Lead Score Assignment

PPT

Problem Statement:

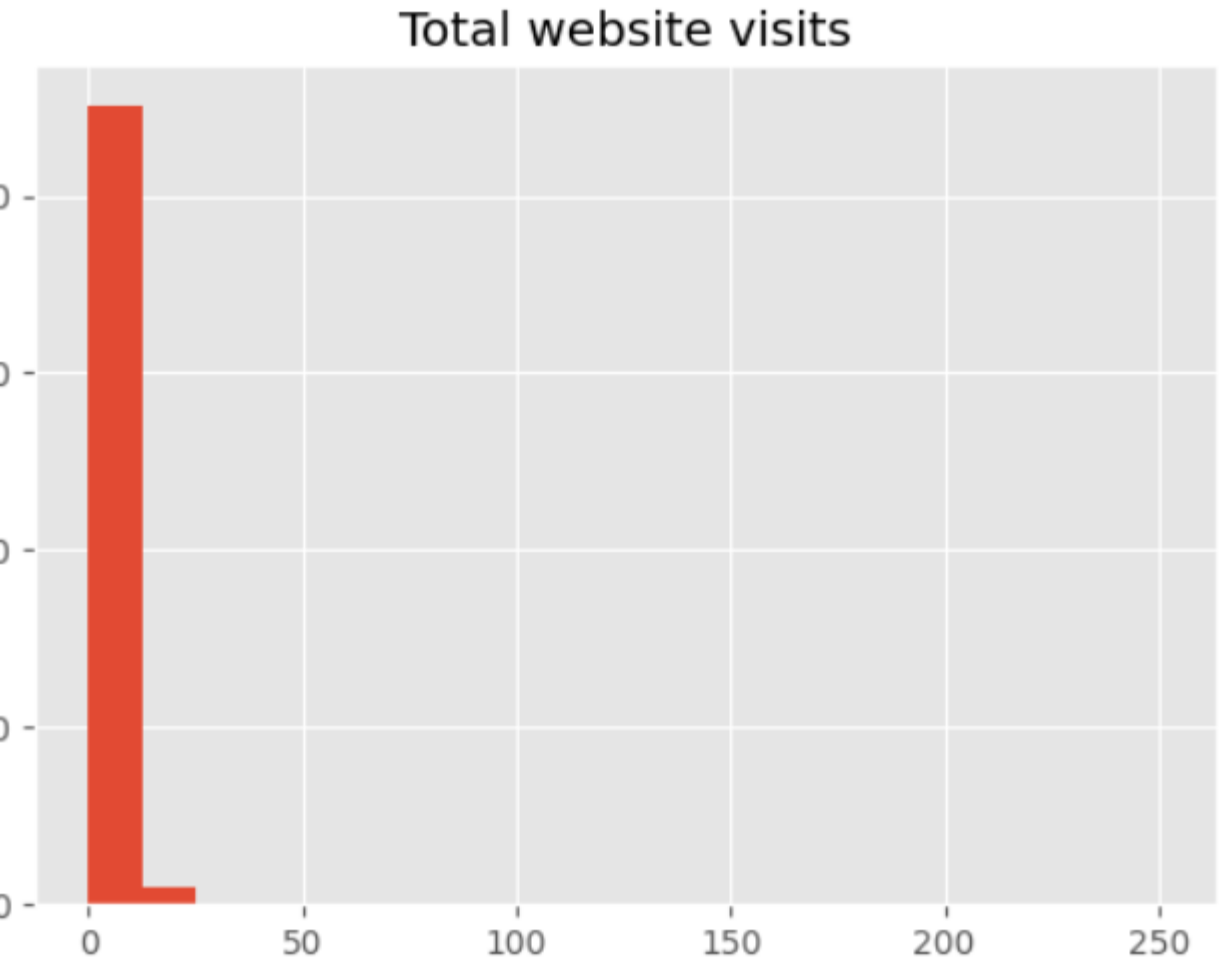
X Education sells online courses to industry professionals. The company markets its courses on several websites and search engine like Google.

