LEAD SCORING CASE STUDY

MEENAKSHI.S, SHITAL PRASHANT JADHAV, RAHUL JALODHAR PATRA

PROBLEM STATEMENT

- X Education sells online courses to industry professionals.
- 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.

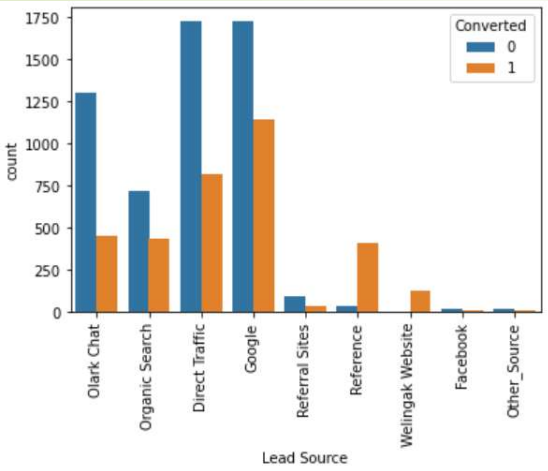
BUSINESS OBJECTIVES

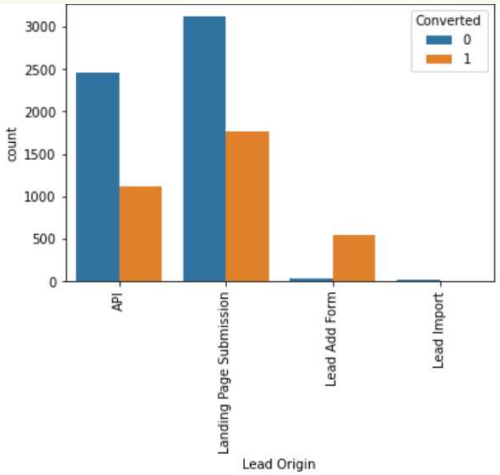
- Create a Logistic Regression model to predict the Lead Conversion probabilities.
- Decide on a probability threshold value above which a lead will be predicted as converted, where as not converted if it is below it.
- Assign a lead score value between 0 and 100 to each of the leads which can be used by the company to target potential leads.

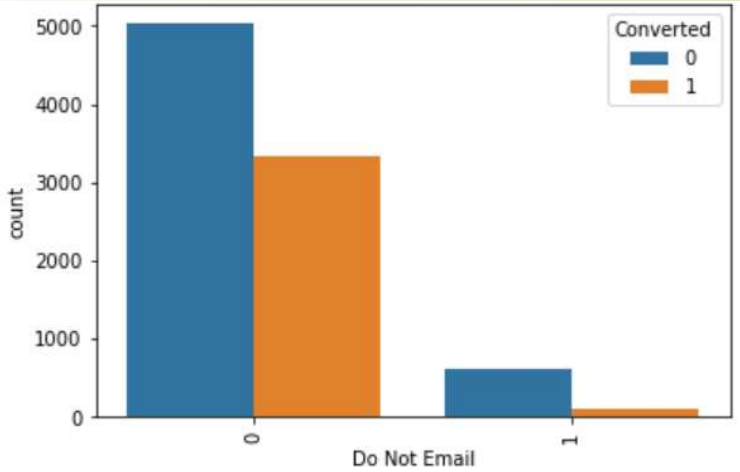
STEPS TO REACH OBJECTIVES

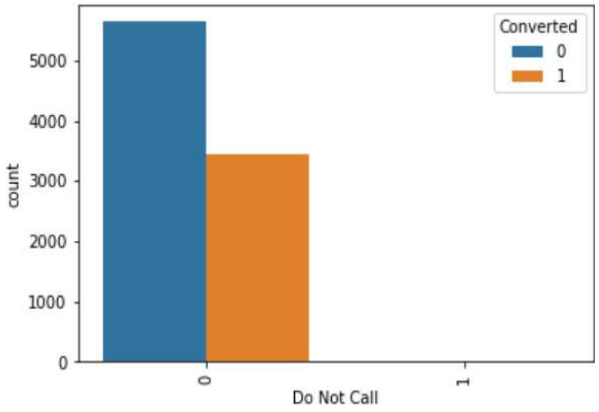
- Data Cleaning and preparation
  - 1. Check for Duplicates
  - 2. To convert the "Select" into the NaN
  - 3. Dropping columns with more than 70% null values
  - 4. Handling of Missing Value
  - 5. Categorical variables encoding
- EDA
- Feature scaling & Dummy variables creation
- Model Building & Evaluation
- Conclusion & Recommendations

