X Education-Lead Scoring Case Study

Deetection of Hot leads to improve conversion rates for X Education

Table of contents

01

Statement

- X Education sells online course to industry professionals.
- Interested professional land on the website, browse courses, fill forms, or watch videos.
- Leads are acquired when people forms
- Sales team contacts the leads through calls and emails
- Typical lead conversion rate is around 30%

Purpose statement

Conversion rate?

Conversion rate is low around 30%.

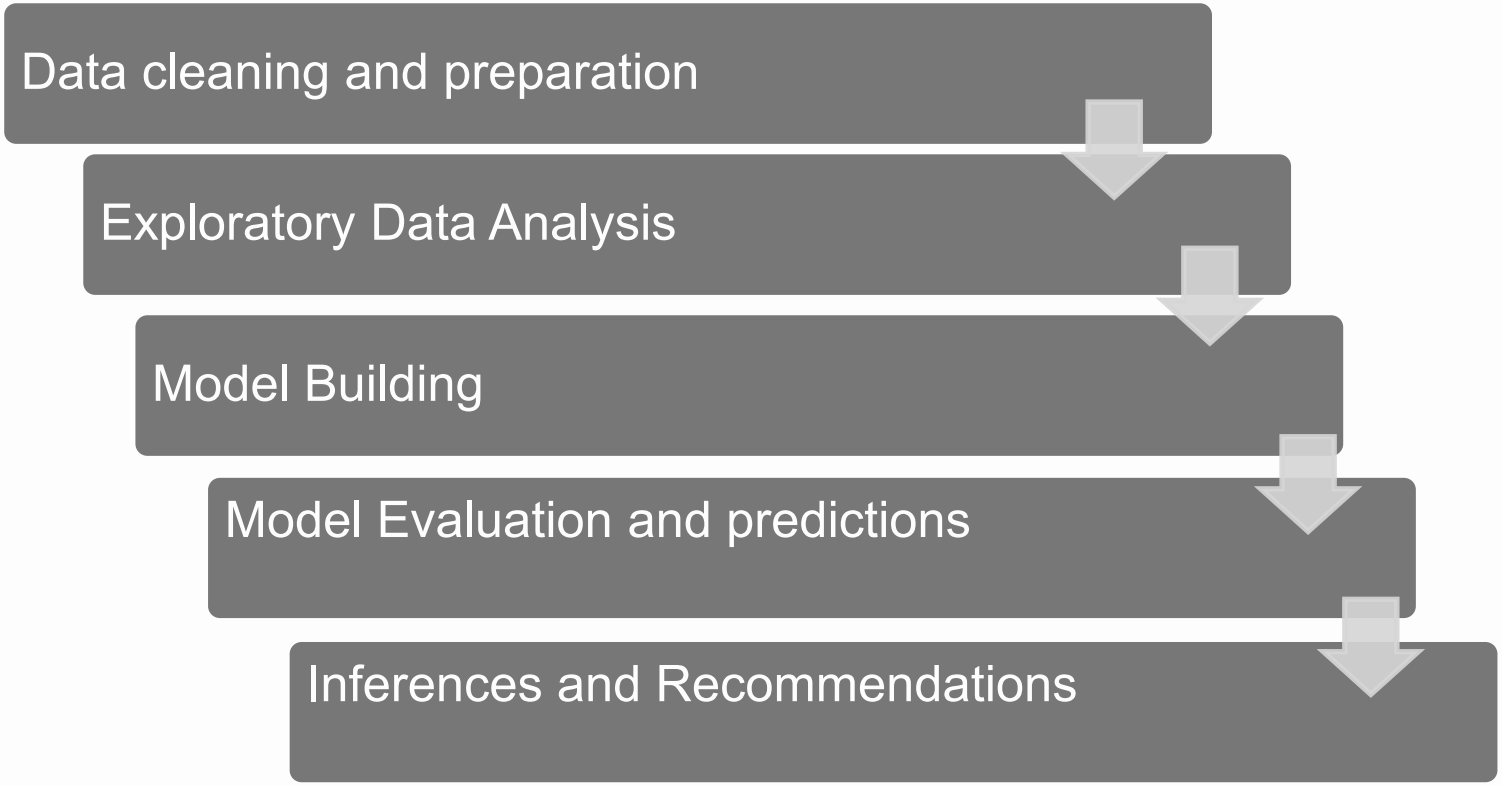
Hot leads?

Need to identify potential leads to focus the sales efforts.