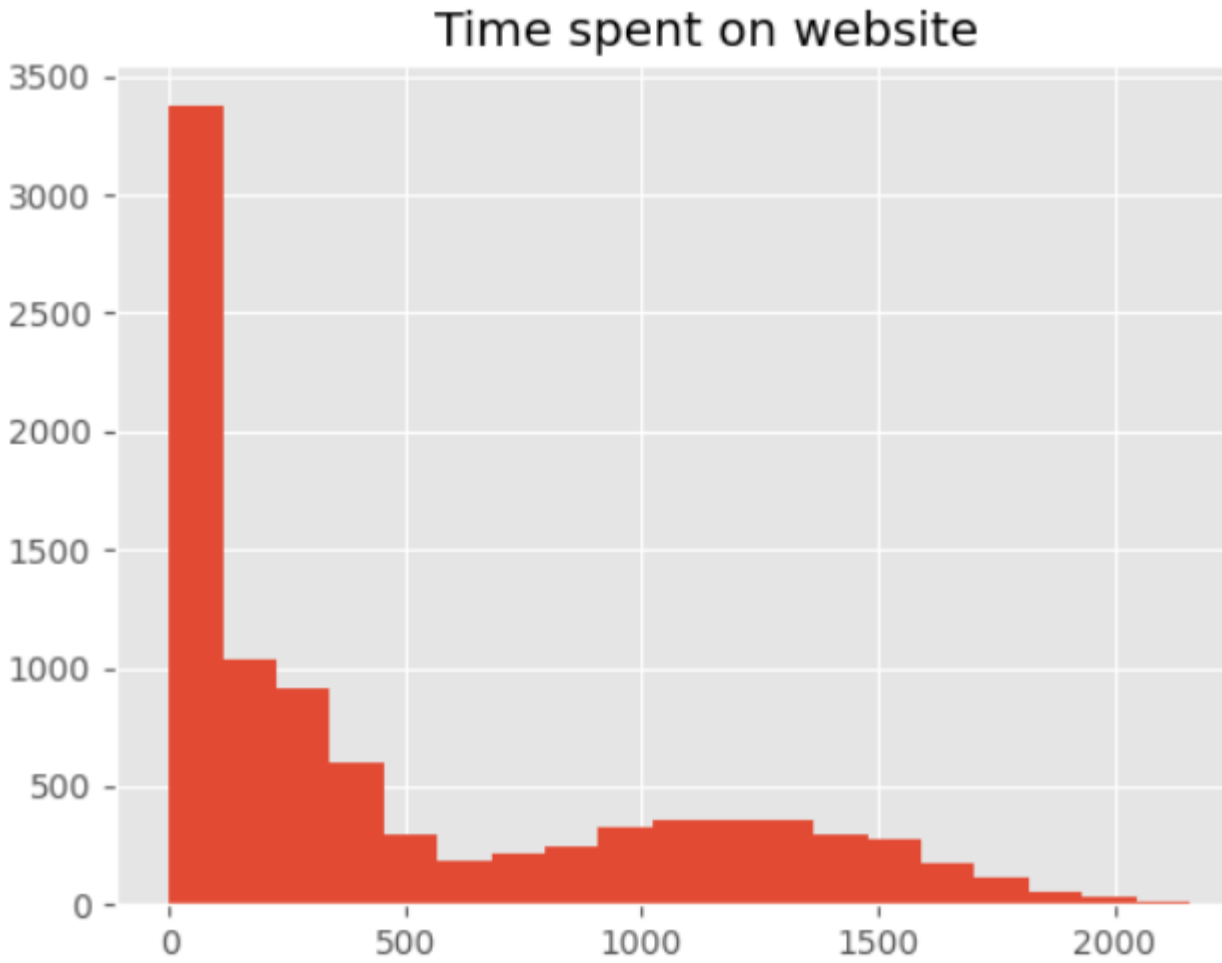
Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos When these people fill up a form providing their email address or phone number, they are classified to be a lead. Moreover, the company also gets leads through past refferals.

