

Lead Score Case Study

PRESENTED BY

KEERTHANA RAJAGOPAL

DHANYA PM

RINA DINDA

Problem Statement

- ▶ X Education sells online courses to industry professionals.
- ▶ X Education gets lots of leads, it's lead conversion rate is very poor. For E.g. 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 identified this sets 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.

Business Objectives

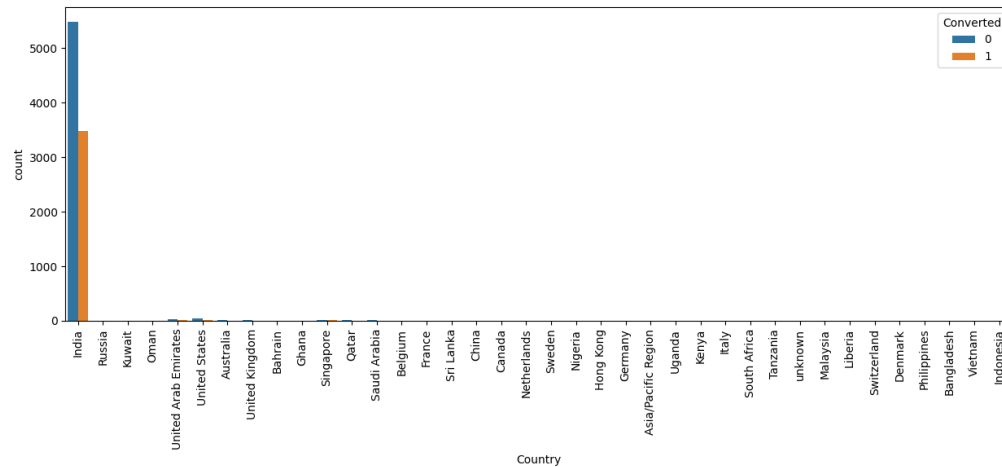
- ▶ X Education wants to know most promising leads.
- ▶ For that they want to build a model which identifies the hot leads.
- ▶ Deployment of the model for the future use.

Solution Methodology

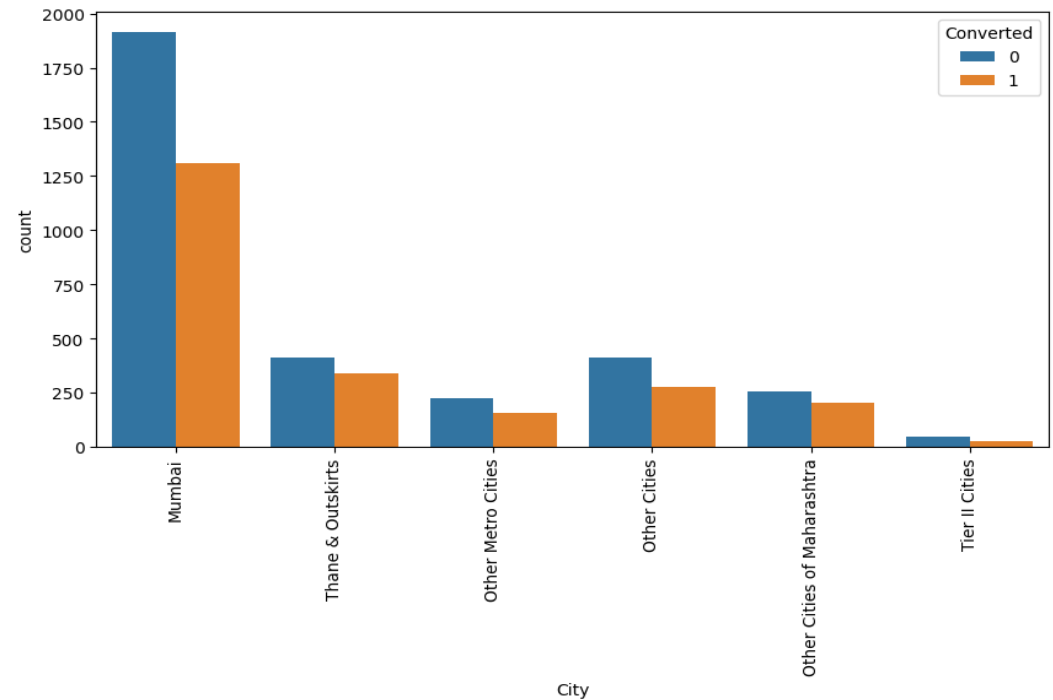
- Data Cleaning and Data Manipulation.
- EDA
- Featuring Scaling & Dummy variables and encoding of data.
- Classification Technique: Logistic regression used for the model making and prediction.
- Valuation of Model.
- Model Presentation.
- Conclusions and Recommendations.