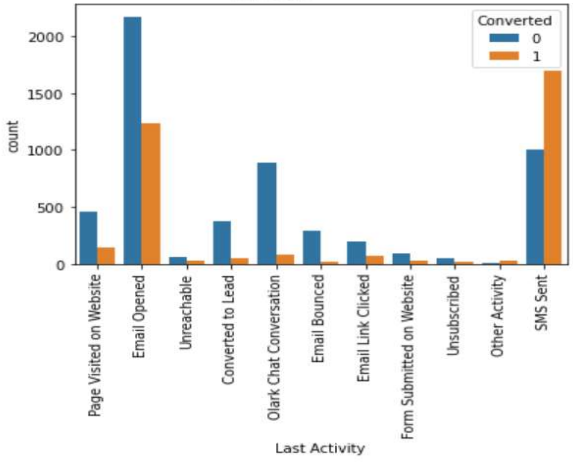
EDA

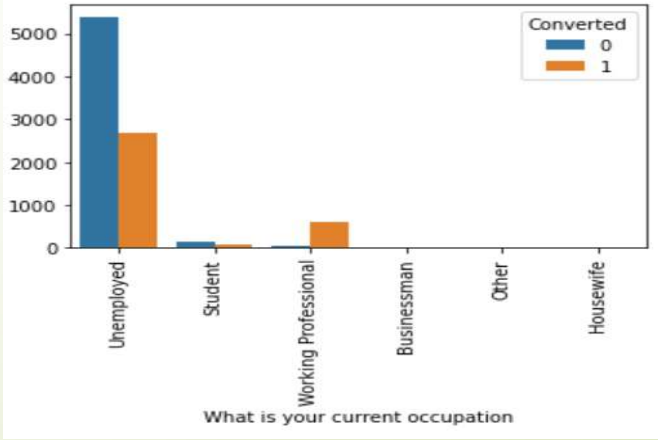

Converted
0
1
count
Olark Chat
Organic Search
Direct Traffic
Google
Referral Sites
Reference
Welingak Website
Facebook
Other_Source
Lead Source

