


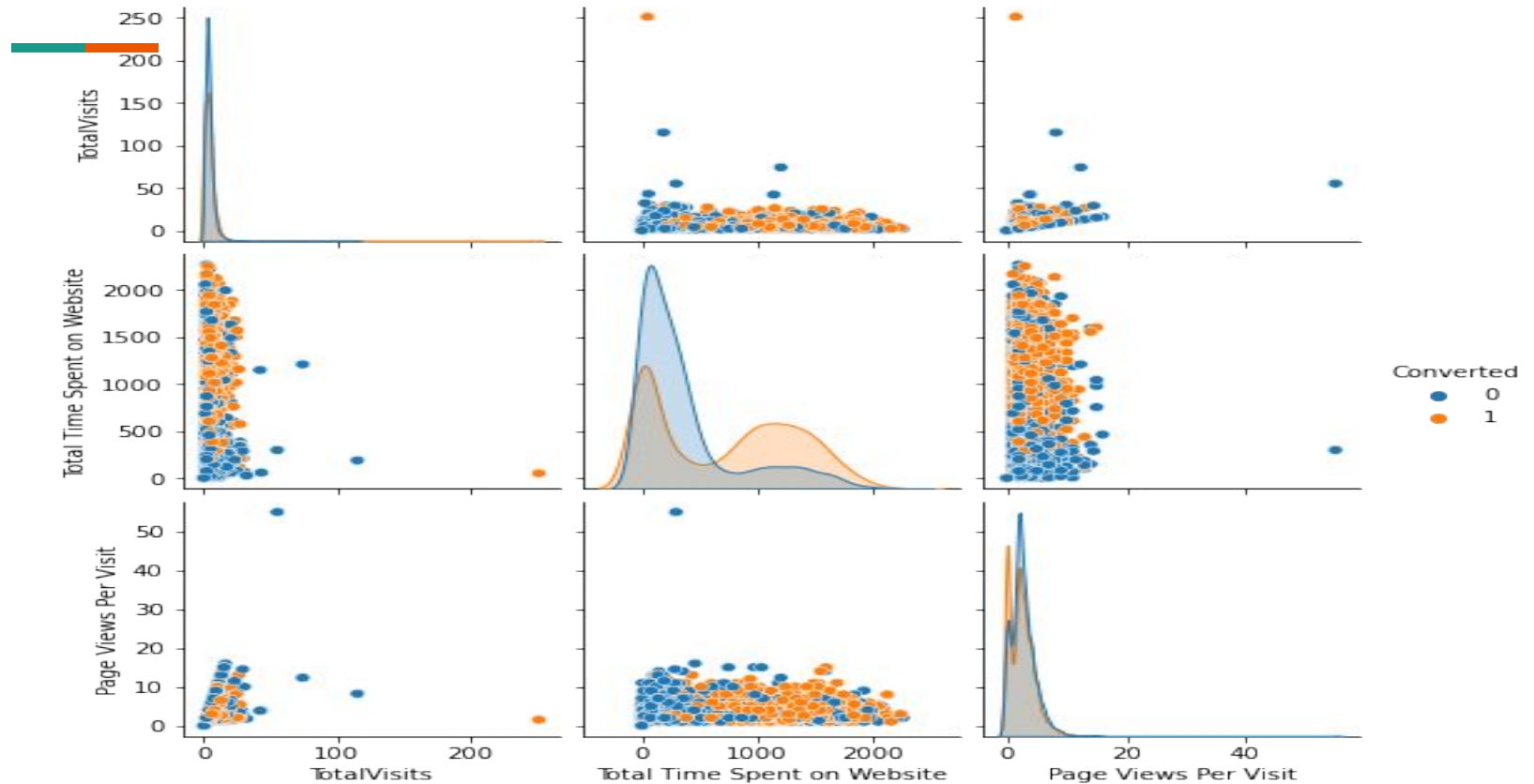
- 
- 1.Arpan pal
 - 2.saurabh yadav



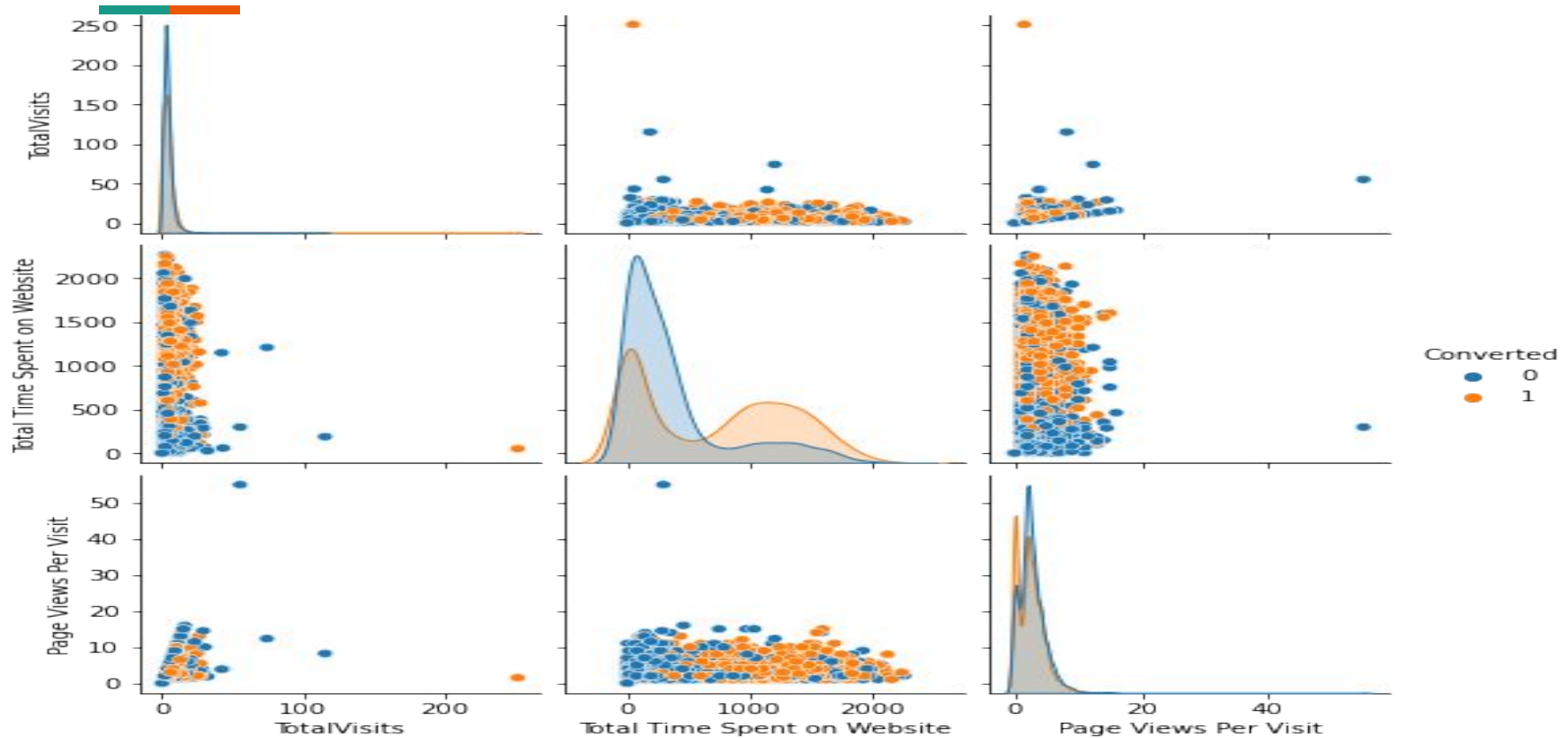
DATA MANIPULATION

1. total row = 9240 . columns = 37
2. checked the “select ” statement and removed them by row
3. removed the prospect id and lead number as not required for the analysis
4. dropping the column more than 32 % of missing value and final data after cleaning was 69%.

Comparing with target variable



Compare target variable with other





Model building

Splitting the data into training and test set 70% 30%

Used RFE for feature selection

Build the model by removing the variable whose p-value is more than 0.5

And vif value is greater than 5.

Accuracy pending



Conclusion

It is found that the most top three variables that matter the most are

1. The total time spent on the website, total number of visits
2. When the lead source was Google, direct traffic, organic search
3. When the current occupation is working professional it has very potential to convert the lead.