

LEAD SCORE CASE STUDY

- Submission By:
 - Sourabh Chouhan
 - Supritha Chakraborty
 - Saurabh Behl



PROBLEM STATEMENT

- X Education sells online courses to industry professionals.
- X Education is able to source lot of leads, but its lead conversion rate is very poor which is only around 30% and is concern for the company.
- To make the conversion process more efficient, the company wishes to identify the most potential leads, also known as 'Hot Leads'.
- With successful identification of the leads, the objective is to result in high conversion rate. Company CEO's target being 80%.
- Such identification of the deals, would facilitate the sales team on communicating with the potential leads rather than making calls to everyone, including non-prospective leads.



PROBLEM STATEMENT

- X education, is targeting to identify most promising leads.
- To accomplish the above, they intend to build a Model which identifies the hot leads.
- CEO is targeting to achieve 80% lead conversion rate.

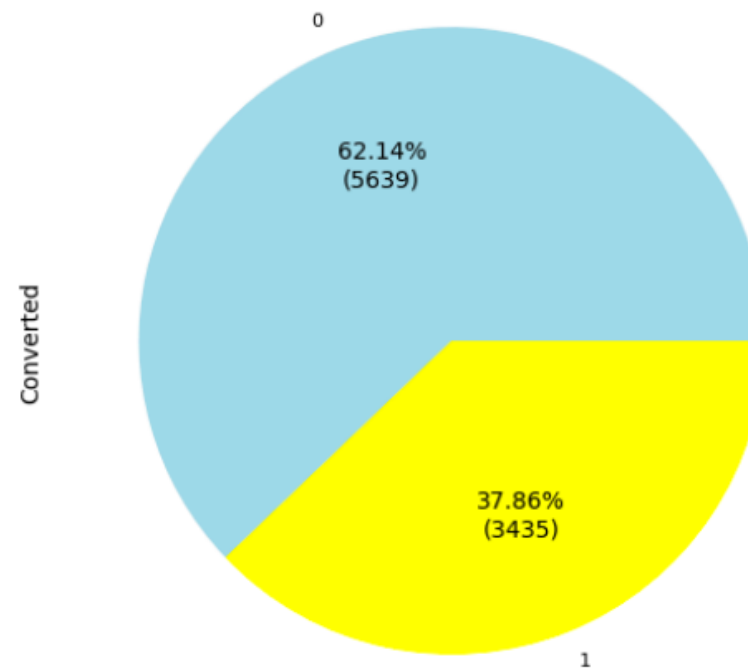


SOLUTION METHODOLOGY

1. **Data Cleaning:** Loading, Understanding and Cleansing of Data Set.
2. **EDA:** Check Imbalance, univariate and bivariate analysis.
3. **Data Preparation:** Dummy variables, test-train split, feature scaling.
4. **Model Building:** RFE for the top 15 feature, manual feature reduction and finalizing the model.
5. **Model Evaluation:** Confusion matrix, cutoff selection, assigning lead score.
6. **Predictions on the Test Data:** Compare train vs test metrics, assign lead score and get top features.
7. **Recommendations:** Suggest top 3 features to focus for higher conversion areas.



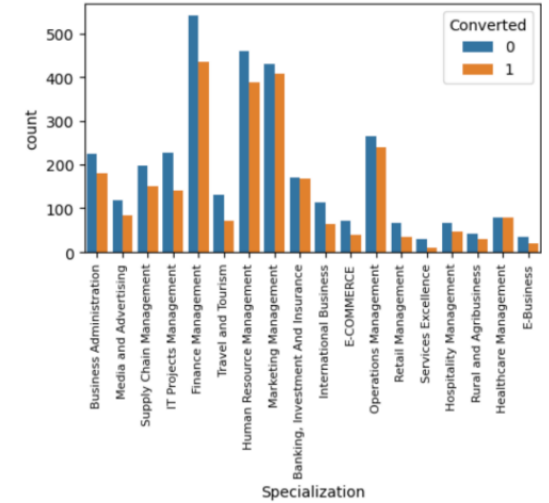
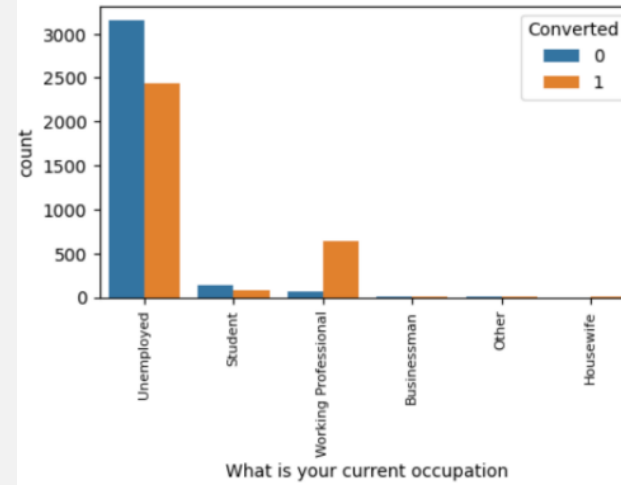
Converted Vs Non-Converted - distribution



