# Lead Scoring Case Study-Subjective Questions

1. Which are the top three variables in your model which contribute most towards the probability of a lead getting converted?

After conducting logistic model building and evaluation, Exploratory Data Analysis of a given data set, we have found the following three variables that are most relevant in lead conversion probability.

Please note along with data analysis, domain knowledge, marketing strategies and story building is also used for analysis of variables. These variables are by no means the only ones that determine the probability. As per our research, these variables seemed most relevant.

a) Lead Source :
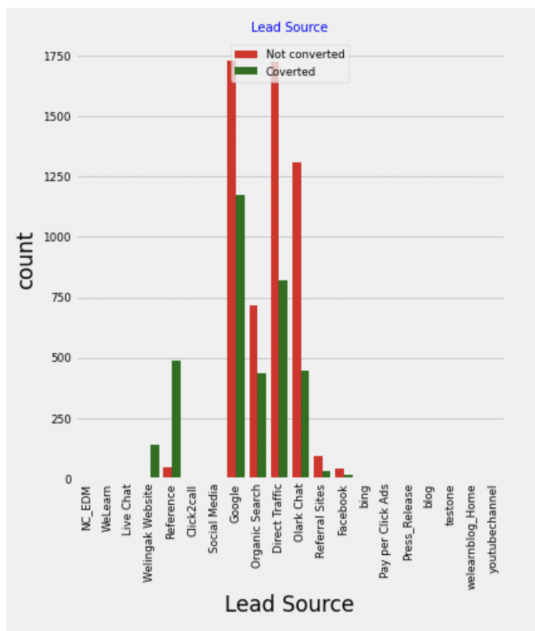
- This variable tells the source of the lead. Includes Google, Organic Search, Olark Chat, etc.

- Reference: Customers coming through reference has a high conversion rate and generally is a cheaper way of lead generation. Also reference builds a strong customer base due to peer views/trust impact.

- Welingak Website: As can be seen, even with a limited number of leads, customers coming through this website are hot and should be followed up.

- Google and Direct Traffic can generate a lot of potential customers. Individually their conversion rate is 35 to 40%. Hence these customers will require certain convincing from the sales team.

b) Time spent on Website:
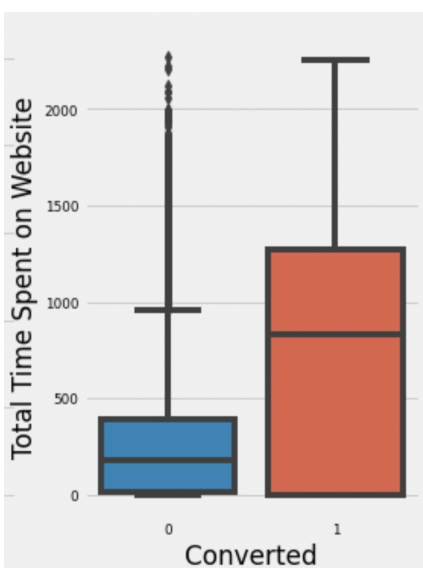
- This variable measures how much time a customer spends on the website.

- If a customer spends more time on the website, the probability of conversion is higher.

- As can be seen from the box plot, customers have spent an average of total 1300 minutes on 'X Education Website', this lead has a high probability of conversion.

c) Last Activity:
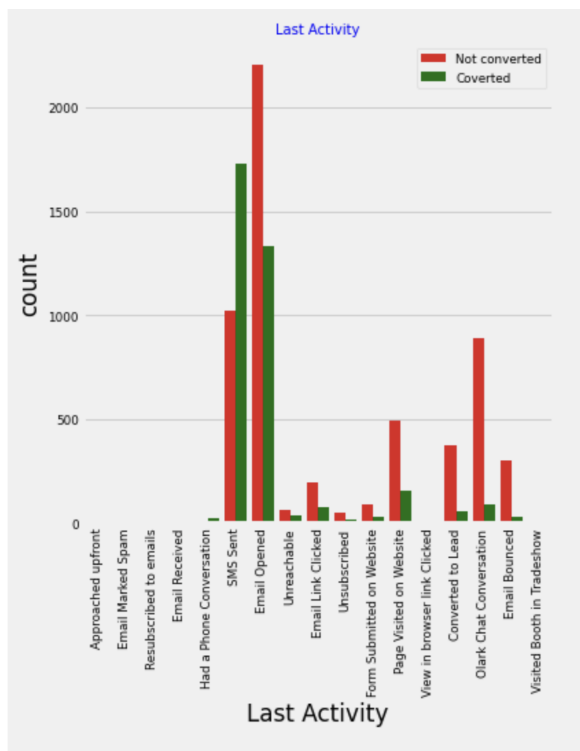
- Last activity performed by the customer. Includes Email Opened, Olark Chat Conversation, etc.

