

Lead Scoring Case study

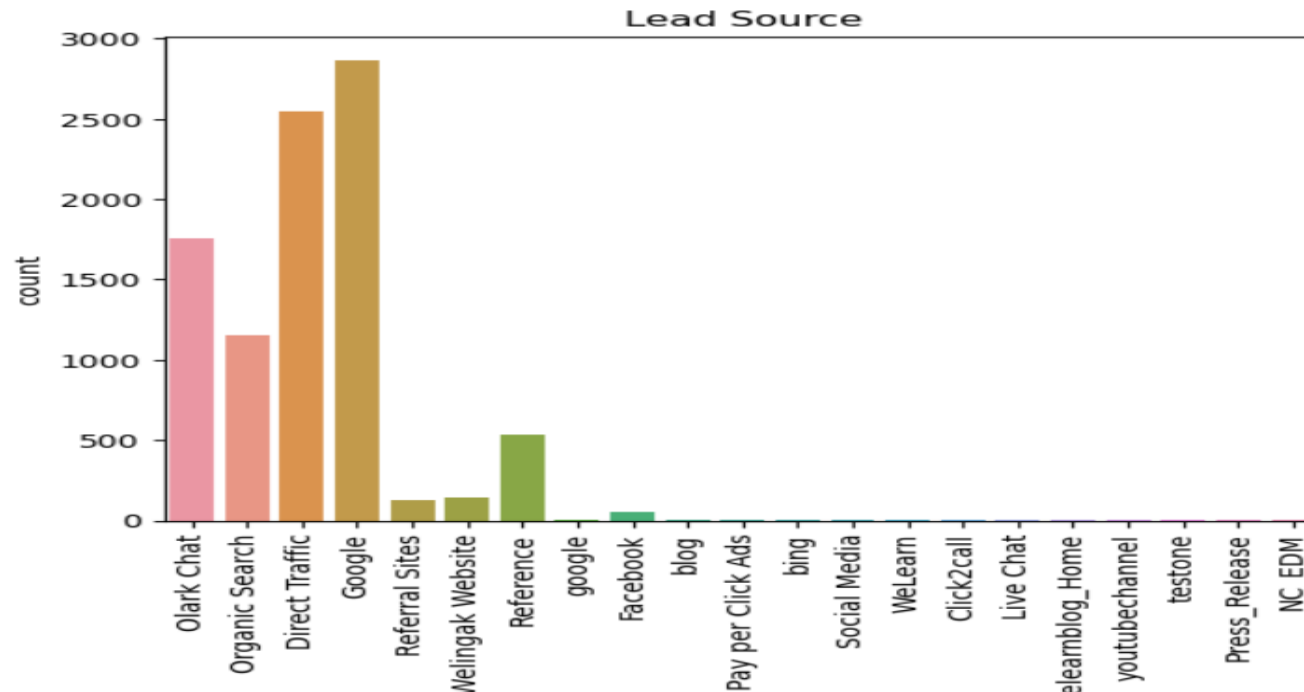
Problem Statement

This analysis is done for X Education and to find ways to get more industry professionals to join their courses. The basic data provided gave us a lot of information about how the potential customers visit the site, the time they spend there, how they reached the site and the conversion rate.

