

Lead Score Assignment

IIIT B - Upgrad

JAGADEESH SHIVAPRASAD

PRITHUL KESHRI

ASHWIN KUMAR



Problem Statement:-

- X Education sells online course to working professional.
- Currently conversion rate of leads is about 30%.
- They want to build Logistic regression model to predict and understand the conversion of potential leads known as Hot-leads.
- Implement strategy to communicate with leads for better conversion

Objective:

- To build a model to identify potential leads for conversion with minimal efforts.
- Predict promising leads.

Methodology:-

1. Data Cleaning and Manipulation

- Check the duplicates – There are no duplicates in current data.
- Check for null values – Dropped columns where percent of null values are more than 40%.
- Impute Data with select with Null values
- Most of the Null values are replaced with values which has more counts in columns.
- In some cases where value counts are less are grouped together based on data.
- Columns which has Data imbalance more than 97-98% are dropped. (“Country”, “Do not Call” etc.)

2. EDA

- Univariate Analysis - Plot count plots, box plot to check outliers. No outliers were dropped.
- Bivariate Analysis – Correlation Matrix, comparison

3. Creating Dummy variables And Scaling Numerical variables

4. Logistic Regression using RFE (Recursive Feature Elimination)

5. Validation of Model using ROC, Accuracy, Sensitivity, Specificity

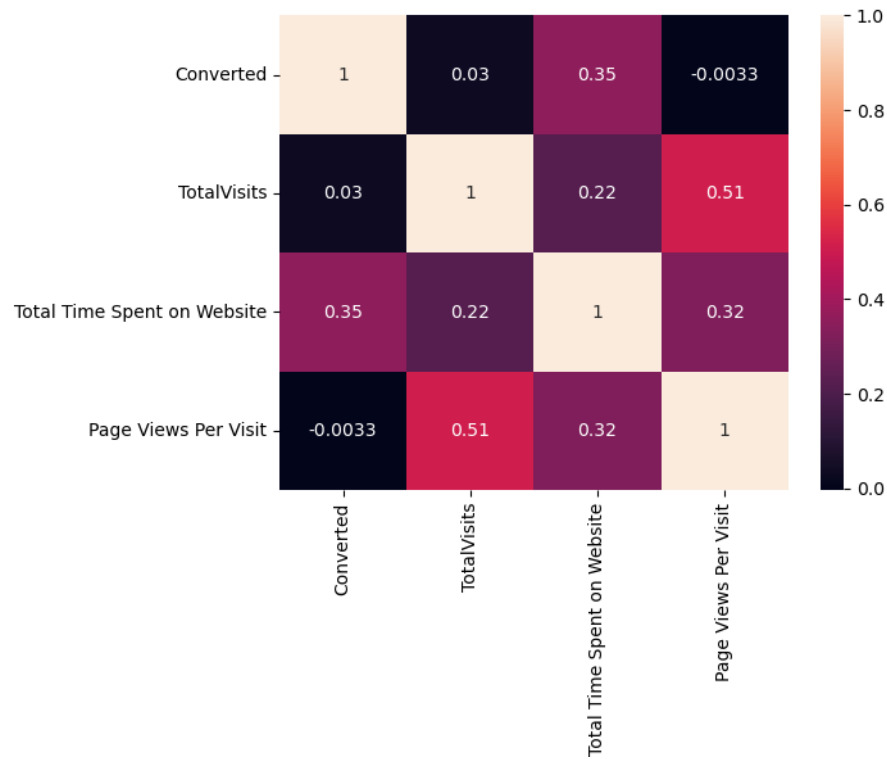
6. Conclusion and Recommendations

Data Cleaning and Manipulation:-

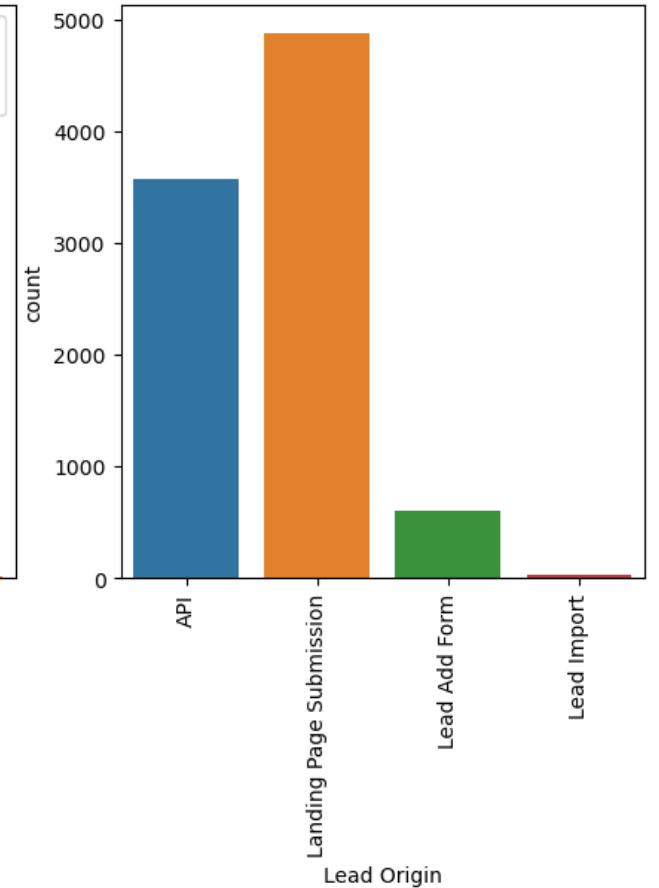
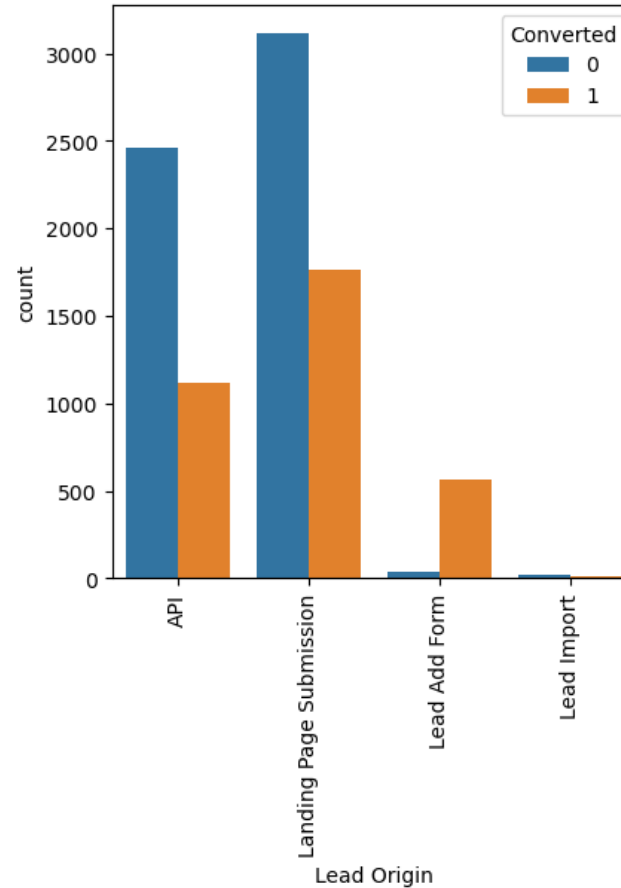
- There are Rows = 9240 and columns = 37.
- Columns with more than 40% null values after imputing 'select' with Null values are dropped.
- Columns with Data imbalance are dropped.
- Columns name "Prospect Id" and "Lead Number" are dropped as it has unique names.
- Thus, finally we have data with columns number reduced to 13.
- Dummies are created using get_dummies function.
- Scaling is done for numerical data.