- Lead conversion rate is 38%.
- While 62% folks did not convert to leads

EDA

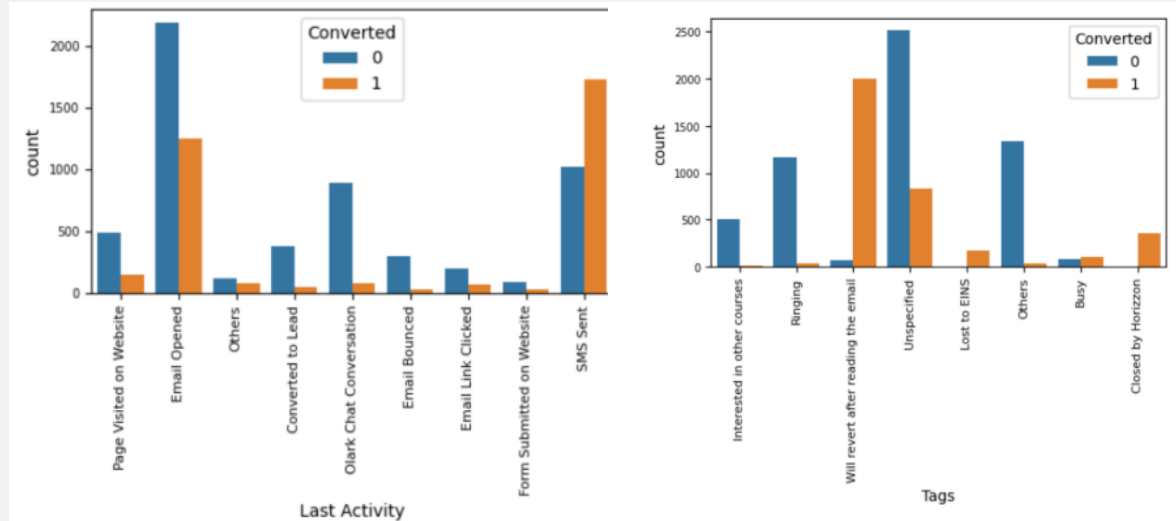
..CONTINUED..



- Unemployed and working professionals have a high conversion.
- Management type “Specializations” have maximum counts and lead conversions. Hence, these should be the focus areas.

EDA

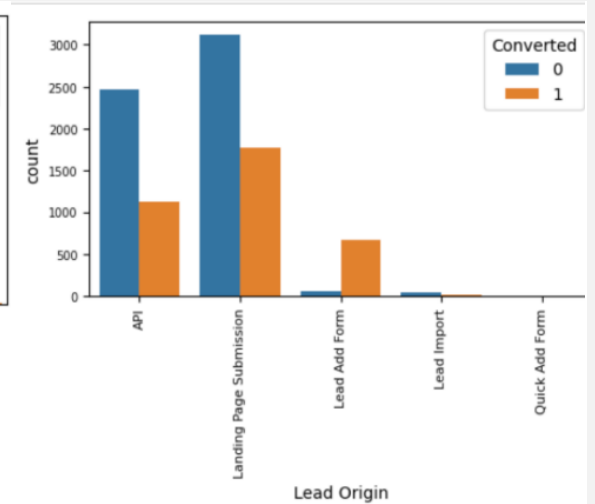
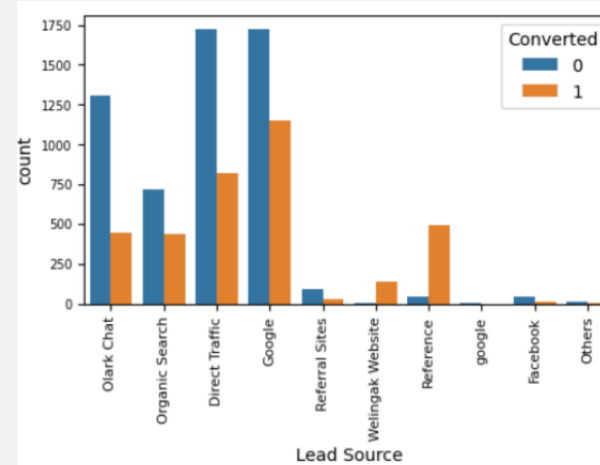
..CONTINUED..