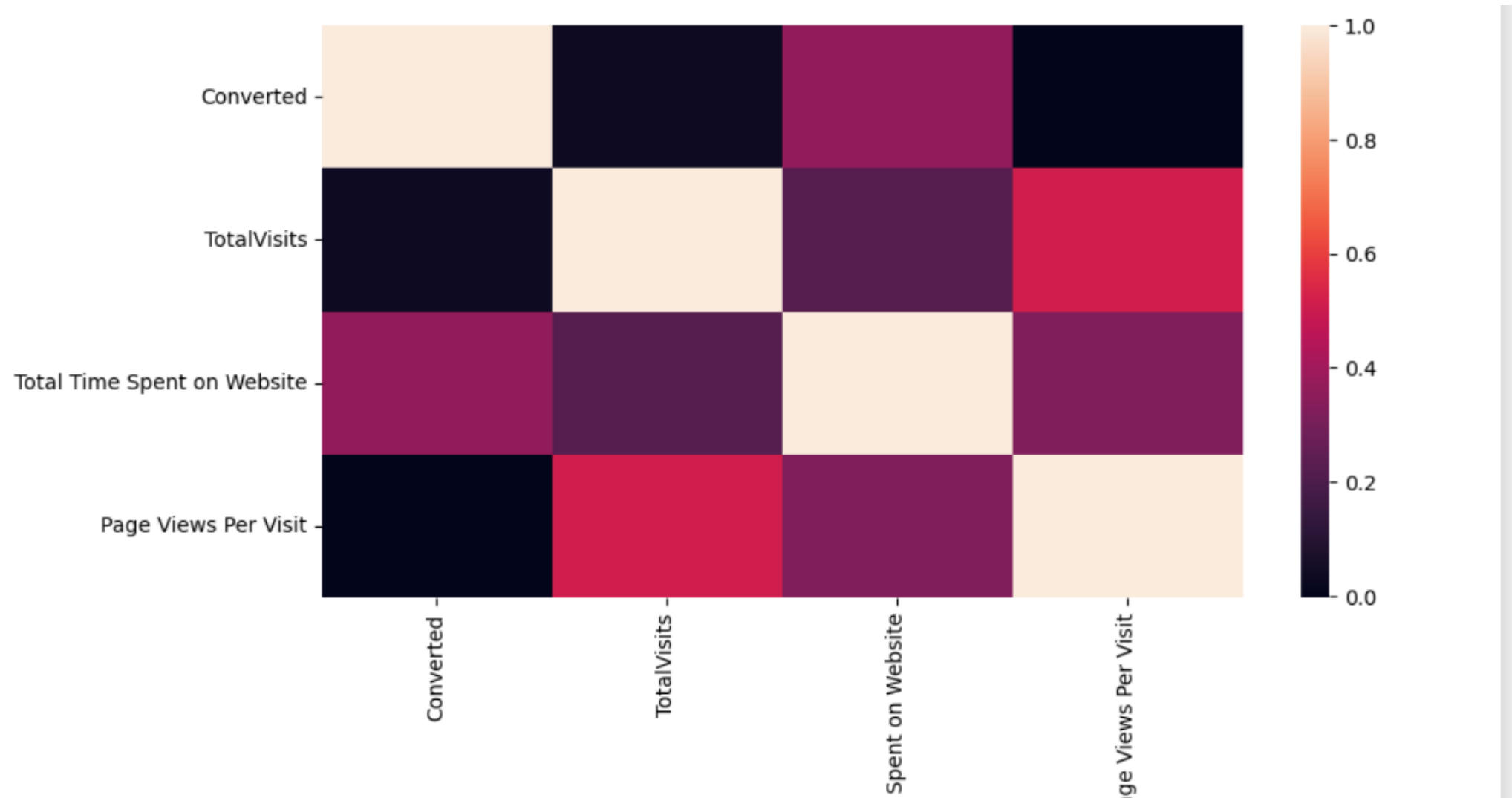
1. For this case study have clean up data and found some column which are not relevant

- Total Time Spent on Website
- Total Visits
- Lead Source with elements Google

- 2. The lead source where we can target is Google



3. From heat map we can see that time spent on website is very much related with converted variables so, in this case we can give more focus how the content of website can be more reliable so that customer can rely on that and so conversion rate can be increased.



4. after that checked multicollinearity using VIF method

5. And split data into train data and test data

```
# Split the data into train and test
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.7, test_size=0.3, random_state=10)
```

6. After creation of model and testing it got the accuracy score as 75%

Conclusion:

That means the variables which have been taken into consideration can lead to increase in conversion rate :

1) Total time spent on website:

We can give more importance to how we can improve content on website so that so that we can check whether total time spent on website has been increased or not if content is good then total time will be more and indirectly it can increase conversion rate

2) Lead source:

As we have seen Google is the lead source that means it will be good if we can target customers on Google for advertisement because it is a leading source of knowing this can be used for advertisement