Finally, we will have Rows for Analysis = 9103 ; Columns for Analysis = 51

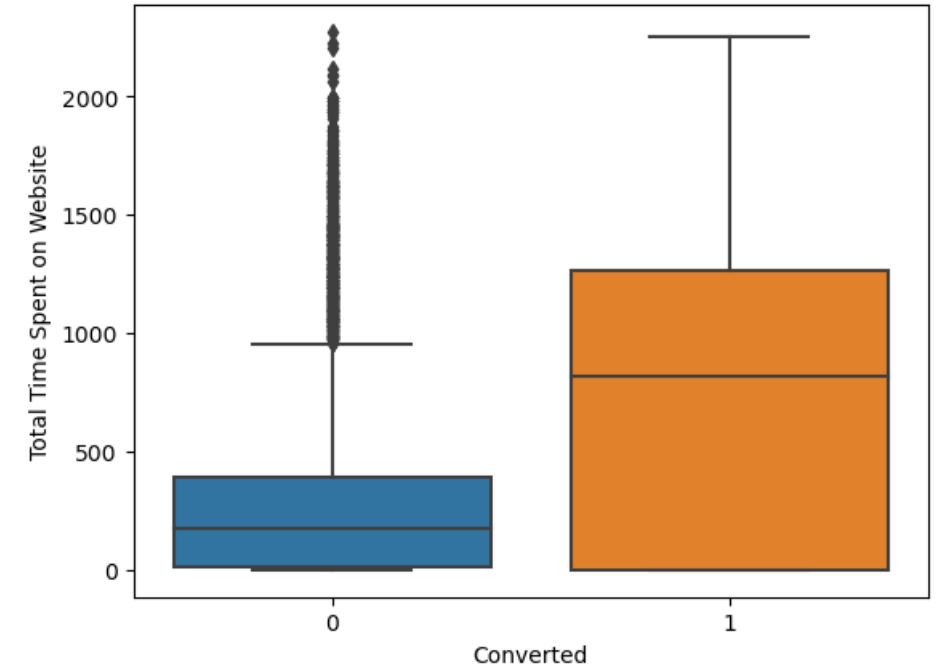
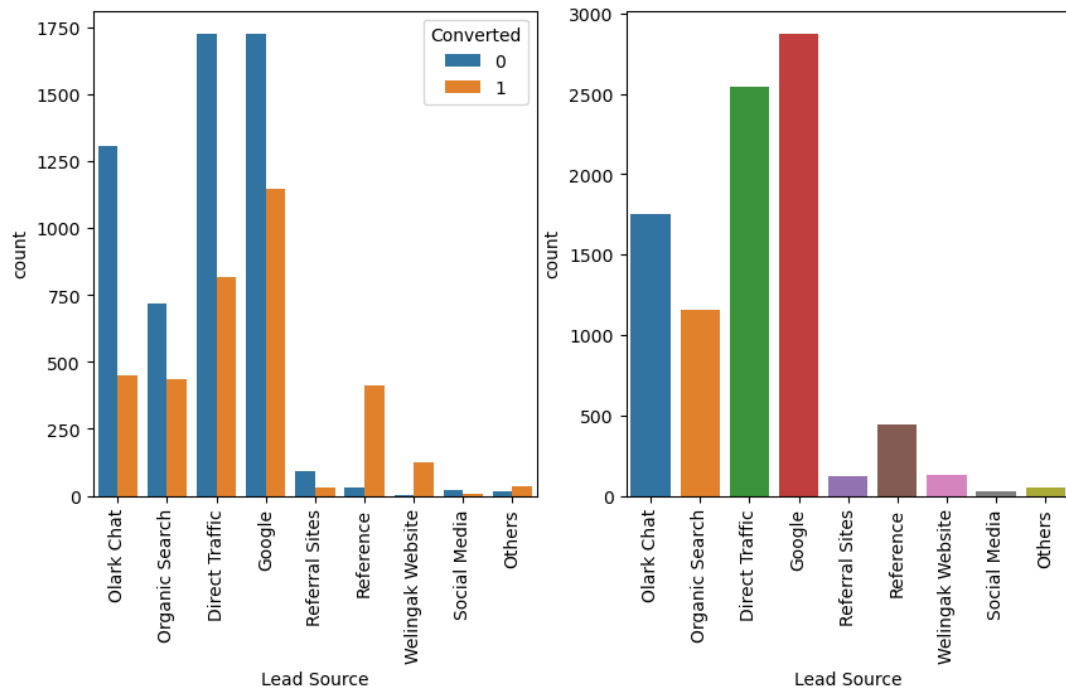
EDA:-



We find good correlation with Total visits on websites and number of page per visits.