Once these leads are acquired, employees from the sales team start making calls, writing emails, etc. Through this process, some of the leads get converted while most do not. The typical lead conversion rate at X education is around 30%.

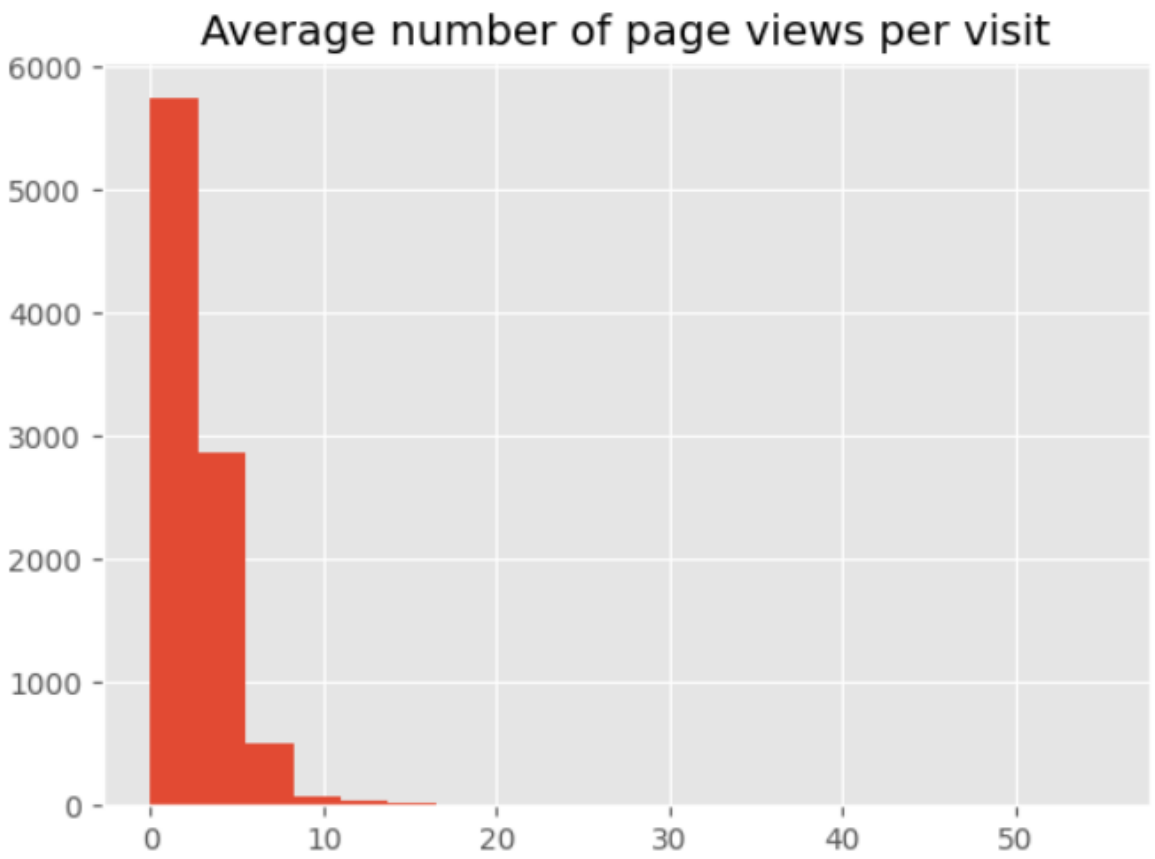
Strategy:

- Source the data for analysis
- Clean and preapare the data
- Exploratory Data Analysis
- Feature Scaling
- Splitting the data into Teat and Train dataset
- Building a logistic Regression model and calculate lead score.
- Evaluating the model by using different metrics- Specificity and sensitivity or precision and Recall.
- Applying the best model in Test data based on the Sensitivity and Specificity Metrics.

