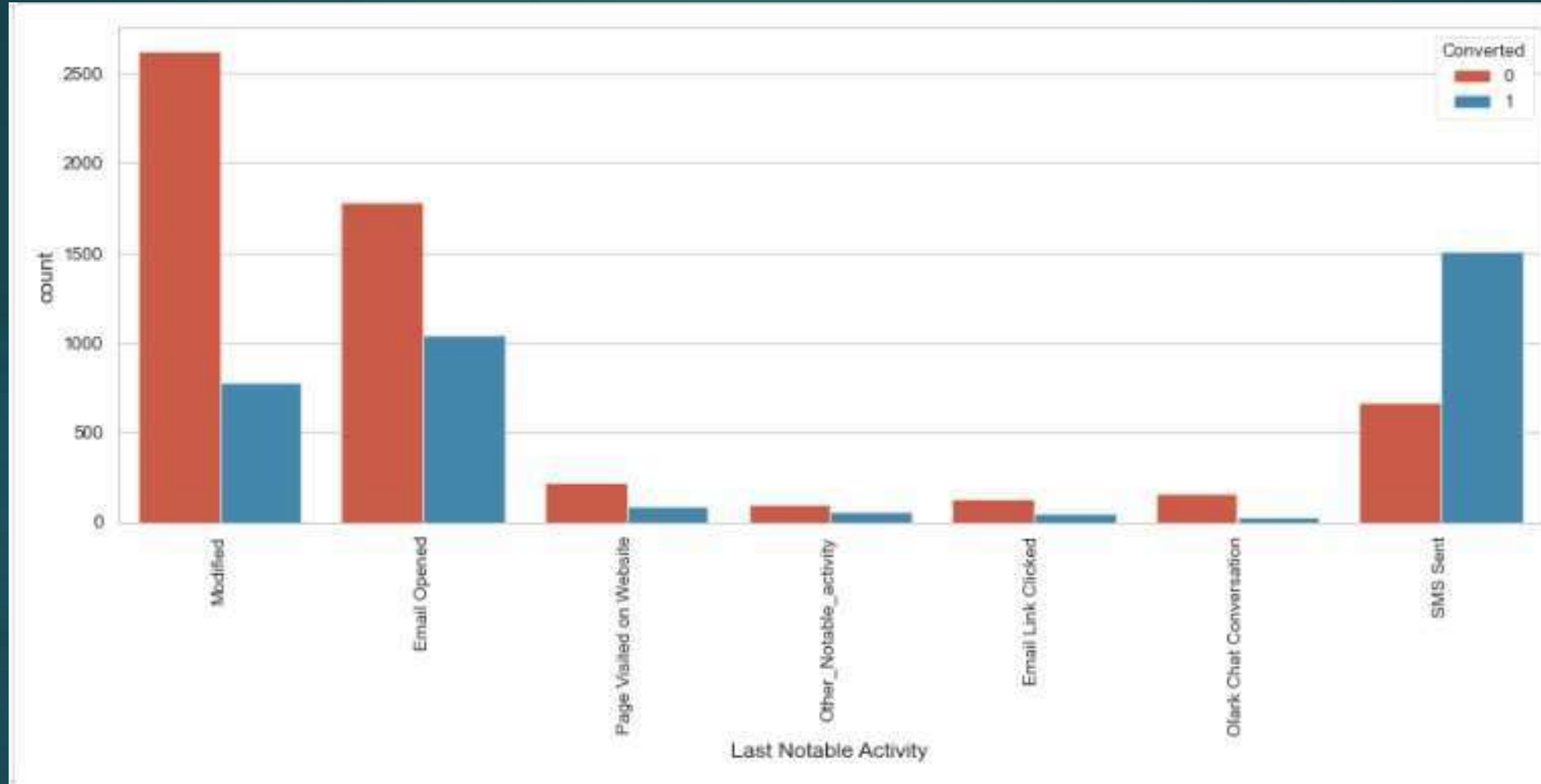