- 'Email Opened', 'SMS Sent' have the highest lead counts and shows Highest converted leads.
- 'Will revert after reading the email' have almost > 90% conversion rate.

EDA

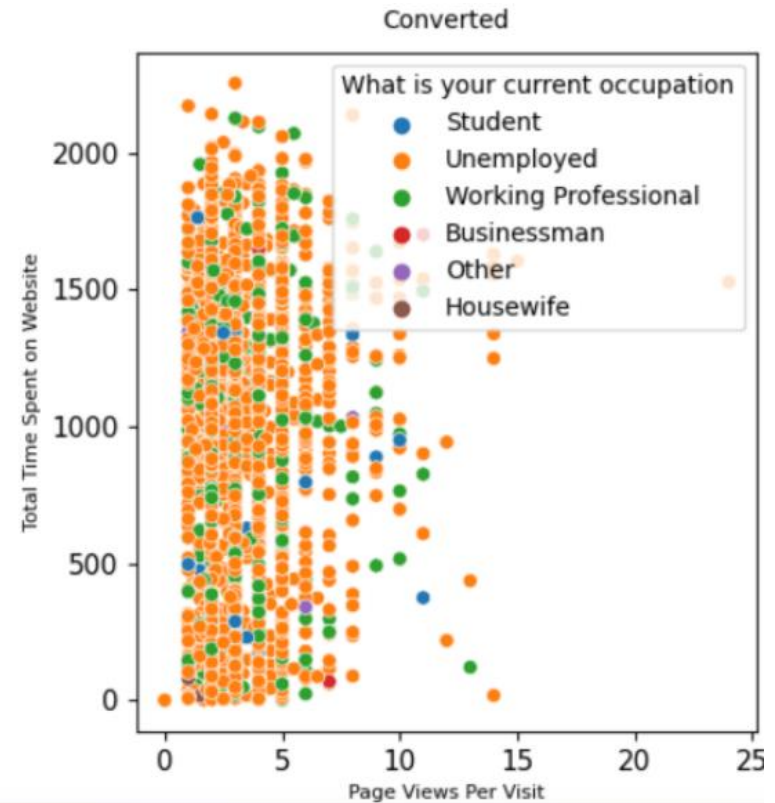
..CONTINUED..



- Lead Origin : Lead Add Form has high conversion.
- Reference & Welingak Website category has highest lead conversion

EDA

..CONTINUED..