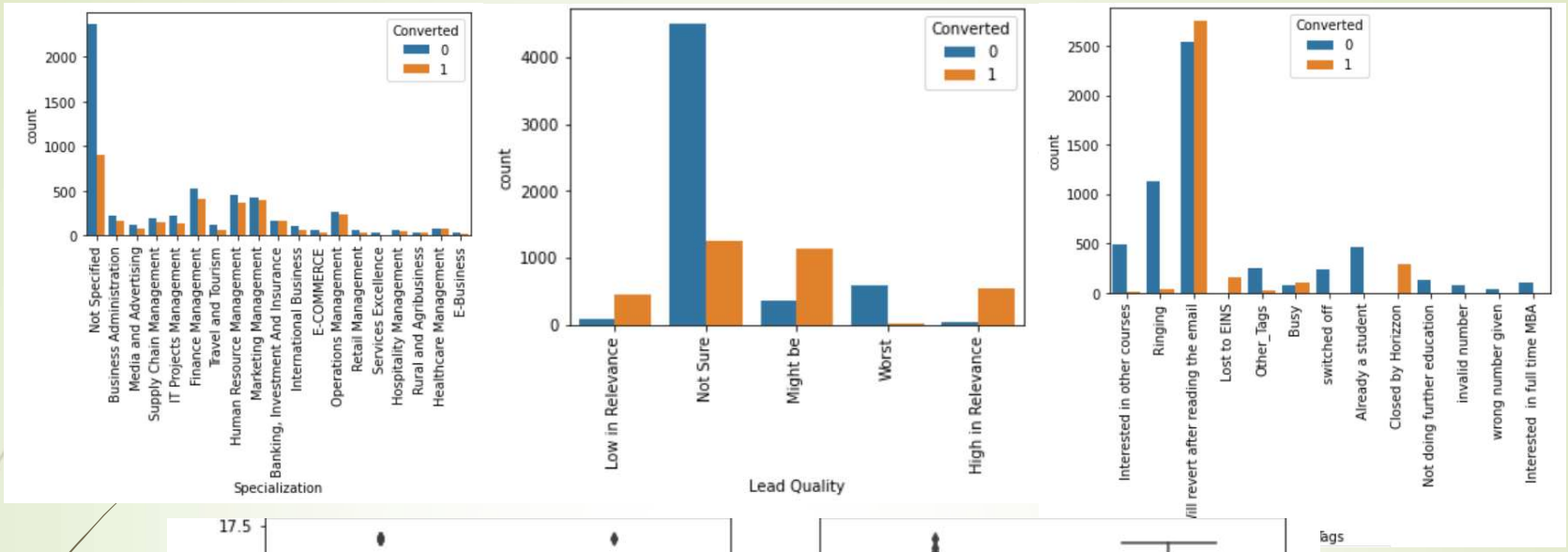

Converted
0
1
count
API
Landing Page Submission
Lead Add Form
Lead Import
Lead Origin

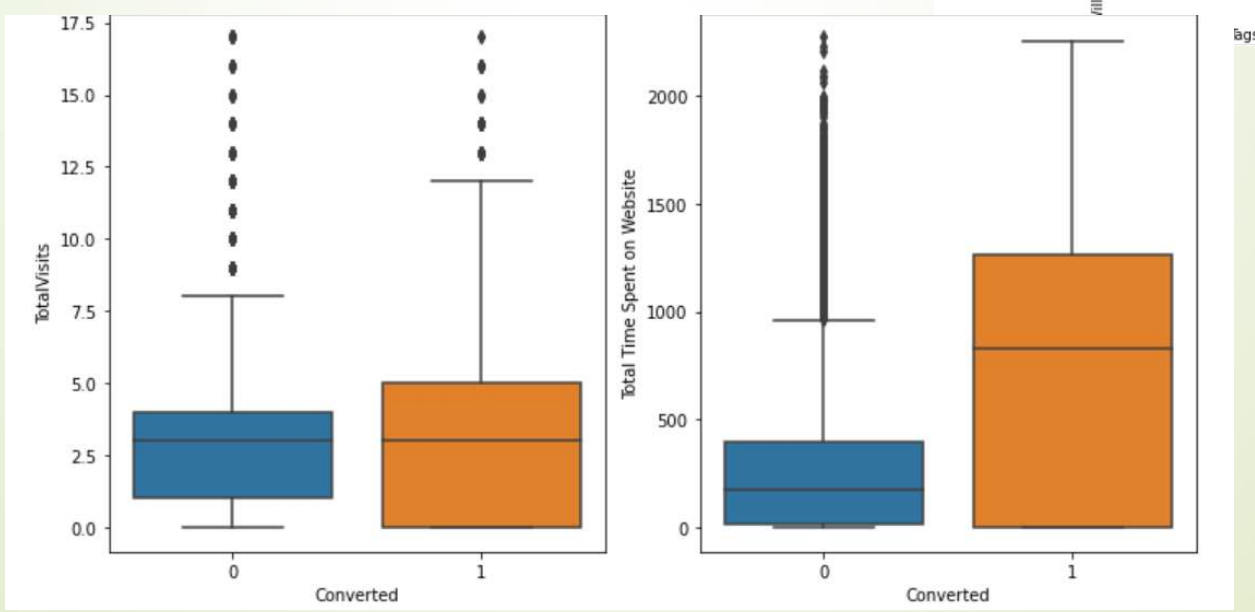

Converted
0
1
count
0
1
Do Not Email


Converted
0
1
count
0
1
Do Not Call


Converted
0
1
count
Page Visited on Website
Email Opened
Unreachable
Converted to Lead
Olark Chat Conversation
Email Bounced
Email Link Clicked
Form Submitted on Website
Unsubscribed
Other Activity
SMS Sent
Last Activity


Converted
0
1
Unemployed
Student
Working Professional
Businessman
Other
Housewife
What is your current occupation


Converted
0
1
count
Specialization
Not Specified
Business Administration
Media and Advertising
Supply Chain Management
IT Projects Management
Finance Management
Travel and Tourism
Human Resource Management
Marketing Management
Banking, Investment And Insurance
International Business
E-COMMERCE
Operations Management
Retail Management
Services Excellence
Hospitality Management
Rural and Agribusiness
Healthcare Management
E-Business
Lead Quality
Low in Relevance
Not Sure
Might be
Worst
High in Relevance
Interested in other courses
Ringing
Will revert after reading the email
Lost to EINS
Other_Tags
Busy
switched off
Already a student
Closed by Horizzon
Not doing further education
invalid number
wrong number given
Interested in full time MBA