EDA

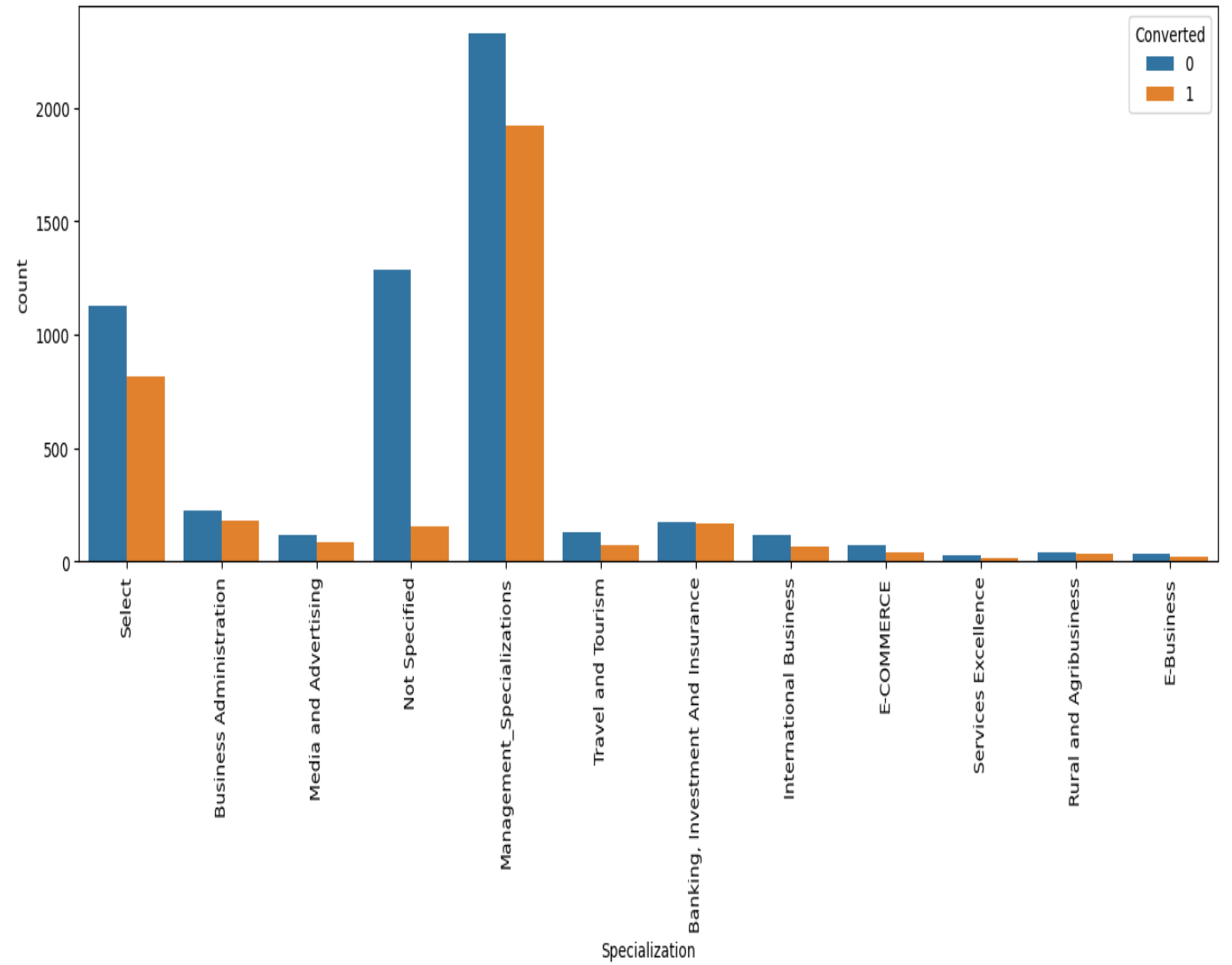
We can see the Number of values for India are quite high which is nearly 97% of the Data.



If we look city wise in India, Mumbai holds a very high rate in Conversion.

