Lead Scoring – Case Study Using Logistic Regression

Contributors:

- 1. Abhay Ranjan
- 2. Rajalakshmi Harhiharan
- 3. Rahul Garg

Contents:

- **■** Problem Statement
- **■** Problem Approach
- **EDA**
- **►** Model Evaluation
- Observations
- **■** Conclusion

Problem Statement:

- An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses.
- The company markets its courses on several websites and search engines 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 referrals. 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%.
- Now, although X Education gets a lot of leads, its lead conversion rate is very poor. For example, 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 identify this set 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 Objective:

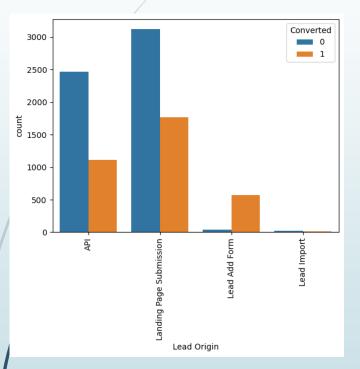
- X Education wants to select the most promising leads, i.e. the leads that are most likely to convert into paying customers.
- The company requires to build a model wherein one needs to assign a lead score to each of the leads such that the customers with a higher lead score have a higher conversion chance and the customers with a lower lead score have a lower conversion chance.
- The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.
- Build a logistic regression model to assign a lead score between 0 and 100 to each of the leads which can be used by the company to target potential leads. A higher score would mean that the lead is hot, i.e. is most likely to convert whereas a lower score would mean that the lead is cold and will mostly not get converted.

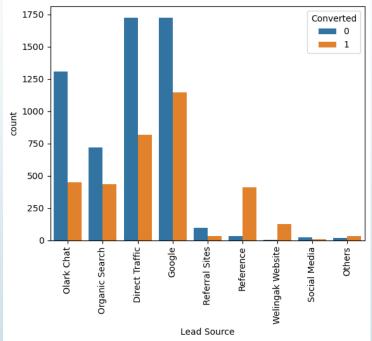
Problem Approach:

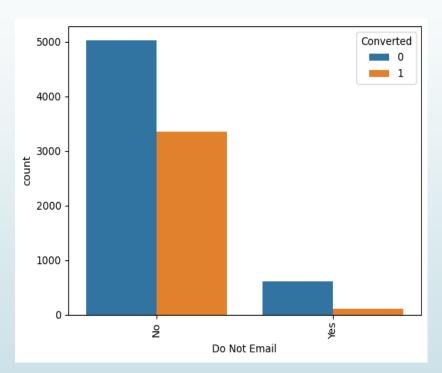
- Loading and inspecting the data
- Data preparation
- EDA
- Dummy variable creation
- Test-Train split
- Feature scaling
- **■** Logistic Regression Model Selction
- Model Building (RFE, VIF and p-values)
- Model Evaluation
- Making predictions on test set

EDA – Univariate:

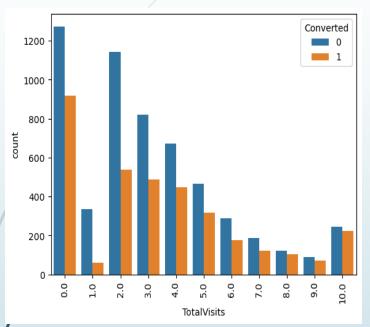
Bar Graph showing converted leads counts for each class label.