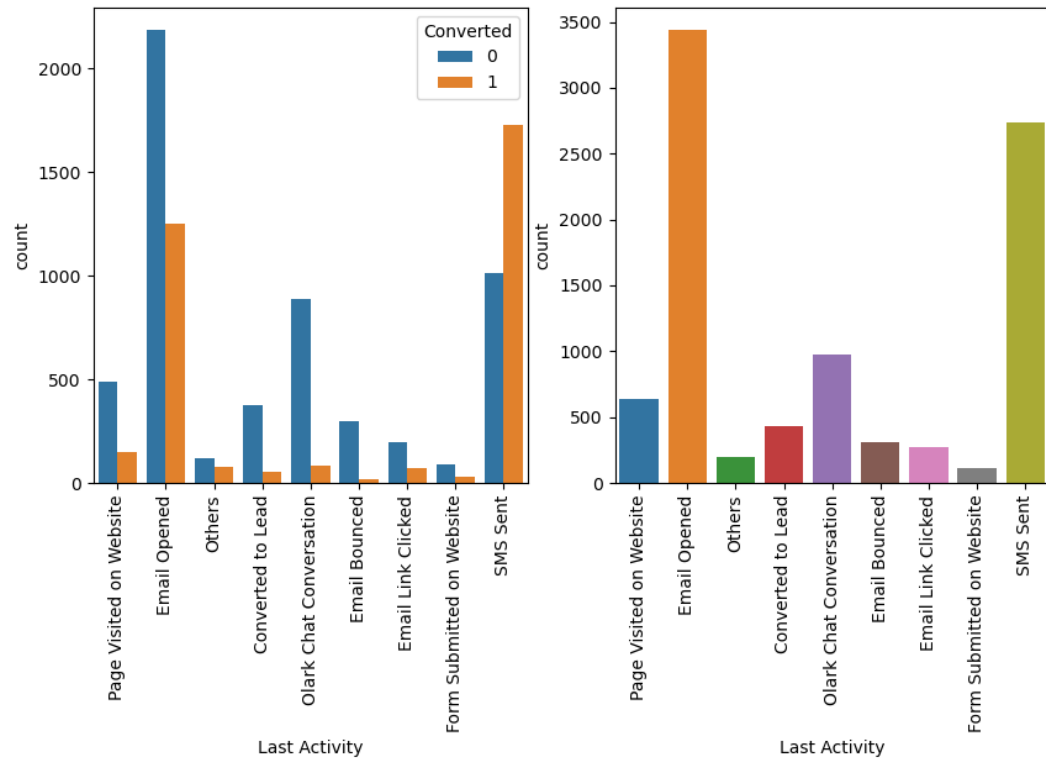


- API and Landing Page has more number of Leads and conversion rate. We need to work on and improve these origin.
- Lead Add Form has high conversion rate but count is less.
- We need to focus more on improving lead conversion of API and Landing Page Submission origin and generate more leads from Lead Add Form to improve.