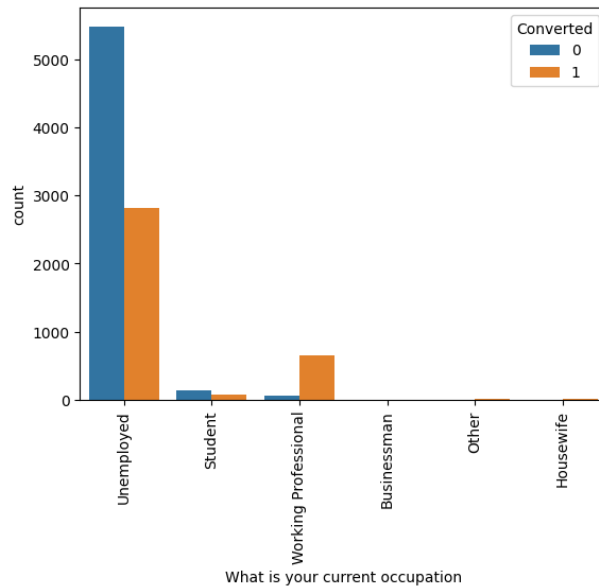


As Per the Graph, we can conclude that among all the specialization Management holds the higher number of leads as converted

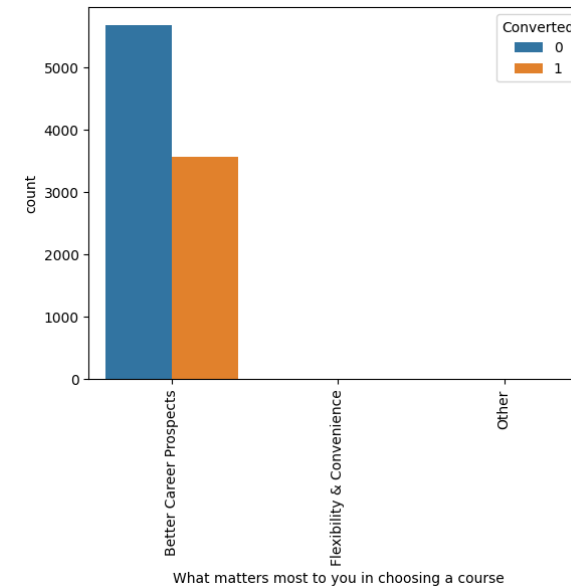


Analysis:-

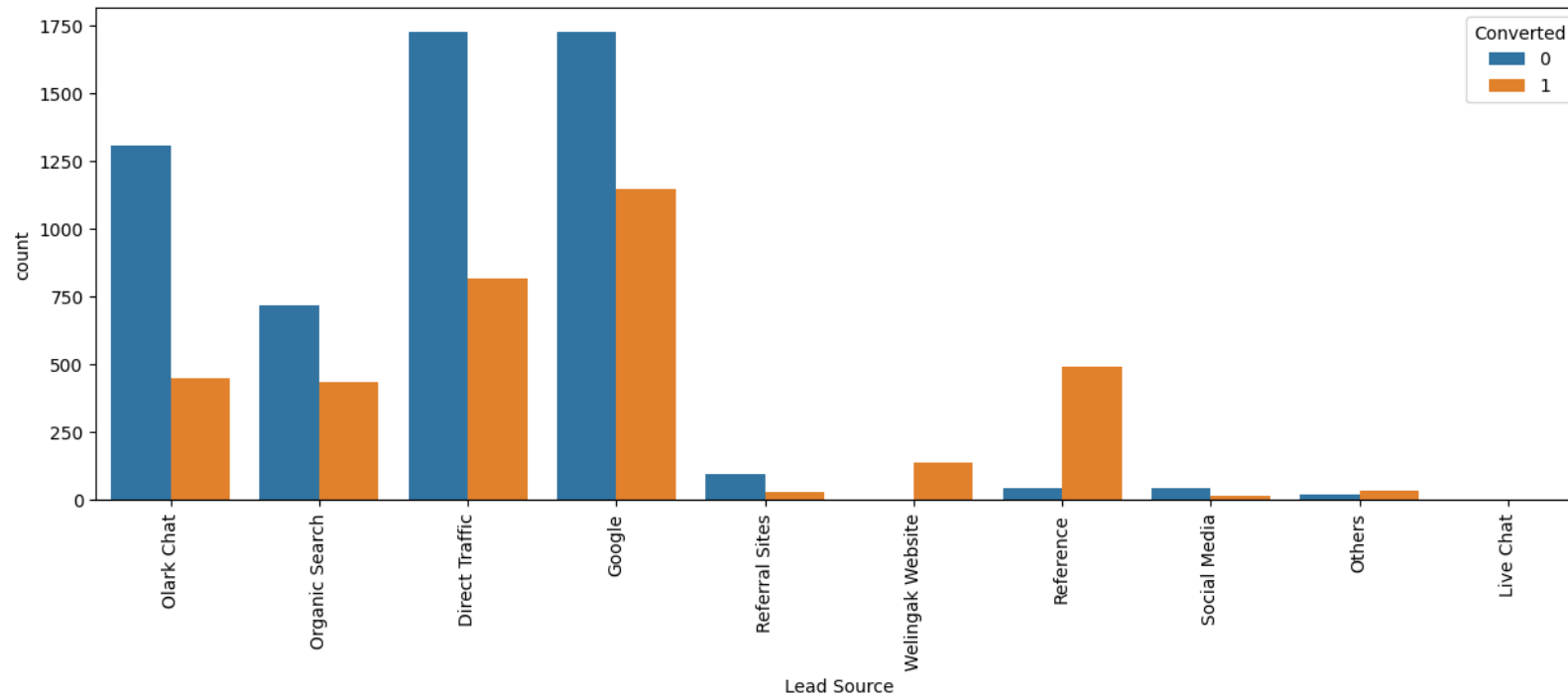
Unemployed Category shows the highest converted leads.