What is required?

Aim for a target lead conversion rate of 80%

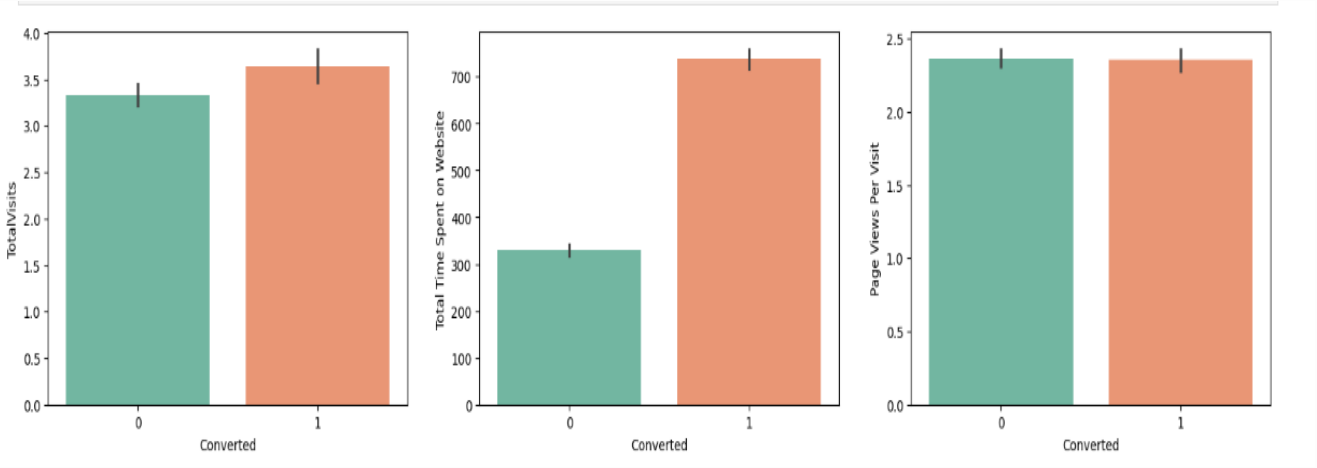
Approch for the case study



DATA CLEANING AND PREPARATION

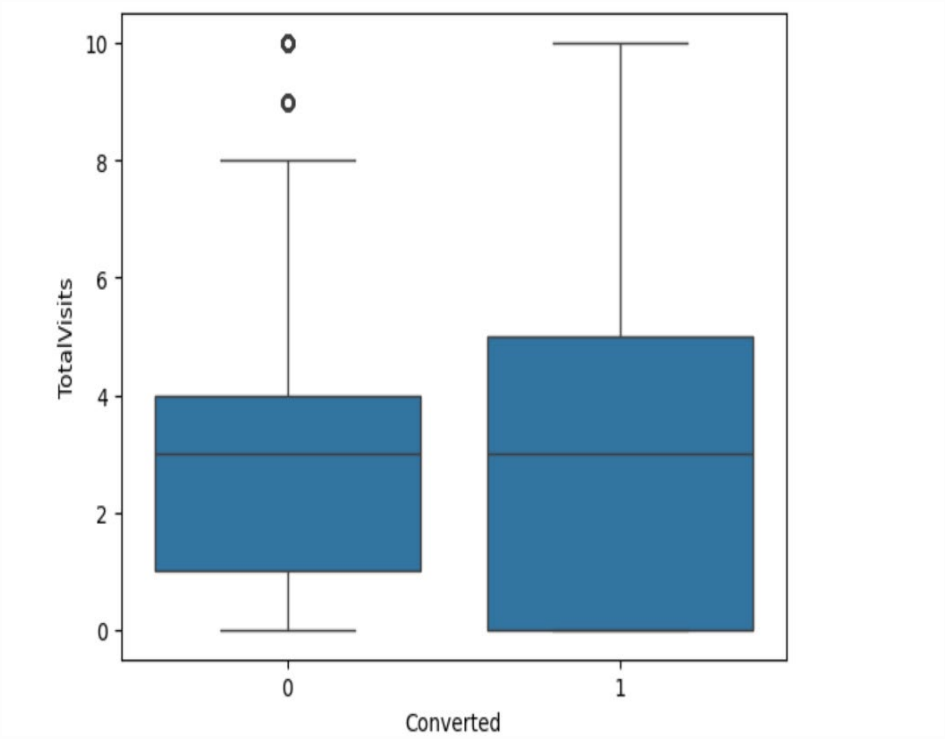
- “SELECT” represents the null values as the customer did not choose any option for that variable.
- Dropping the columns with over 30% null values.
- Replacing the missing values with their mode in categorical columns.
- Imputing numerical data with the mean.
- Standardizing data in columns by checking the case(e.g. Google and google)