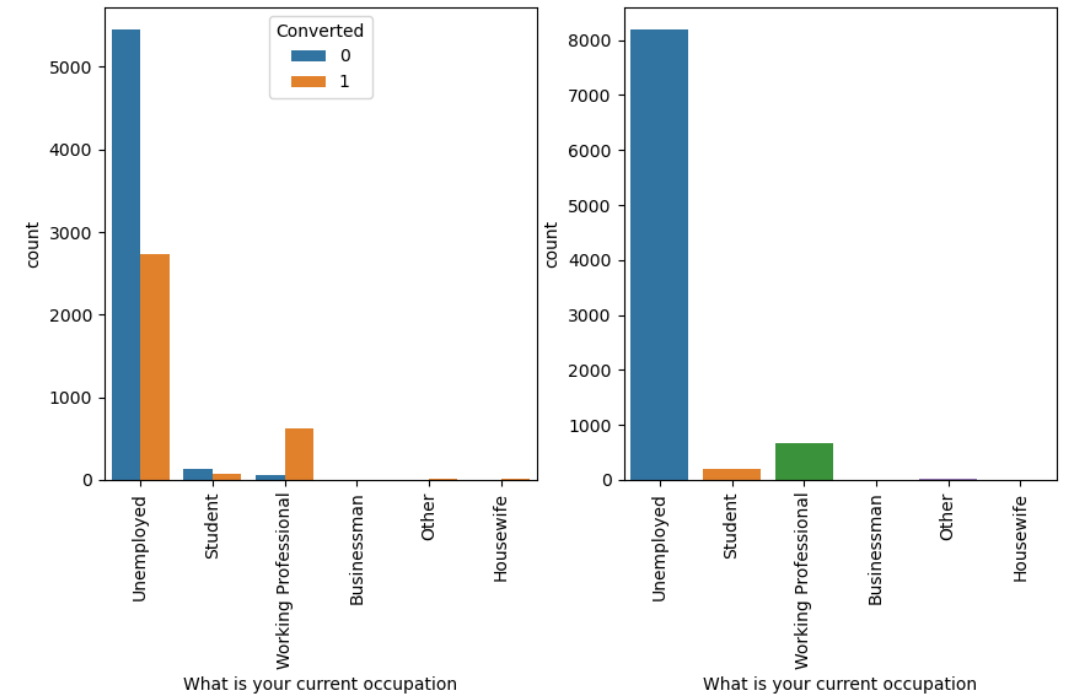


- Google and Direct Traffic has max number of leads almost 30%.
- Conversion rate of Weilingak website is high.

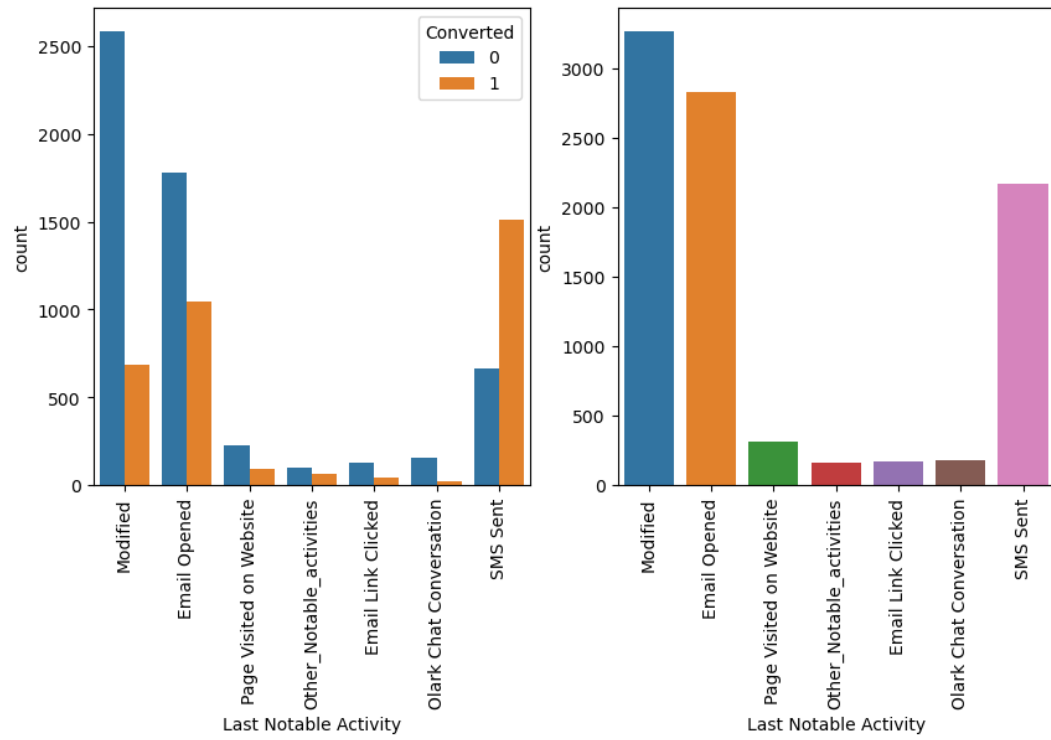
- Leads spending more time on the website are more likely to be converted.



- Most of the lead have their Email opened as their last activity.
- Conversion rate for leads with last activity as SMS Sent is almost 60%.