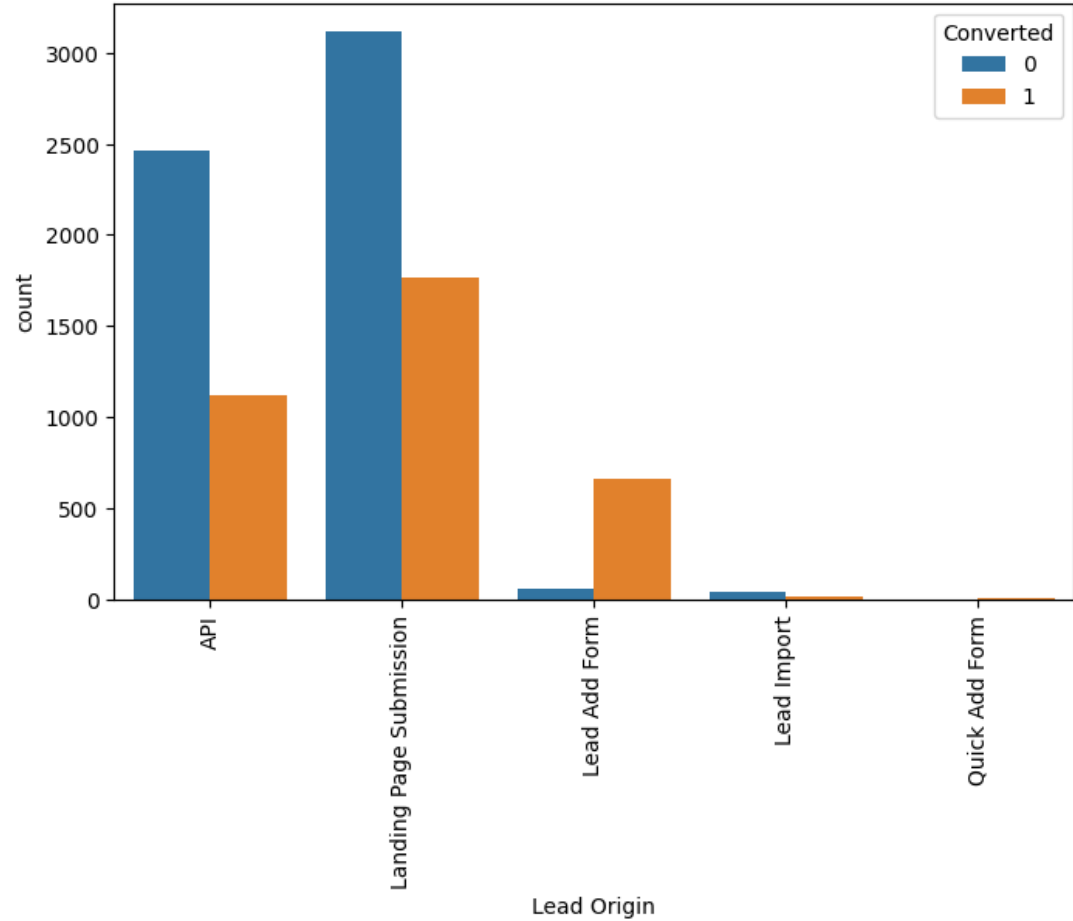
Reason: For Better Career Prospects



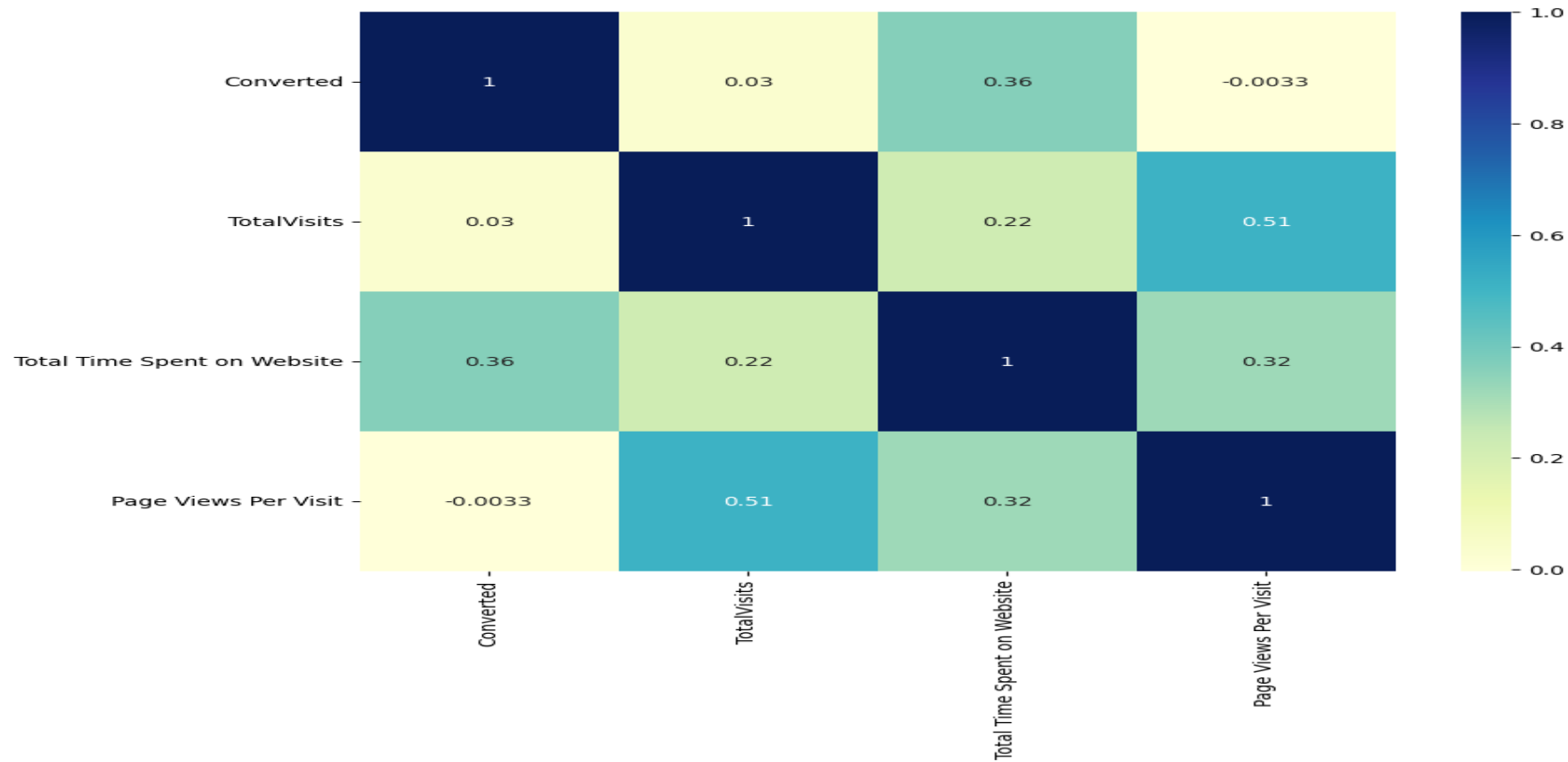
Maximum number of leads are generated by Goggle and Direct Traffic.



- ✓ API and Landing Page submission brings higher number of leads as well as conversion rate.
- ✓ Lead Add Form has a high conversion rate but the count of leads is very low.
- ✓ Lead Import and Quick Add form get very few leads.