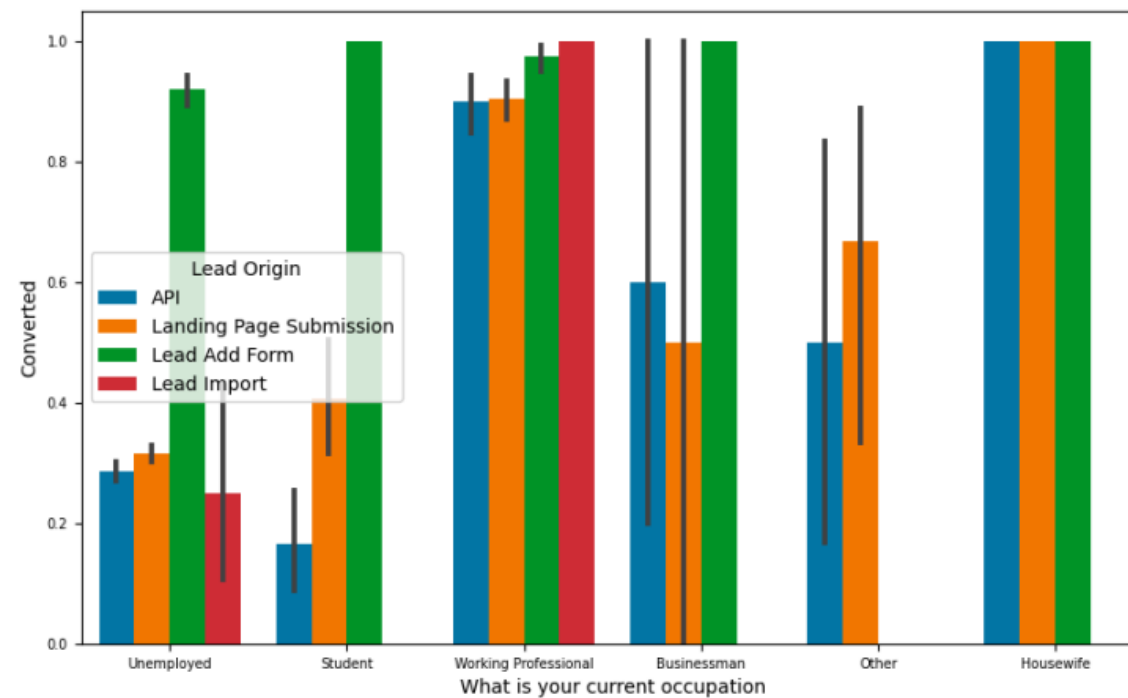


- Unemployed and Working professionals spend most time on the website and have a high lead conversion rate.
- High conversion rates have fewer page views per visit

EDA

..CONTINUED..

Current occupation VS Lead Origin for Converted



- Housewife and Working Professional show highest and similar conversion across 'Lead Origins'.
- Lead Add Form has highest conversion amongst all occupations

