

# LEAD SCORE CASE STUDY

---

BY ANJALI MENON

# 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 Objective

---

- X education wants to know most promising leads.
- For that they want to build a Model which identifies the hot lead.
- Deployment of the model for the future use

# Case Study Approach

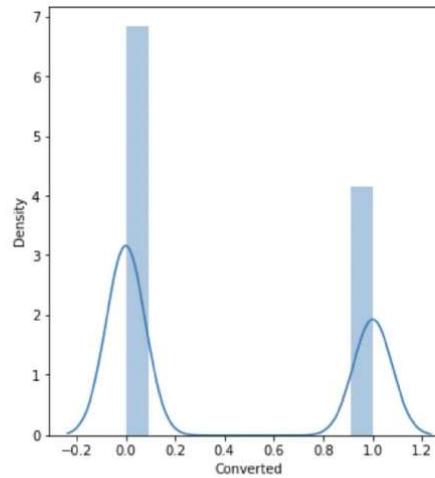
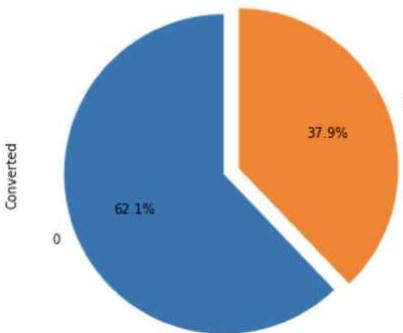
---

- *Understanding Problem statement*
- *Understanding Data*
- *Data Cleaning by handling missing values and unique variables.*
- *Exploratory Data Analysis by performing Univariate , Bivariate and Multi-variate analysis*
- *Data interpretation*
- *Data preparation for Modelling*
- *Logistic regression Model building*
- *Model Evaluation*
- *Conclusion*

# Data Cleaning and Manipulation

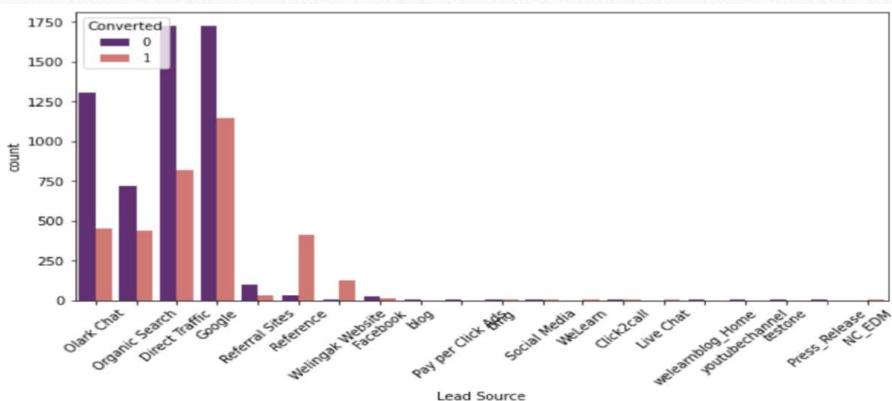
- Total Number of Rows =37, Total Number of Columns =9240.
- Single value features like “Magazine”, “Receive More Updates About Our Courses”, “Update me on Supply”
- “Chain Content”, “Get updates on DM Content”, “I agree to pay the amount through cheque” etc. have been dropped.
- Removing the “Prospect ID” and “Lead Number” which is not necessary for the analysis.
- After checking for the value counts for some of the object type variables, we find some of the features which has no enough variance, which we have dropped, the features are: “Do Not Call”, “What matters most to you in choosing course”, “Search”, “Newspaper Article”, “X Education Forums”, “Newspaper”, “Digital Advertisement” etc.
- Dropping the columns having more than 35% as missing value such as ‘How did you hear about X Education’ and ‘Lead Profile’.

# Exploratory Data Analysis

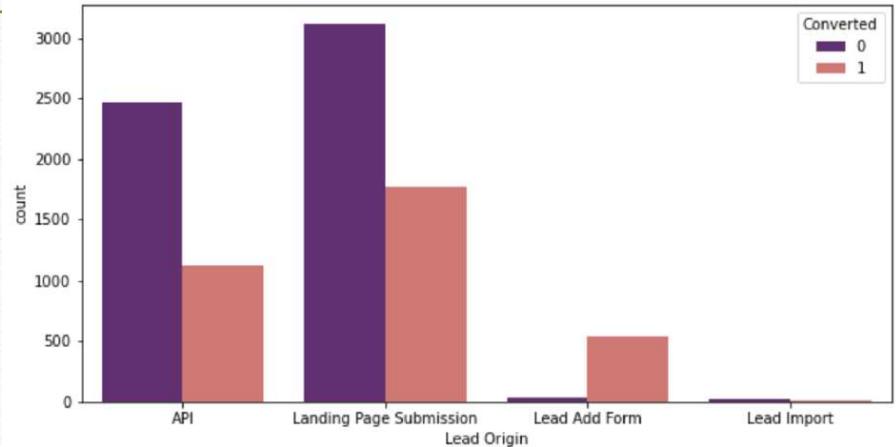


- Hence we can say that around 39% of the leads are converted whereas 61% of the leads are not converted