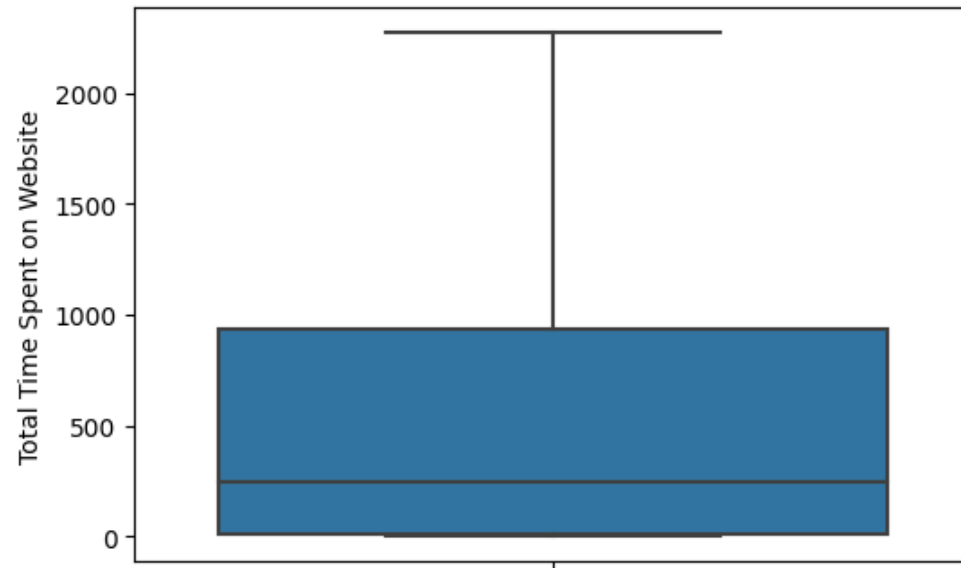


Correlations of Numeric Values:-

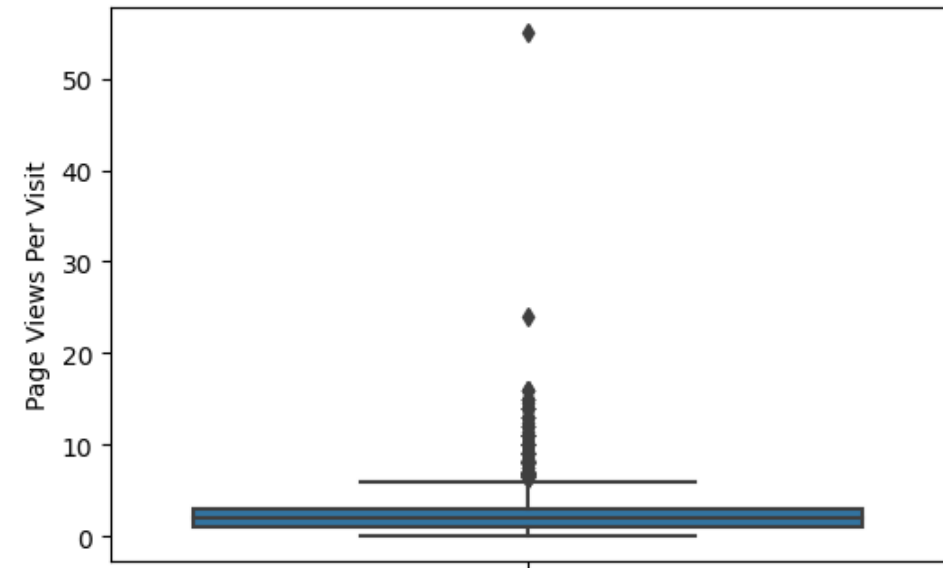


Views:-

Total time spent on Website

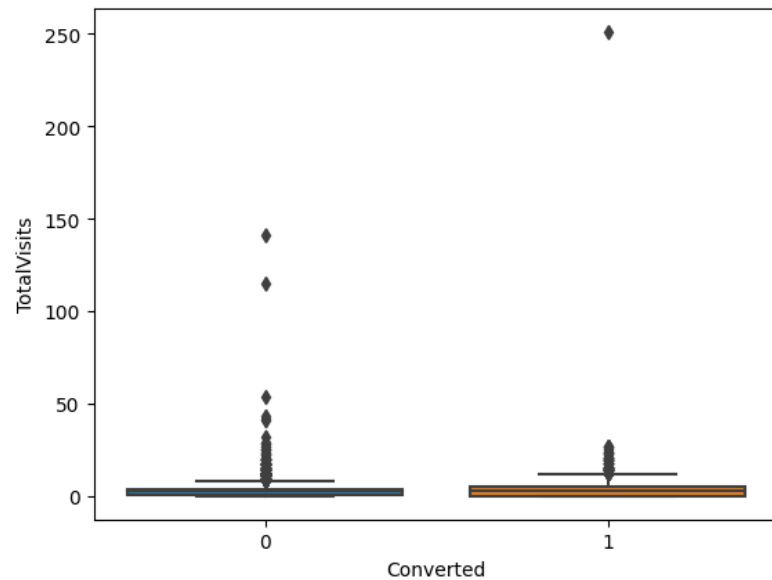


Page View per Visit

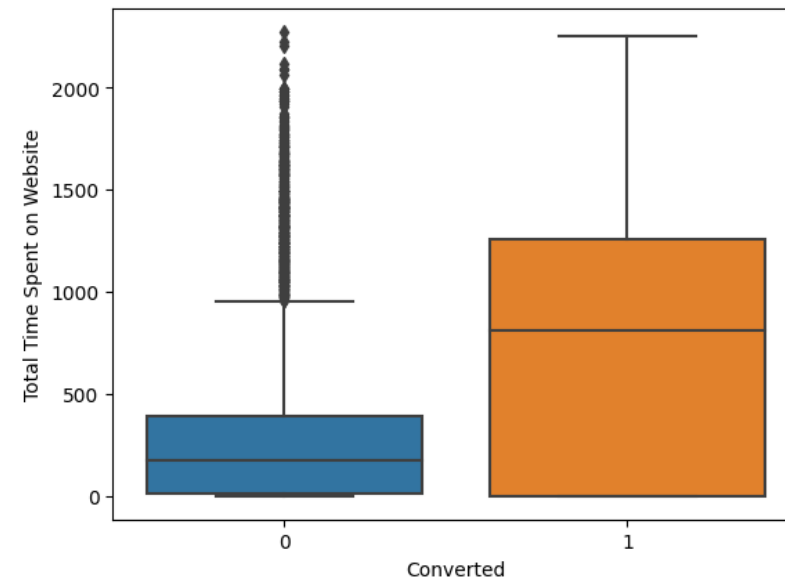


Comparison with Converted Variable

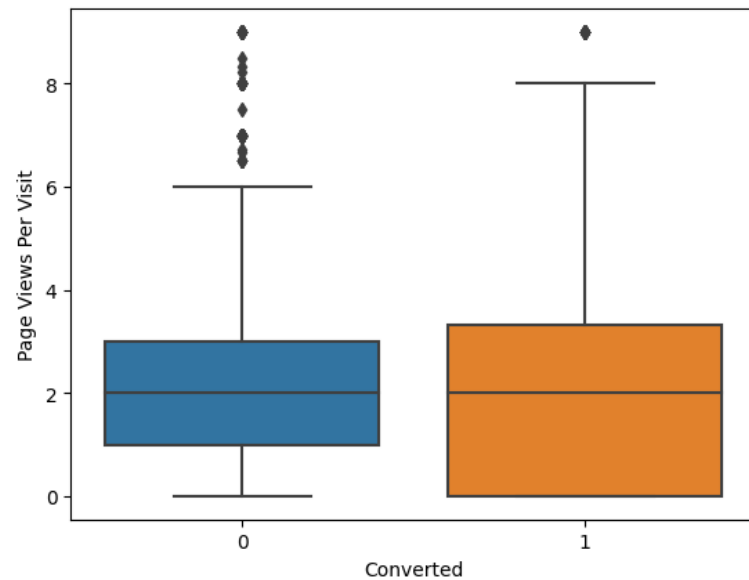
Total Visits v/s Converted Variable



Total time spent v/s Converted Variable



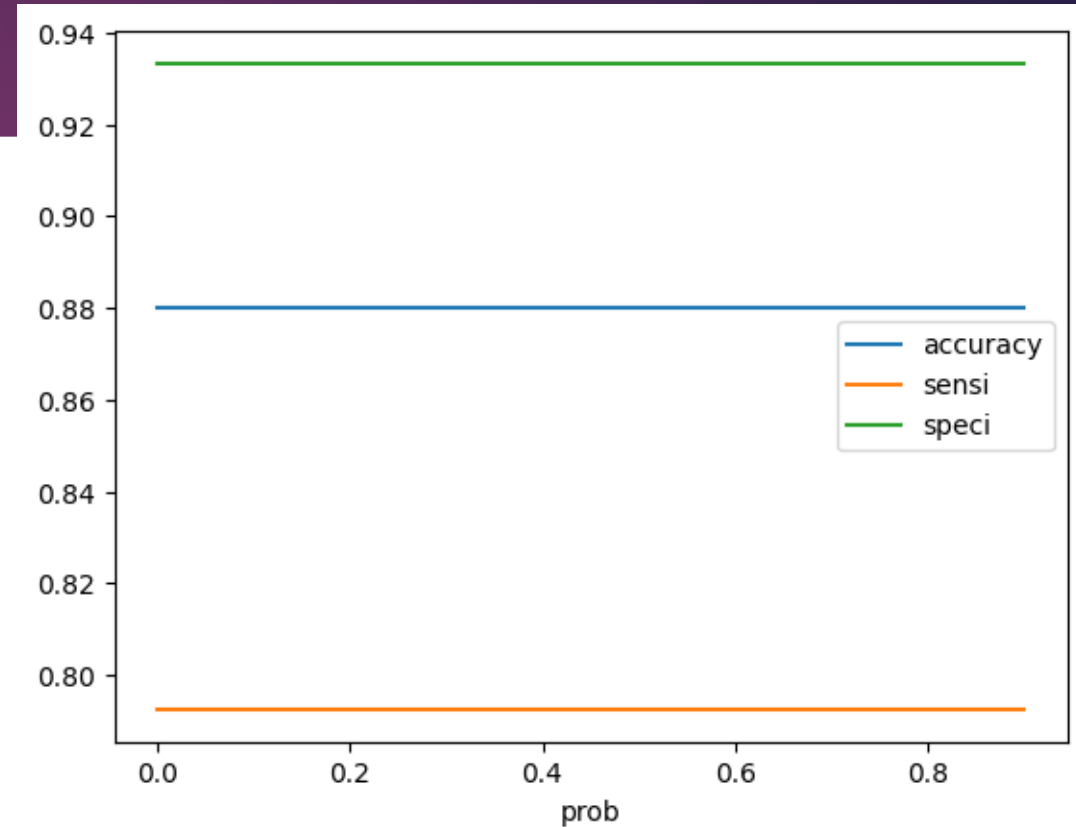