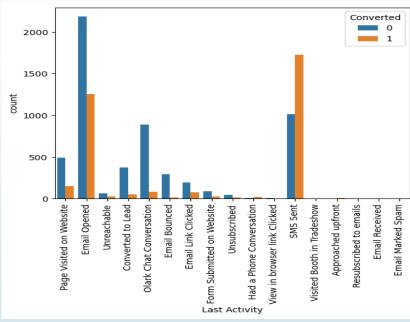


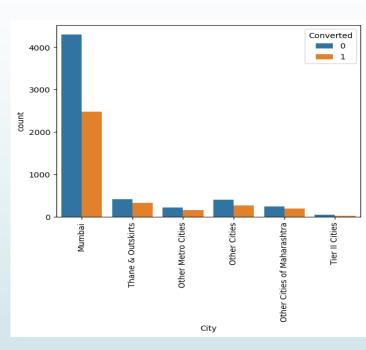




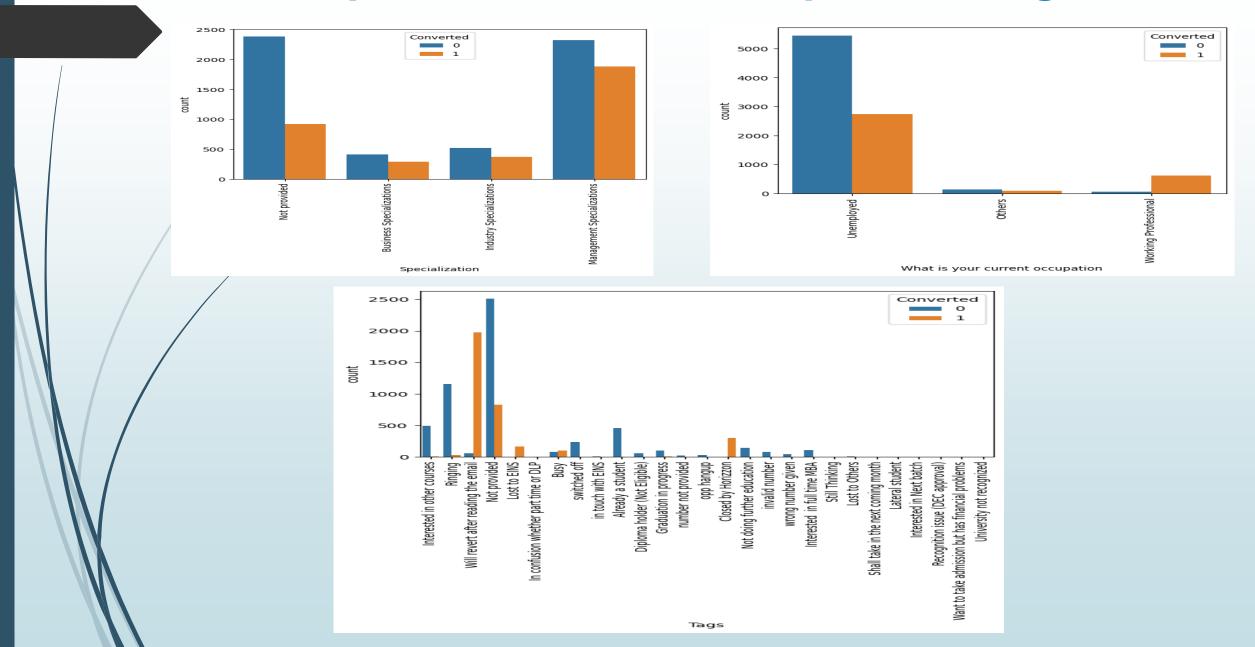
EDA - Last Activity, Total Visits and City:



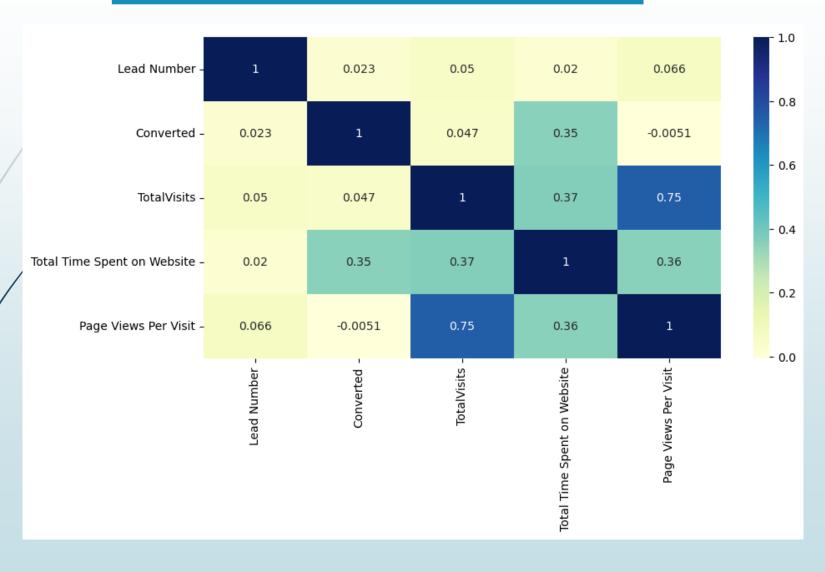




EDA – Specialization, Occupation, Tags

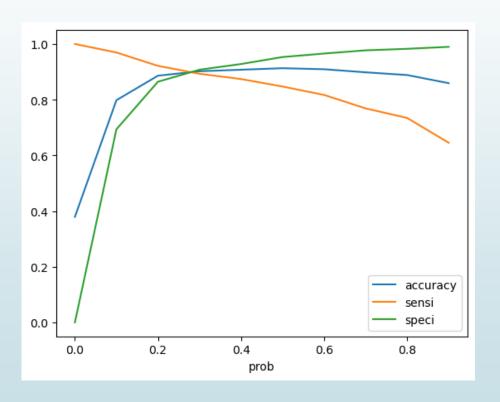


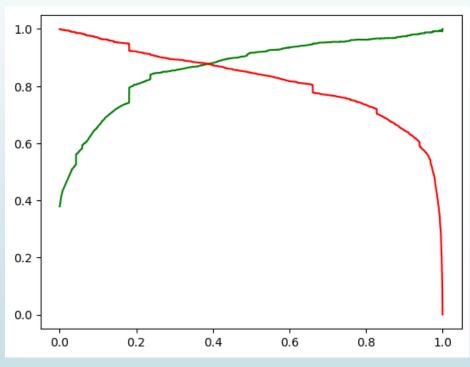
Bivariate - Correlations:



<u>Model Evaluation – ROC Curve:</u>

0.28 is the tradeoff between Precision and Recall Hence, we can consider Prospect Lead with Conversion Probability higher than 28 % to be a hot Lead





Observations:

Important Features:

- Lead Source_Olark Chat
- Lead Origin_Lead Add Form
- Lead Source_Welingak Website
- Total Time Spent on Website
- Lead Origin_Landing Page Submission
- Tags_Will revert after reading the email
- Last Notable Activity_SMS Sent
- What is your current occupation_Working Professional
- Last Notable Activity_Page Visited on Website
- Tags_Lost to EINS

Train Data:

Accuracy: 91%

Sensitivity: 87%

Specificity: 88%

Test Data:

Accuracy: 90%

Sensitivity: 87%

Specificity: 93%

Conclusion:

- Total visits as the number of visits to website increase the chances of conversion increases. First timers convert faster
- Last Activity- SMS sent and email opened means the conversion chances will increase.
- Specialization- Management Specializations have high conversion ratio
- Current occupation- percentage conversion of working professional is higher, however maximum leads who convert are unemployed.
- Leads who revert back after reading an email will definitely convert
- Most numbers of leads are from INDIA and maximum conversions come from city Mumbai
- Last activity- SMS and email as communication Source have high conversion ratio.
- The high number of total visits & Total time spent on platform may increasing the chances of lead to be converted.
- The leads are joined course for Better Career Prospects, most of having Specialization from Finance Management. Leads from HR, Finance & marketing management specializations are high probability to convert.
- Talking to last notable Activity, making improvement in customer engagement through email & calls will help to convert leads. As the leads which are opening email have high probability to convert, same as Sending SMS will also benefit.