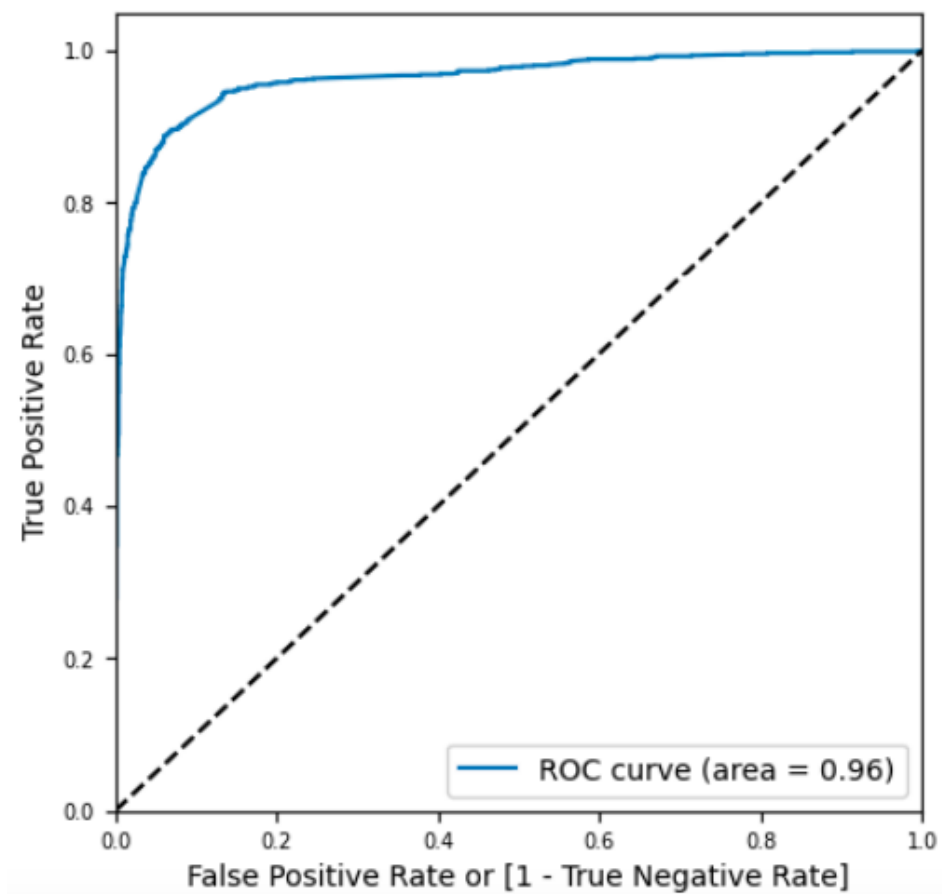


MODEL BUILDING

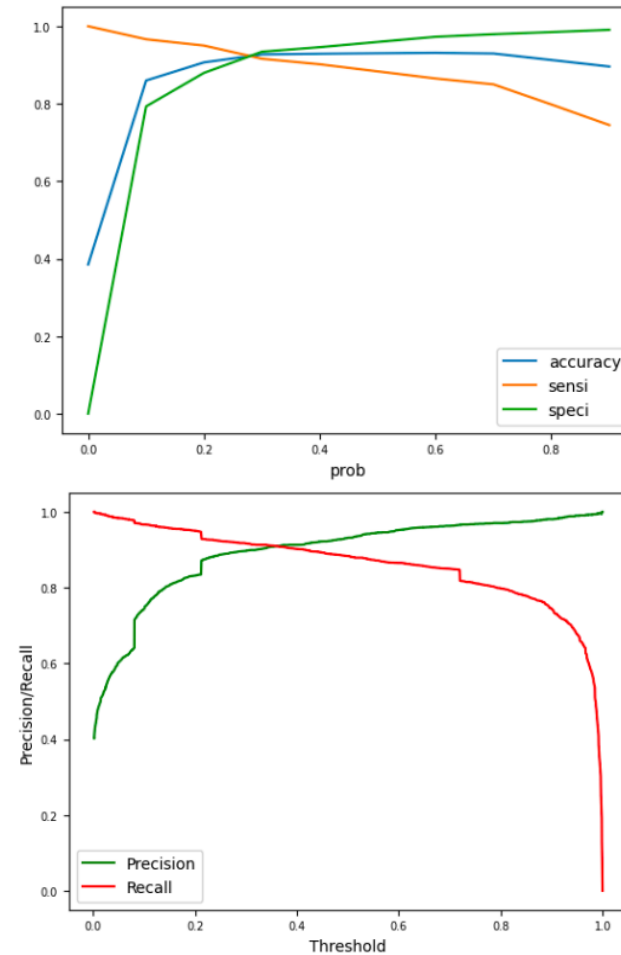
- ✓ Splitting the Data into Training and Testing Sets.
- ✓ The first basic step for regression is performing a train-test split, we have chosen 70:30 ratio.
- ✓ Use RFE for Feature Selection.
- ✓ Running RFE with 15 variables as output.
- ✓ Building Model by removing the variable whose p-value is greater than 0.05 and VIF value is greater than 5.
- ✓ Predictions on test data set.

ROC CURVE

Receiver operating characteristic example



ROC CURVE ..CONTINUED..



- Optimum Probability is 0.3.
 - Hence any Prospect Lead with Conversion Probability higher than 30% is a hot Lead.
- Precision and recall trade-off is approx. the same.



MODEL SUMMARY

- Metrix Comparison:
 - Train Data Set:
 - Accuracy : 93%.
 - Sensitivity: 92%.
 - Specificity: 93%.
 - Test Data Set:
 - Accuracy : 92%.
 - Sensitivity: 90%.
 - Specificity: 93%.

Summary:

- The model sensitivity of 92% in the train data set and 90% in the test set, using a cut-off value of 0.3.
- Sensitivity indicates how many leads the model could identify correctly.
- The CEO of X Education had set a target sensitivity of around 80%.
- The model also achieved an accuracy of 93%, which beats the threshold.



CONCLUSION

It was found that the variables that mattered the most in the potential buyers are:

- The Total Time Spent on Website.
- Total number of Visits.
- When The Lead source was:
 - *Olark chat*
 - *Wellingak Website*
- When the Last Activity was :
 - *SMS*
 - *Olark Chat Conversation*
- When the lead origin is Lead add Form.
- When the current occupation was:
 - Working Professionals
 - Student
 - Unemployed
 - Other
- Focusing on the above areas, X education can improve their conversion rate and achieve the threshold targets.