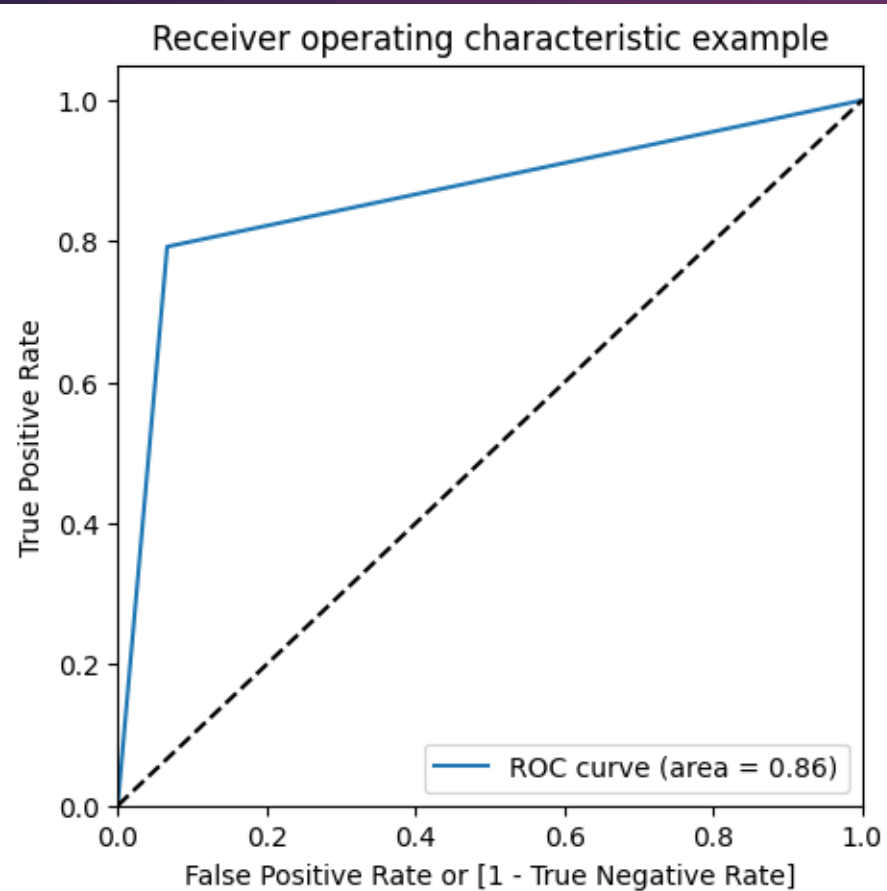
Page views per visit v/s Converted variables



Model Building

- ▶ Splitting the Data into Training and Testing sets.
- ▶ Use RFE for Feature Selection.
- ▶ Predictions on test data set.
- ▶ Overall Accuracy 91%.

ROC Curve



Probability where we get balanced sensitivity and specificity.
From the graph it is visible that the optimal cut off is at 0.3

Conclusion:-

- It was found that variables that mattered the most in the potential buyers are:-
- ▶ Total time spent on the Website.
- ▶ Total number of visits.
- ▶ When the Lead source was:
 - a. Google
 - b. Direct Traffic
 - c. Organic Search
 - d. Welingak Website
- ▶ When their current occupation is as a Working Professionals.

Keeping these in mind the X Education can flourish as they have a very high chance to get all the potential buyers to change their mind and buy their courses.