METHODOLOGY

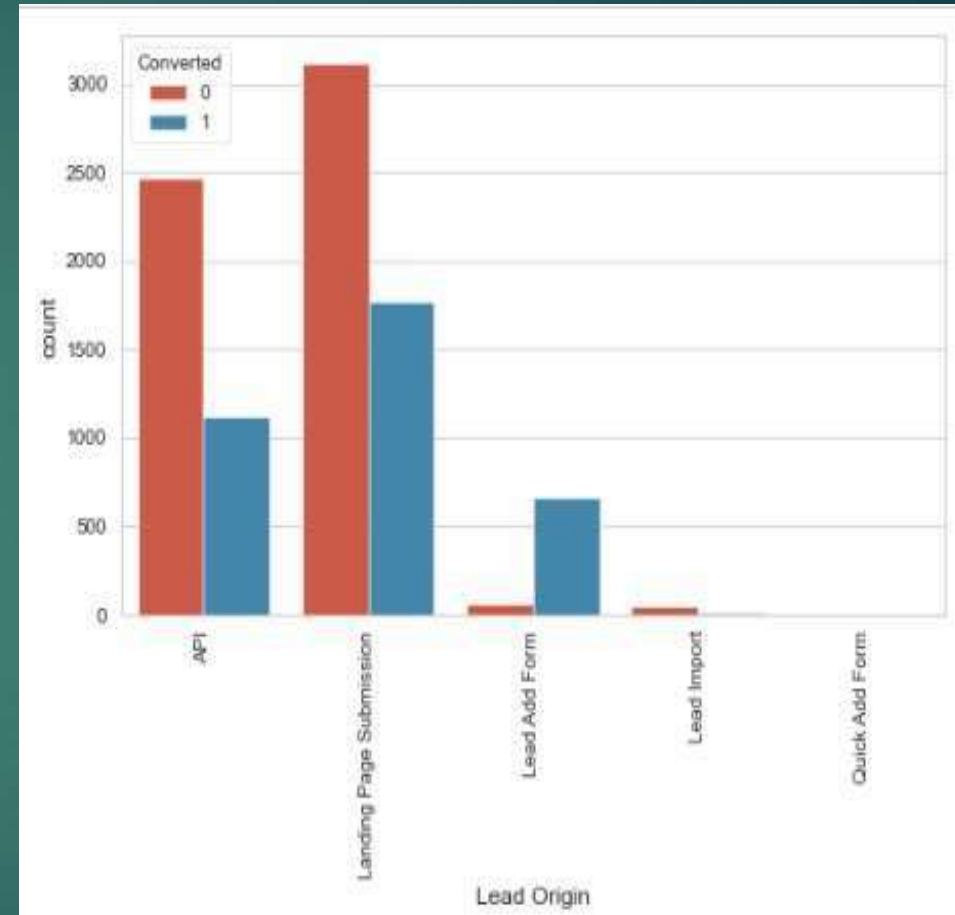
