Lead Scoring Case Study

Ali Faiza
Abhishek Saxena
Anurag Shukla

Problem Statement

- X Education is an education firm selling online courses to students and professionals.
- They generate leads through their website and other methods, but their lead conversion rate is very low, ex. if they generate 100 leads only 30 get converted.
- To improve their conversion rate they need mode which can guide them well and help them convert their hot leads.
- It will help the interns and sales department on focusing on leads which can be converted rather than on leads which are waste of time.

Objective

- X education wants to deploy a model for their department.
- They want to identify most promising leads.

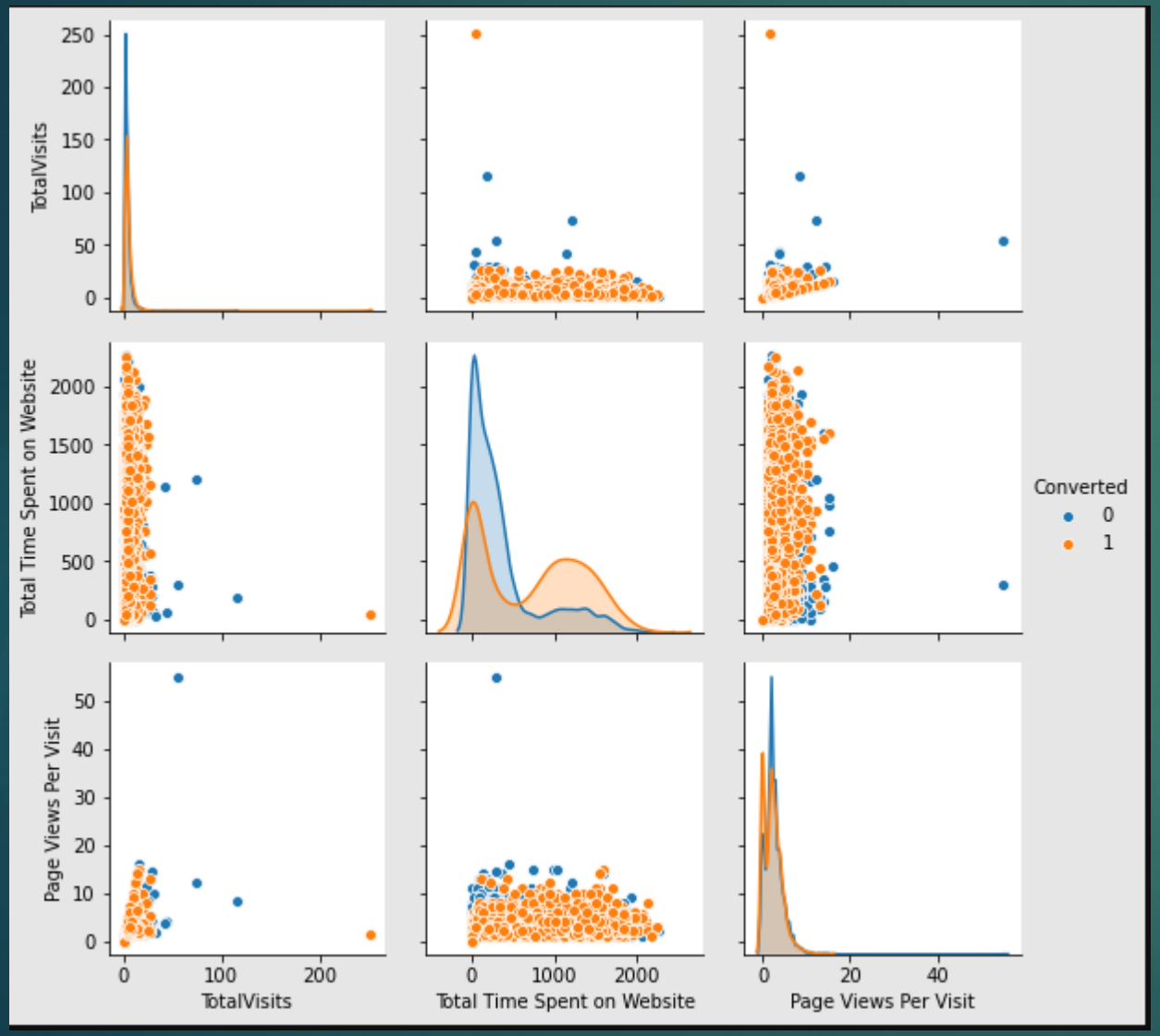
Methodology

- Reading the data checking the head, description and info about it.
- Secondly checking the null values in columns using .isnull().sum().
- Dropped the features which were having more than 3000 null values.
- Checked outliers if necessary.
- Removed all null values from features which were necessary and cannot be dropped.
- EDA
- Features were scaled and dummy variables were created for regression model.