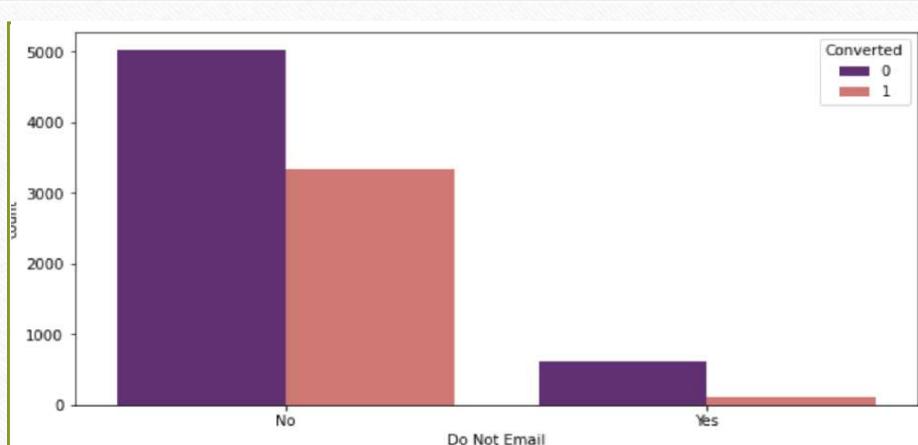
# Lead Score



- The count of leads from the Google and Direct Traffic is maximum
- The conversion rate of the leads from Reference and Welingak Website is maximum

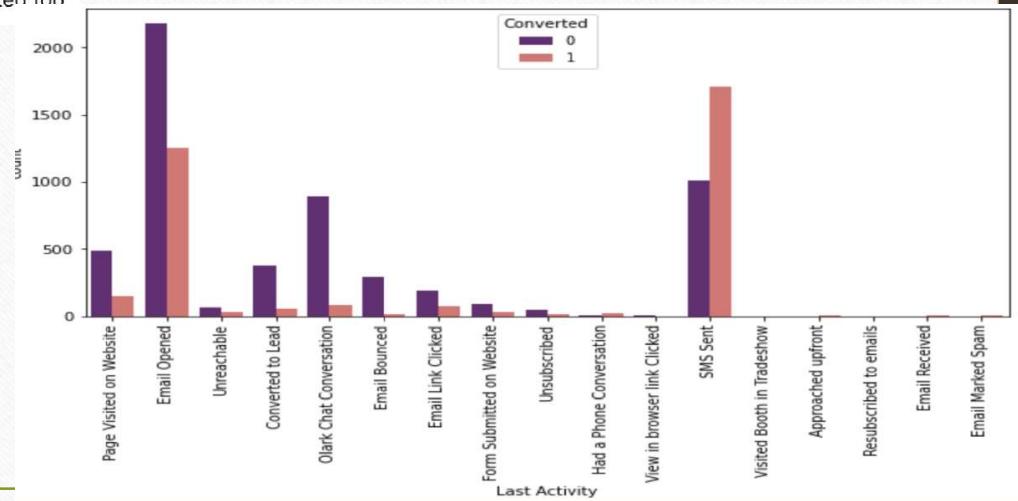


- The count of leads from the Lead Add Form is pretty low but the conversion rate is very high