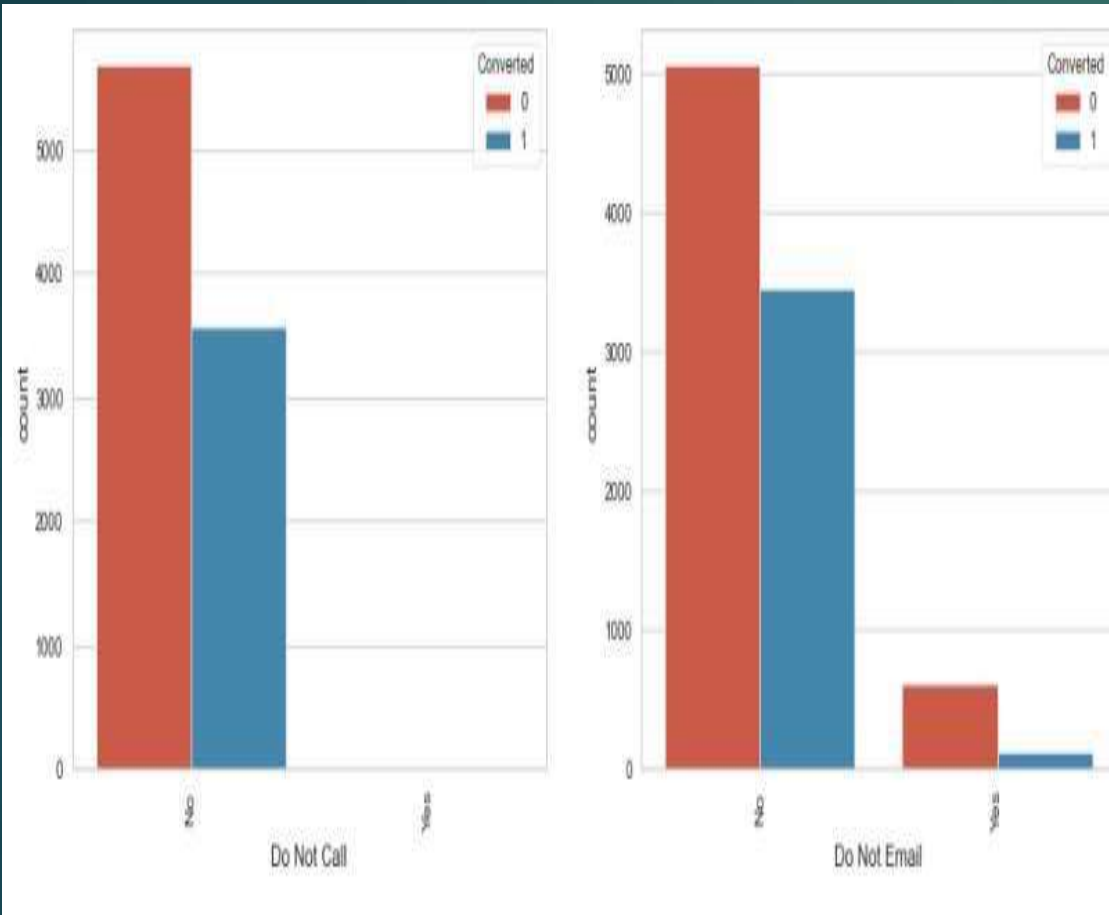
4