Exploratory Data Analysis



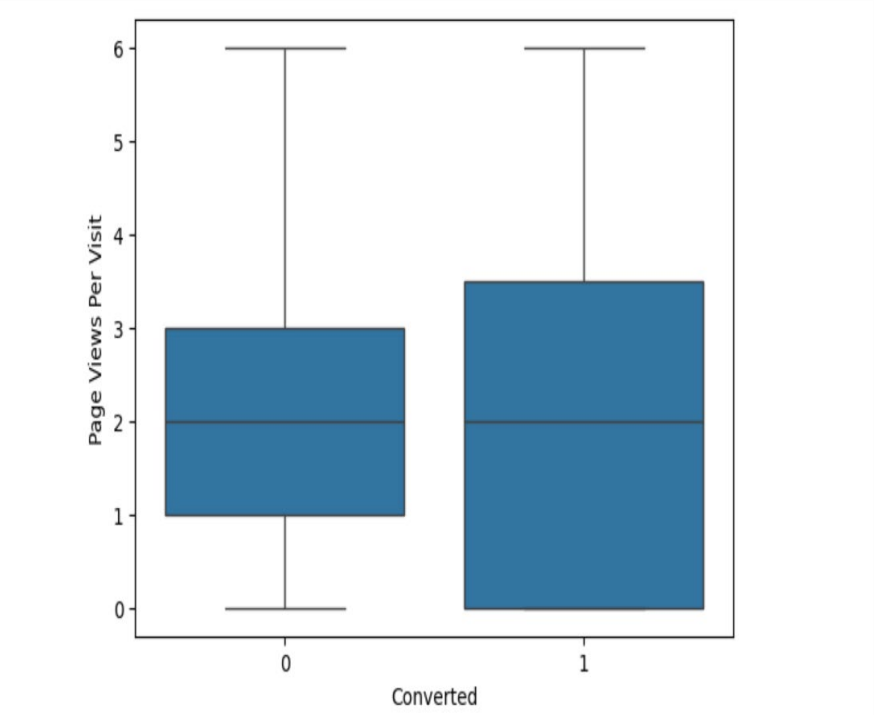
The conversion rate is high for Total Visits and Total Time Spent on Website.

Analysis of the results



Since the median is the same for both converted and non-converted leads, no definitive conclusions can be drawn from the variable TotalVisits.

Analysis of the results



Since the median is the same for both converted and non-converted leads, no definitive conclusions can be drawn from the variable Page views PerVisit.

Data Preparation

After creating dummy variables and mapping them we move on to splitting the data into train test set.

- Splitting train test set in 7:3
- MinMaxScaler is used to scale the features.
- ‘Converted’ column is our targeted column
- Highly correlated variables are dropped as they have high VIF.

Mapping the binary categories

Mapping binary categories from ‘Yes’ to 1 and ‘No’ to 0. Same with True and False

Dummy variables

Creating dummy variables for categorical features like Lead Origin, Lead Source, Last Activity, What is your current occupation and Last Notable Activity

Conclusions

Mars

Despite being red, Mars is actually a cold place. It’s full of iron oxide dust

Venus

Venus has a very beautiful name and is the second planet from the Sun

Mercury

Mercury is the closest planet to the Sun and the smallest one in the Solar System

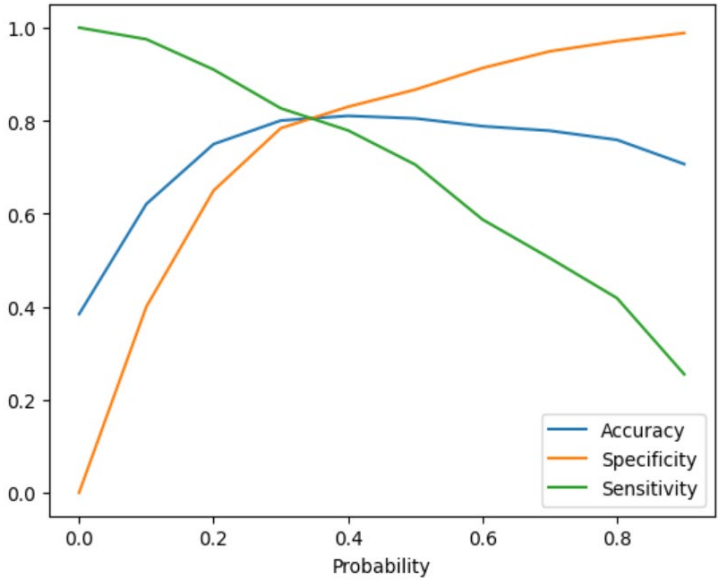
Neptune

Neptune is a ig planet. It is the fourth-largest planet by diameter in the Solar System