TotalVisits
Converted
Total Time Spent on Website
Converted

EDA - INFERENCES

- Lead Score
  - Maximum number of leads are generated by Google and Direct traffic.
  - Conversion Rate of reference leads and leads through welingak website is high
- Lead Origin
  - API and Landing Page Submission has less conversion rate but considerable counts of the leads.
  - Conversion rate is very high from the Lead Add Form but count of leads is low.
- Users spending more time on the website are more likely to get converted.
- Last Activity
  - The conversion rate of "SMS sent" as last activity is maximum.
  - "Email Opened" activity is having maximum count.
- High conversion rate observed in Working professionals.
- Tags
  - 'Will revert after reading the email' and 'Closed by Horizon' have high conversion rate.

MODEL BUILDING

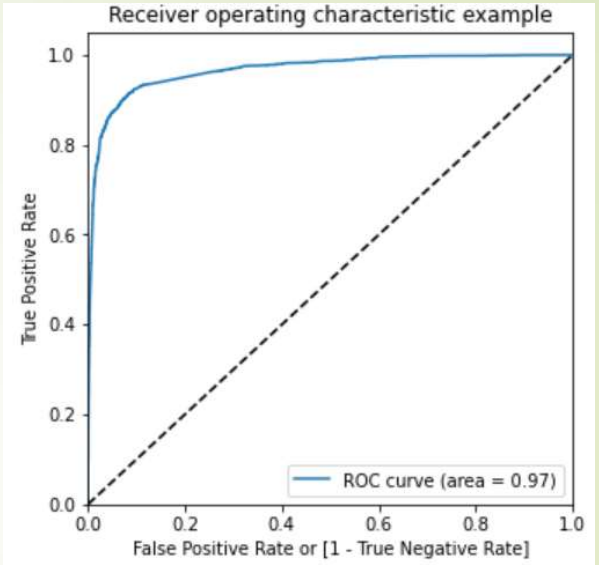
- Splitting data into train and test set (70:30 ratio)
- Use RFE to choose top 20 feature variables for model building
- Repeated model building by removing variable whose p-value is more than 0.05 and VIF value greater than 5
- Final model is built and with probability threshold value of 0.5, below are evaluation metrics of the model,
  - Accuracy – 91.9 %
  - Sensitivity – 87.7%
  - Specificity – 94.5 %

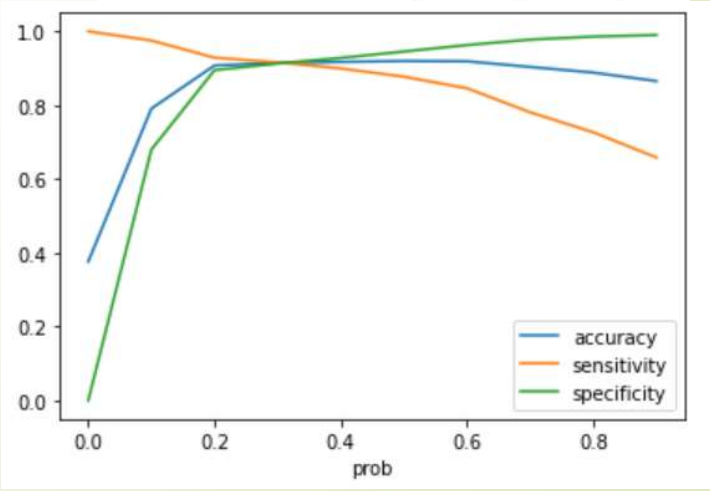
MODEL BUILDING (Cont.)

- Balancing all metrics, 0.3 is taken as optimal cut off point
- With 0.3 as cut off point, below are evaluation metrics for train & test data set

	Train set	Test set
Accuracy	91.4 %	90.7 %
Sensitivity	91.6 %	92.1 %
Specificity	91.2 %	89.9 %

- **Average Lead score for converted customers is <u>90.</u>**





CONCLUSION & RECOMMEDATIONS

- Below are variables that impact to positive conversation rate,
  - Total Visits
  - Total Time spent on website
  - Lead origin - Lead Add form
  - Lead Source
    - Olark Chat
    - Welingak website
  - Current occupation – Working professional
  - Tags
    - Closed By Horizon
    - Busy
    - Lost to EINS
    - Will revert after reading email
  - Last notable activity – SMS sent

With the above information, X Education can improve their sales and increase profit by targeting potential leads.

Higher conversion rate with optimal use of resources can be achieved with different threshold value strategies based on company's targets.