EDA

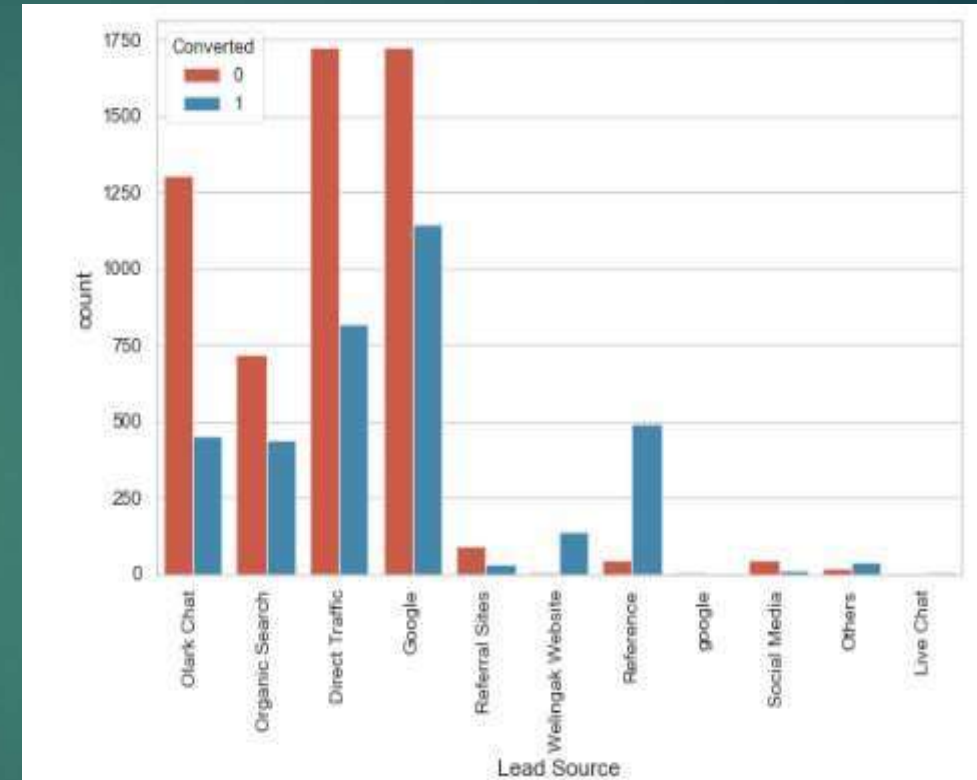
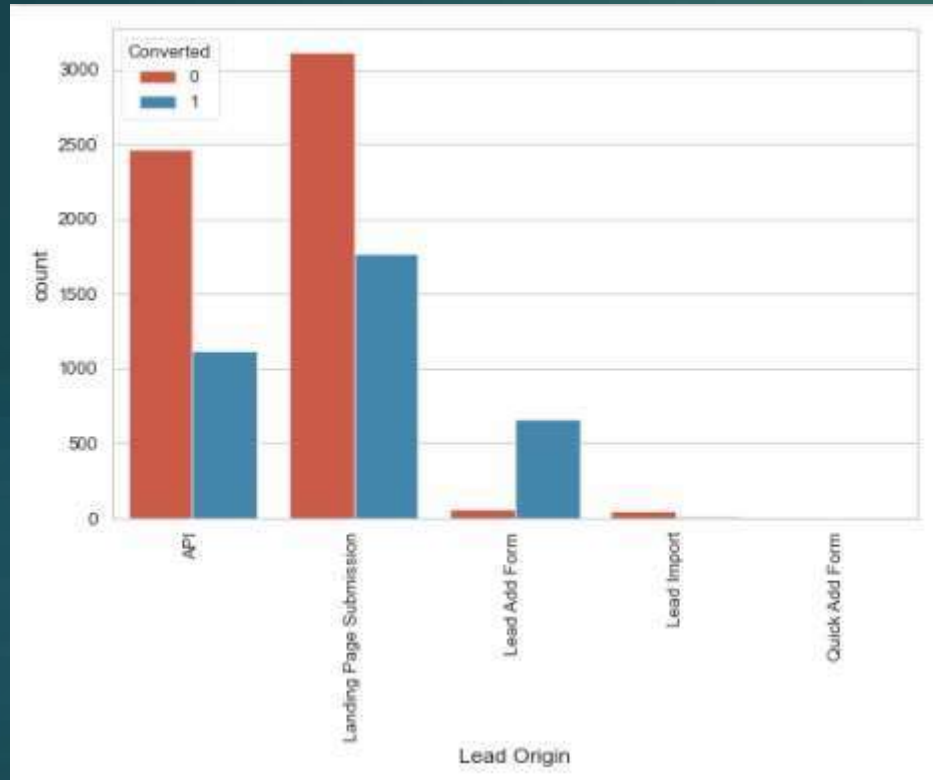
5