MODEL BUILDING

- The data set has a lots of dimension and large number of features making it important to perform RFE.
- RFE was used to select the important features.
- After using the RFE we reduced features from 48 to 15.
- Later, we reduced the features manually by dropping the high p-value variables and VIF>5.

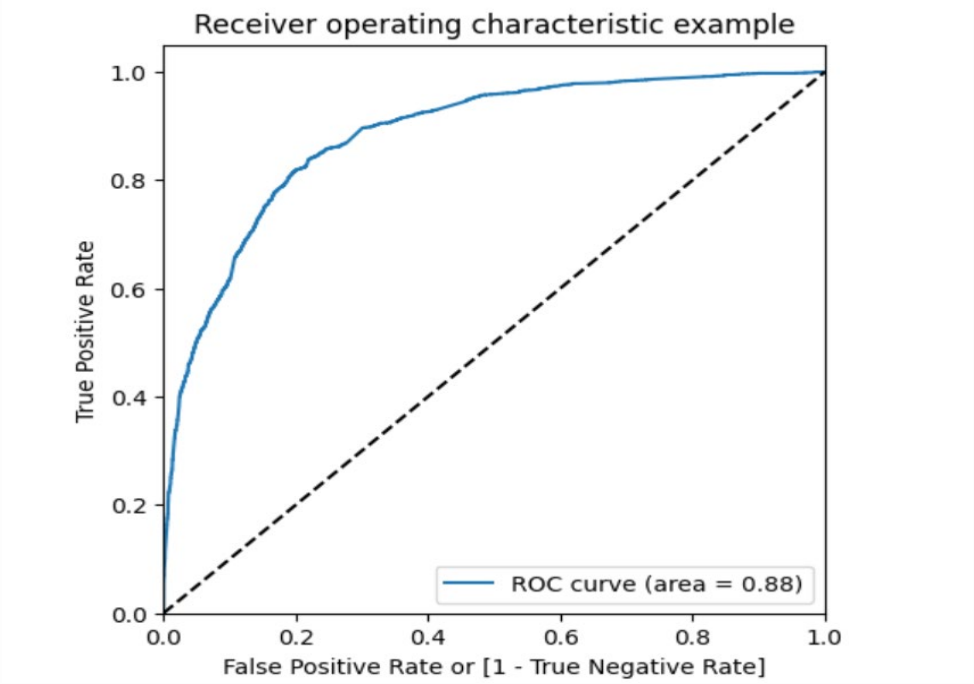
MODEL EVALUATION



	Probability	Accuracy	Specificity	Sensitivity
0.0	0.0	0.384045	0.000000	1.000000
0.1	0.1	0.620748	0.399849	0.975040
0.2	0.2	0.749536	0.649598	0.909823
0.3	0.3	0.800247	0.783886	0.826490
0.4	0.4	0.810297	0.829819	0.778986
0.5	0.5	0.804731	0.866717	0.705314
0.6	0.6	0.788033	0.913153	0.587359
0.7	0.7	0.778293	0.949046	0.504428
0.8	0.8	0.758658	0.970884	0.418277
0.9	0.9	0.706555	0.988454	0.254428

0.35 as cutoff was selected after checking evaluation metric and plot.

MODEL EVALUATION



ROC AUC score is 0.88 which represents that model can produce relative scores to discriminate between positives and negative instances across all classification thresholds.

.

Inferences and recommendations

To increase lead conversion rate:

- Focus on positive coefficient features for targeted marketing.
- Working professionals who visits website are highly likely to be converted.
- Optimize the digital marketing to increase the traffic to the website.

To indetify areas of improvement:

- Analyze negative coefficients like Last Notable Activity

Thanks!

Do you have any questions?

Made by:
Harshita Kapoor
Harsimran Singh
Himabindu Thondamanti