Details On Data

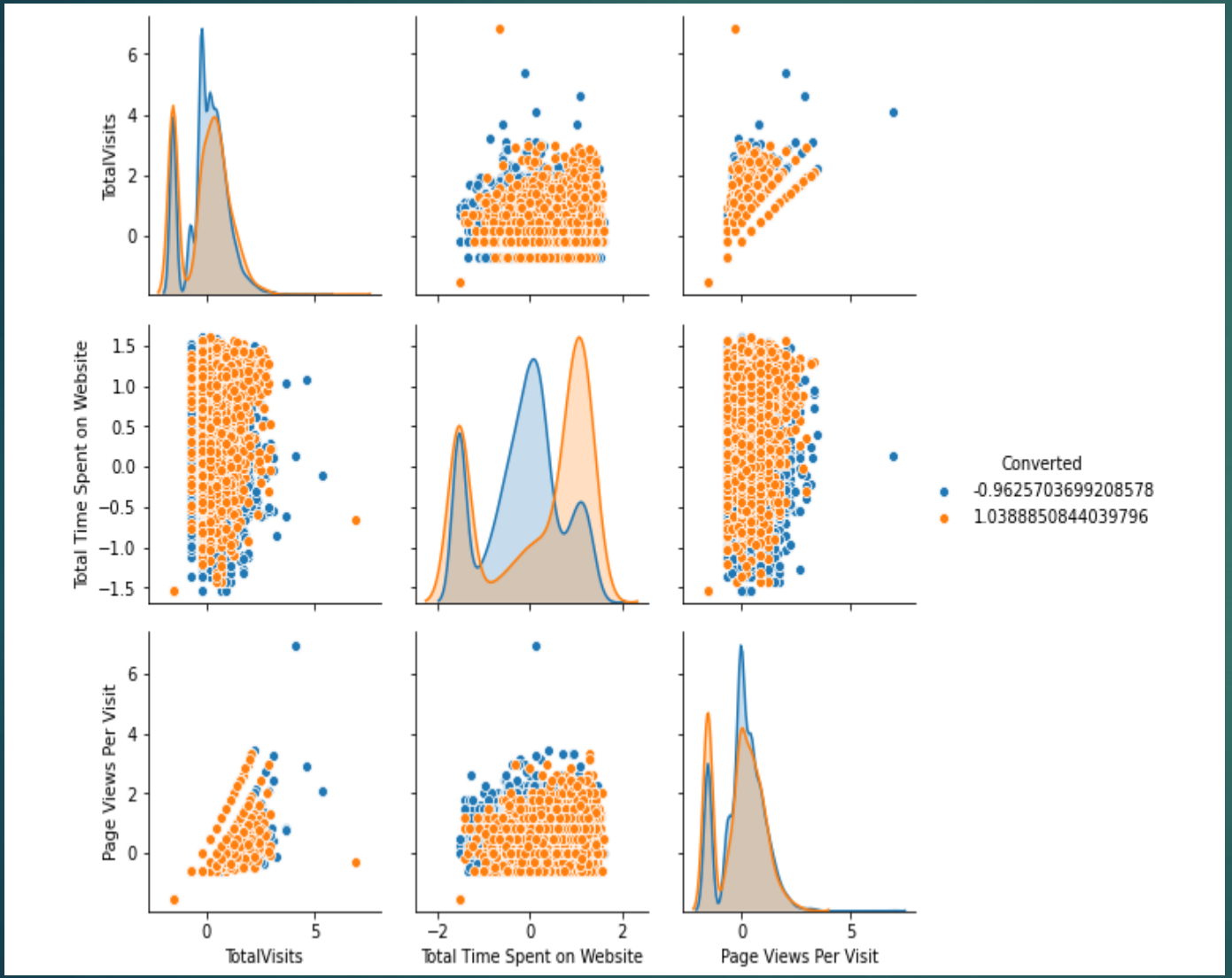
- It has a shape of (9240,37)
- Some features like ‘How did you hear about X Education’, ‘What is your current Occupation', 'What matters most to you in choosing a course’ had the most number of null values.
- All features with more than 3000 null values were dropped such as *Tags, Lead Quality, Asymmetric Activity Index* and some more.
- Prospect ID and Lead number were dropped as they were of no use to us.

EDA



- Checking features against target variable ‘converted.’

EDA when data transformed



Dummy Creation

- Numerical features were normalized.
- Created dummy variables for all object type features.
- Use dtype = int in dummy formula to replace True/False with 1/0 in dummy creation.

Building Model

- Firstly data was split into test and train.
- Split ratio was taken 70:30.
- RFE used for feature selection.
- Used RFE with 15 features selection.
- Features selected must have p-value greater than 0.05 and VIF value greater than 5.
- Making predictions on test data.

CONCLUSION

- Features that are most important are mentioned below
- When the current profession is in line to courses searched on X education X must counsel the candidates for choosing the correct course.
  - The total time spent on website
  - Total no of visits.
  - When Lead source:
    - Google
    - Organic search
    - Traffic
    - Welingak Website