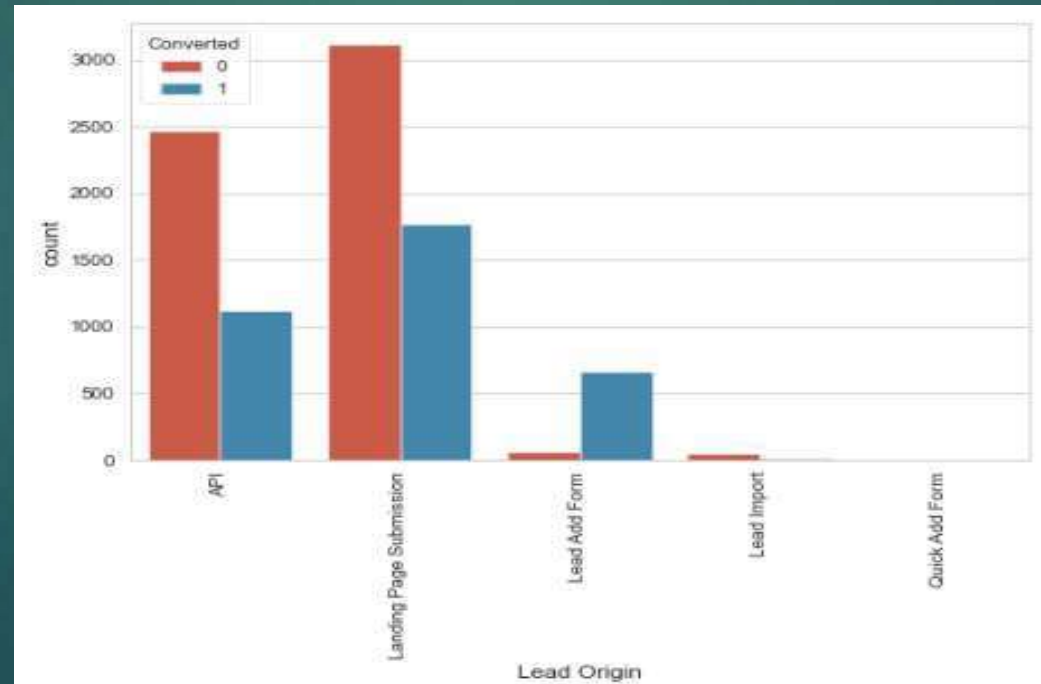
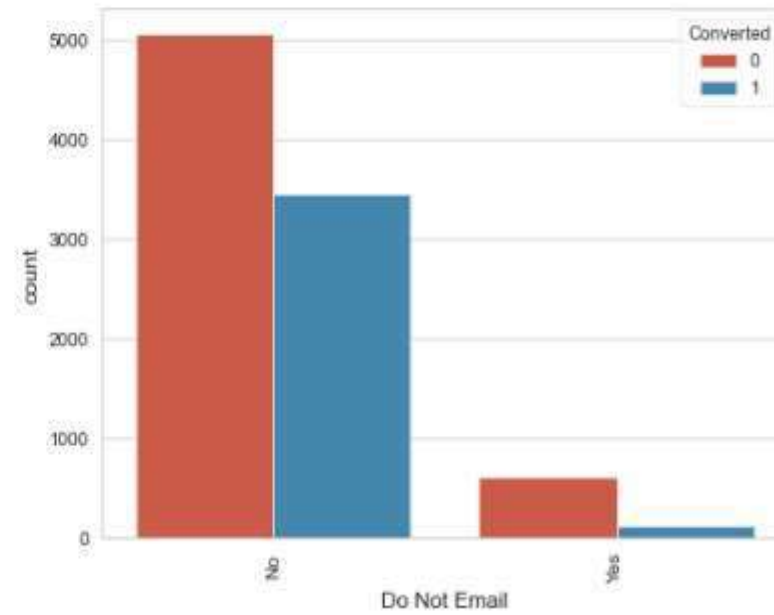
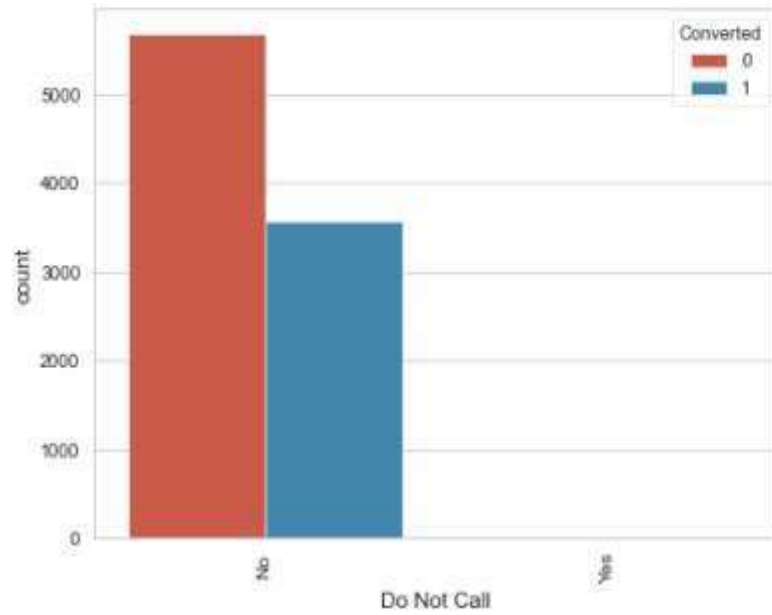