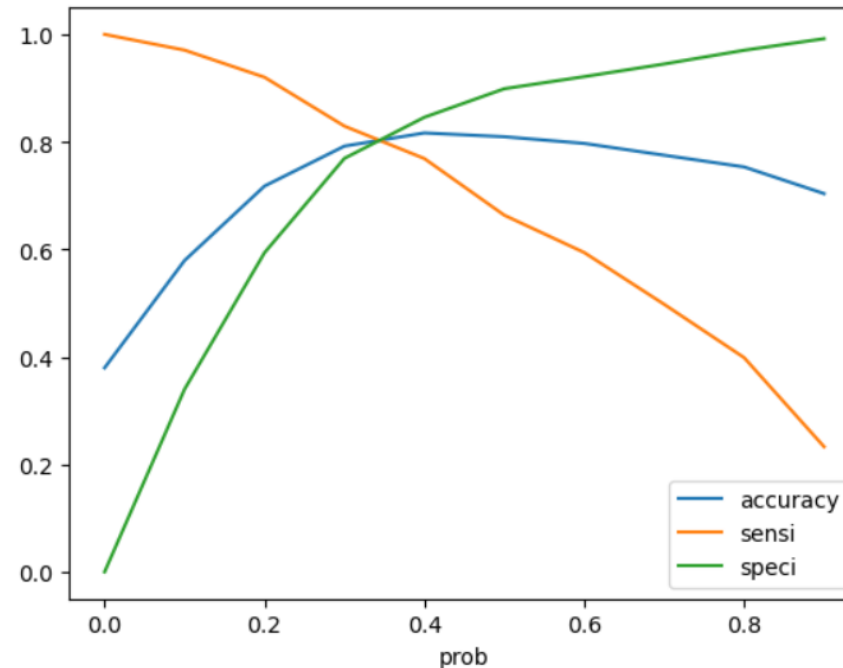
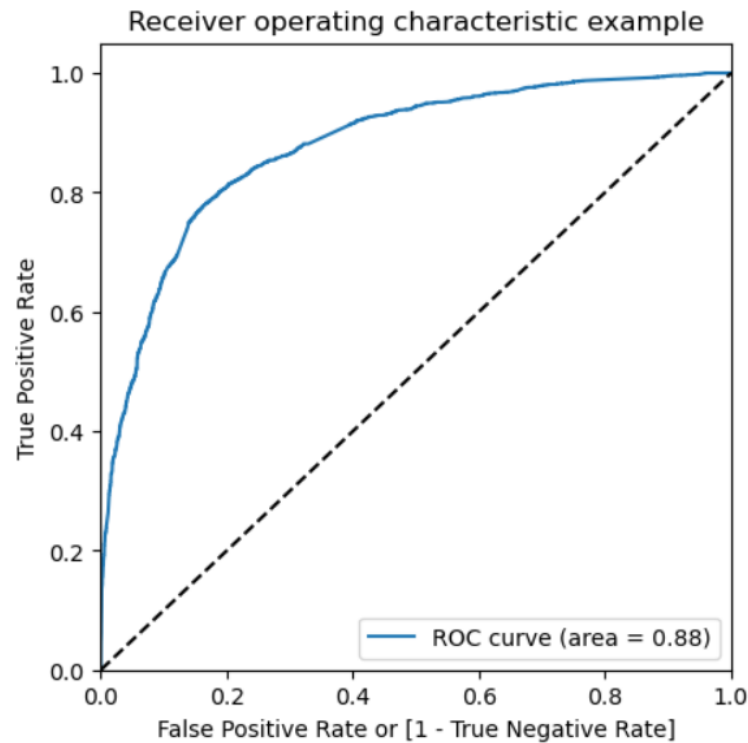
- Working professional has high conversion rate. Most of them are unemployed.



People to whom SMS is sent has more conversion Rate.

Model Building:-

- Split Data into Train and Test Sets in 70:30 ratio
- We used RFE for feature selection with 15 variables as output.
- Building Model by removing the variable whose p-value is greater than 0.05 and vif-value is greater than 5
- Predictions on test data set
- Overall accuracy is 80% on both test and train data. Thus, this model is very reliable.



From the second graph it is visible that the optimal cut off is at 0.34.

Conclusion:-

The company can look for leads for better conversion

- Lead Origin with Add form
- Lead Source
- Occupation like working professional
- Time spent on website.

The company should make calls to:

- working professional
- whose last activity is shown as SMS sent
- Origin is Add Form
- who spent more time on websites
- who use welingak website.

The Company should not make calls to save time and efforts:-

- Who do not have email
- Who did not mention any specialization – maybe these people are illiterate or students.
- Lead origin is Landing page submission

The company, can ask its team to created automated emails and SMS. This will help customers to read and go through course details. It is known that people who read emails and revert are more likely to get converted or buy the course