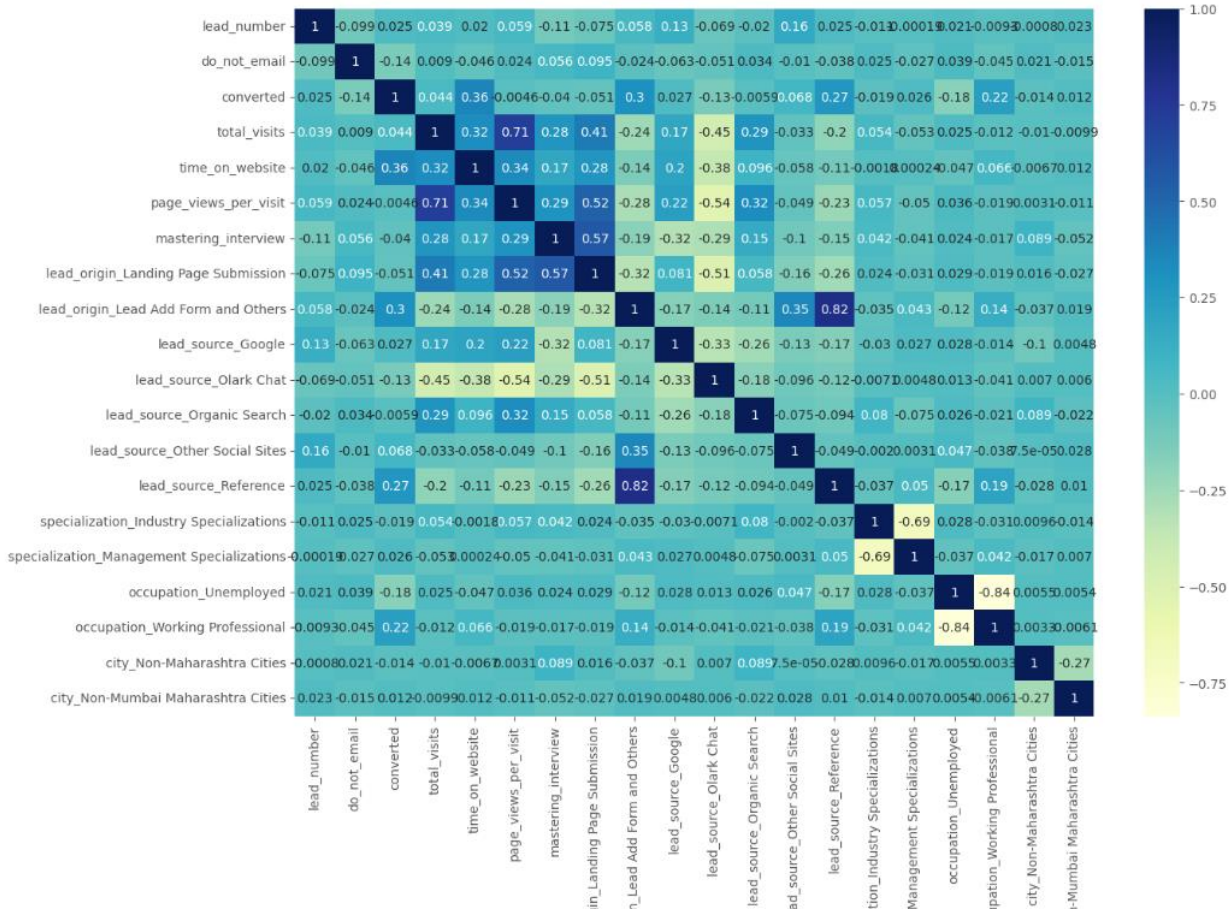

Total website visits
0
50
100
150
200
250


Time spent on website
3500
3000
2500
2000
1500
1000
500
0
0
500
1000
1500
2000

Average number of page views per visit:



Below is the correlation matrix



Conclusion:

- While we have checked both sensitivity-specificity as well as precesion and recall metrics, we have considered the optimal cut off bbased on sensitivity and specificity for calculating the final prediction.
- Also the lead score calculated shows the conversion rate on the final predicted model is around 80% and 79% in test set.
- The top 3 variables that contribute for lead getting converted in the model are
- 1. Total time spent on website.
  2. Lead Add form from lead origin
  3. Had a phone conversion from last notable activity.
- Hence overall this model seems to be good.