CATEGORICAL VARIABLE

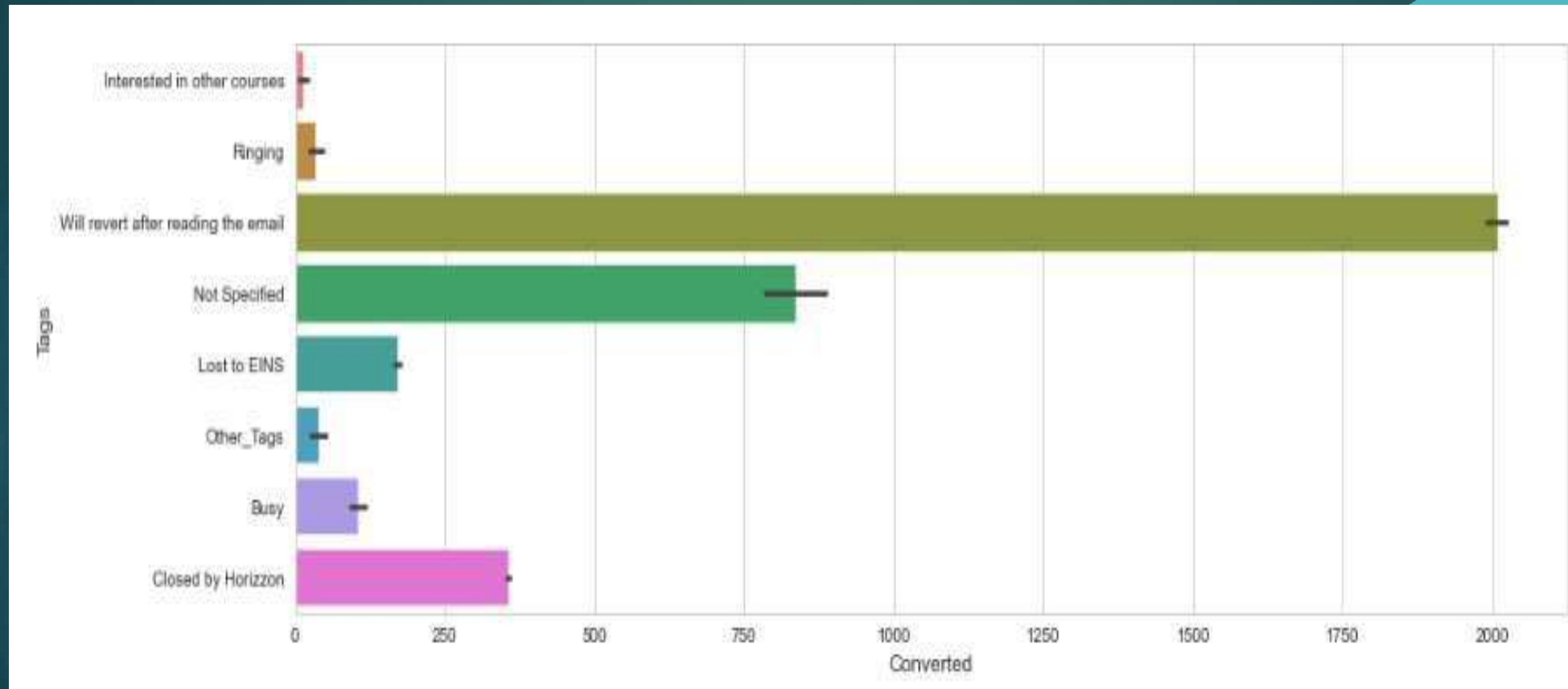
7





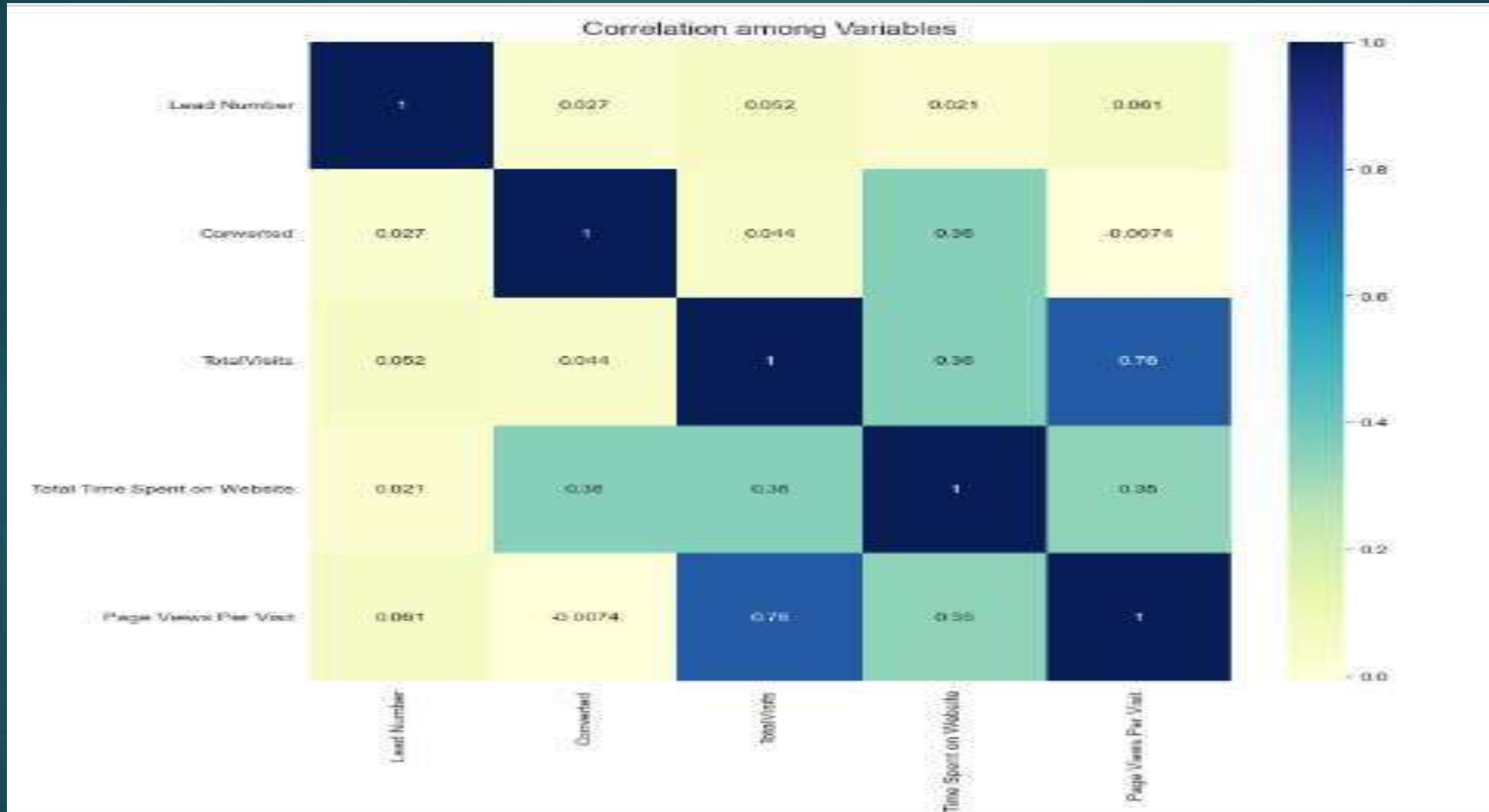
DATA CONVERSION

1. Dummy variables are created for all the object type variables.
2. Numerical variables are normalised.




CORRELATION AMONG VARIABLES

10

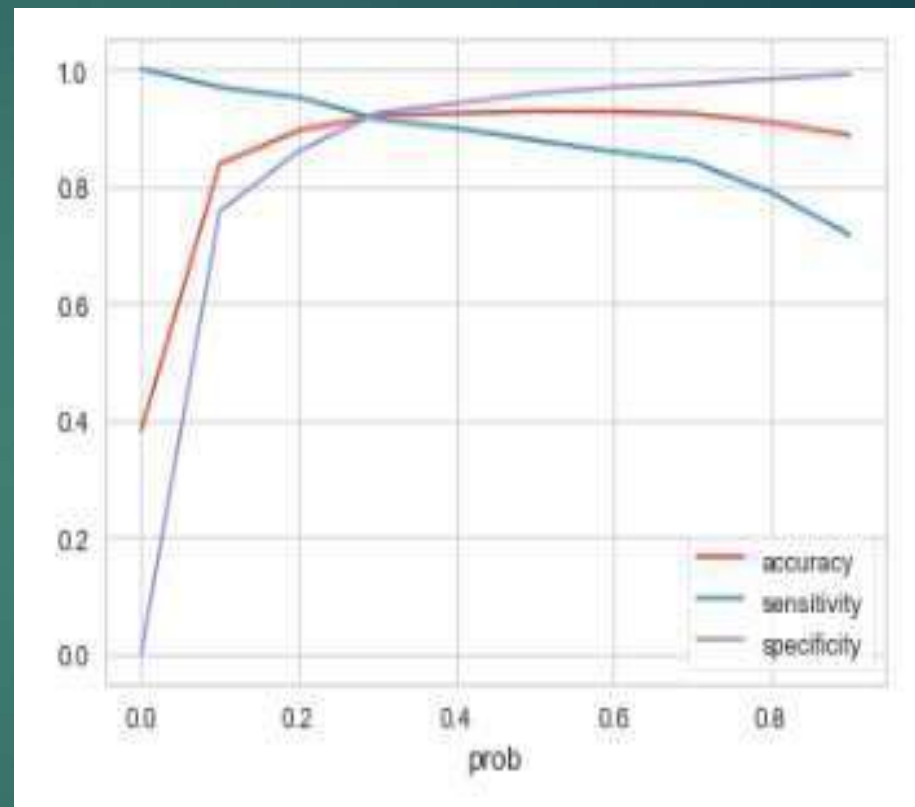
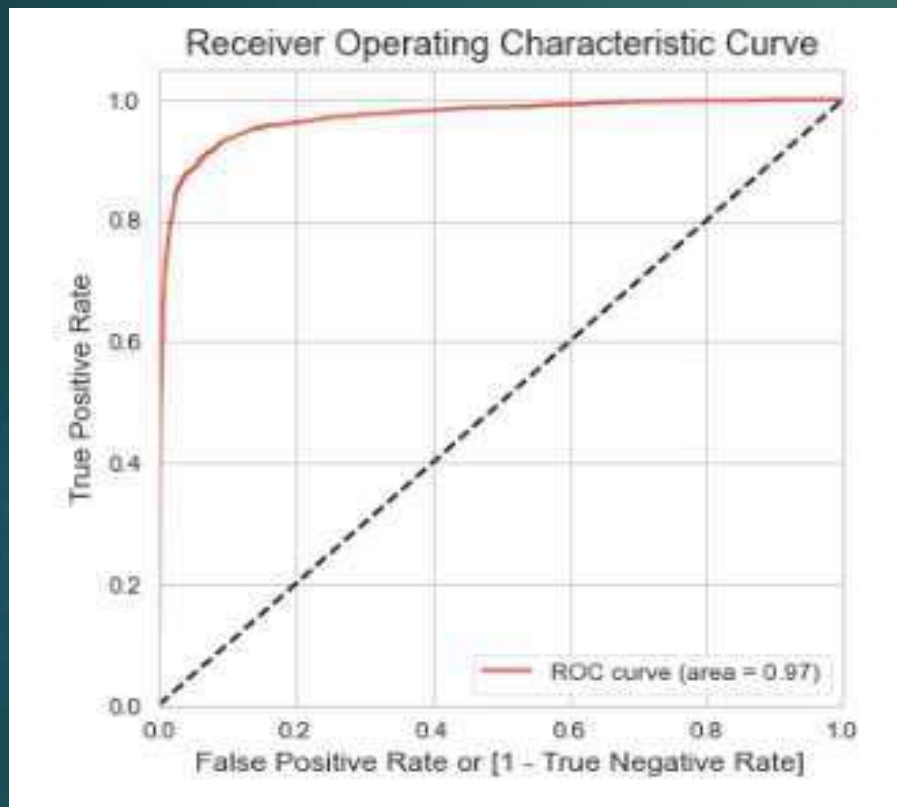


MODEL BUILDING

- ▶ The data is split into test and train split.
 - ▶ Recursive Feature Elimination (RFE) used for feature selection.
 - ▶ RFE are used in 15 variables as output.
 - ▶ Variables are removed whose p-value are greater than 0.05 and VIF value greater than 5. This is done do optimize the model.
 - ▶ Final predictions are done using test data set.
 - ▶ The overall accuracy achieved was 81%.
- 

ROC CURVE

12



The optimal cut off point is inferred as 0.35.

TEST SET

ACCURACY	92.15 %
SENSITIVITY	91.49 %
SPECIFICITY	92.50 %

TRAIN SET

ACCURACY	92.57 %
SENSITIVITY	91.19 %
SPECIFICITY	93.45 %

CONCLUSION

14

The potential buyers can be found using the variables that mattered most.

- ▶ The users who spend maximum time on the website
- ▶ The total no. of website visits
- ▶ The lead origin is lead add
- ▶ Google as lead source followed by direct traffic, organic search and welingak website
- ▶ Working professional as occupation

With the above all variables the X edu company can have higher chance to attract the potential buyers.