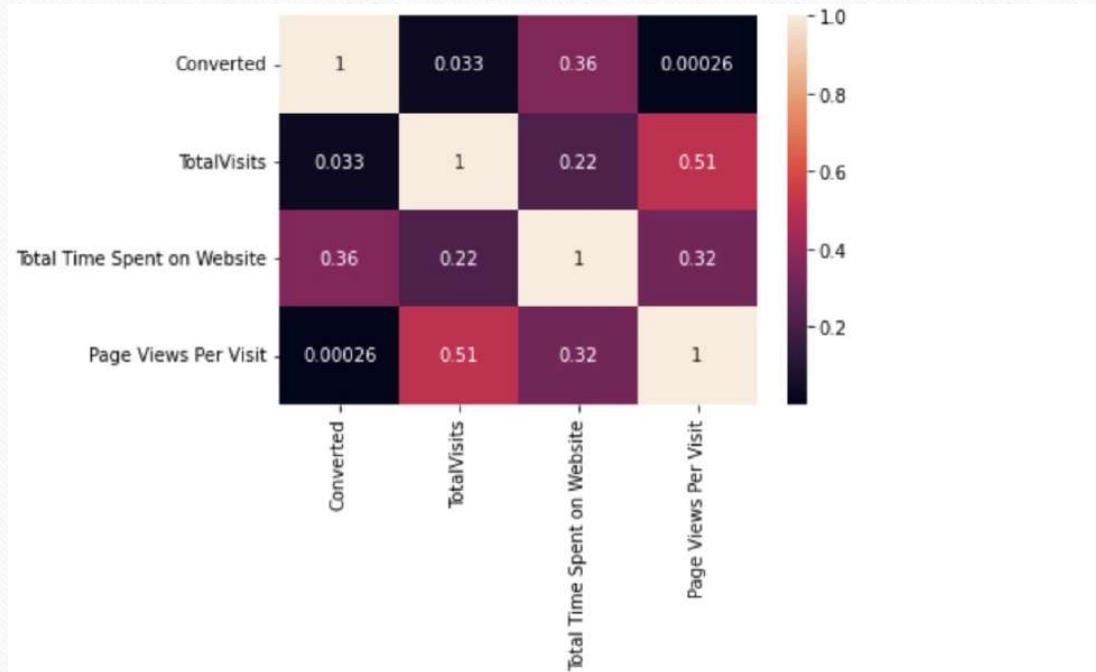


## LEAD PREFERENCE

- As we can see high number of leads opted to email them and most of them are converted too



The Lead Conversion rate is higher when the information sent via SMS



There is relatively a higher corelation for total visits and page views per visit

# DATA PREPARATION

---

- Dummy Variables are created for object type variables
- Total Rows for Analysis: 9074
- Total Columns for Analysis: 20

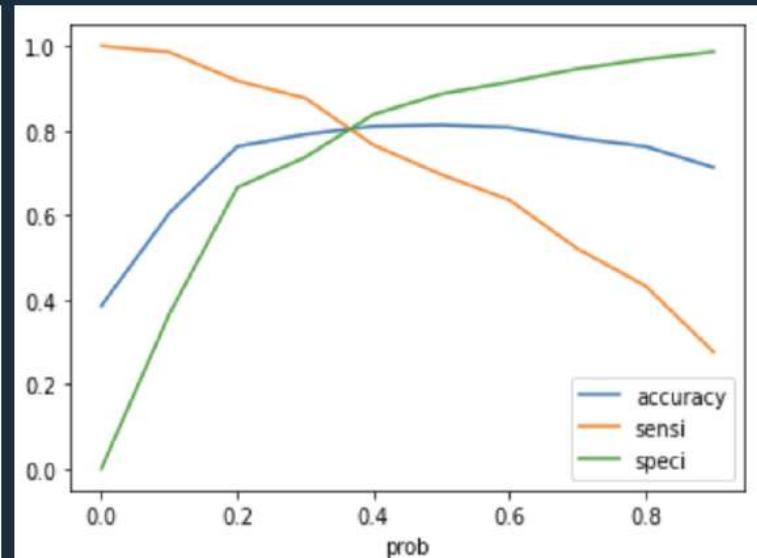
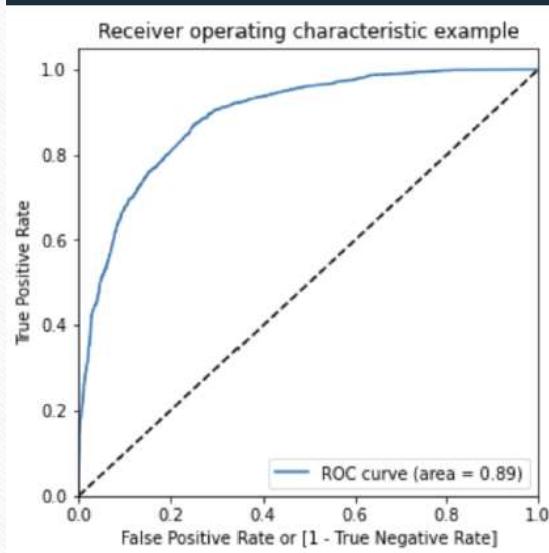
# Model Building

---

- Train - Test split was done at 70% and 30% respectively.
- RFE was done to attain the top 15 relevant variables.
- Removed variables manually depending on the VIF values and p-value.
- Predictions on test data set
- Overall accuracy 81%

# RoC CURVE

From the second graph it is visible that the optimal cut off is at 0.35.



# Conclusion

---

- The probability of lead getting converted is high when 'Lead Origin' is 'Lead add form'
- Large number of leads come from Google and direct traffic but referral sites convert most lead conversions.
- Leads opting for emailing option have more probability of getting converted.
- Conversion rate is higher when the information is sent through SMS.
- Unemployed people have more conversion rate as well as more count.
- The model gives accuracy of 81%.
- The optimal cutoff comes out to be 0.35
- With the help of this model if the sales team contact only the leads with high lead score, their conversion rate will increase to more than 80% as expected.

# Suggestions for Business

---

- When additional interns present:
- Phone calls can be done to people who comes to website repeatedly. Also make the website more interesting to engage potential customers.
- Make calls to leads whose last activity is through SMS or through Olark chat conversation and working professionals.
- When the Target is already achieved:
- Focus more on other methods like automated emails and SMS instead of calls.
- Call only when it is an emergency for customers having very high chance of buying the course.