- Last Activity by the customer can be indicative of how far is into the decision making process. Usually interested candidates will check their emails and send any required sms to X Education Institute.

- As can been from the graph, SMS sent and Email Opened leads to hot candidates, conversion rate is about 40-60%
- Leads whose last activity was 'Olark Chat Conversion' and 'Email Bounced' are cold leads. Usually customers who have email bounced might have given incorrect/fake email ids. Olark customer interaction is obviously a problem since they have fewer conversion rates.

2. What are the top 3 categorical/dummy variables in the model which should be focused the most on in order to increase the probability of lead conversion?
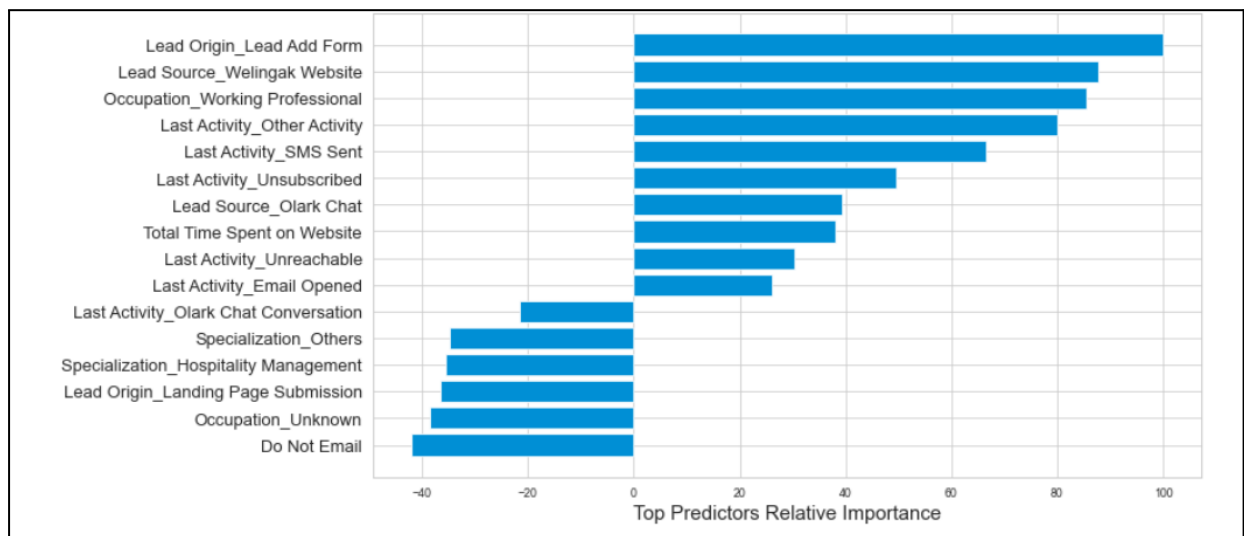
a) Lead Origin_Load Add Form
b) Lead Source_Wellingak Website
c) Occupation_Working Professional

3. X Education has a period of 2 months every year during which they hire some interns. The sales team, in particular, has around 10 interns allotted to them. So during this phase, they wish to make the lead conversion more aggressive. So they want almost all of the potential leads (i.e. the customers who have been predicted as 1 by the model) to be converted and hence, want to make phone calls to as much of such people as possible. Suggest a good strategy they should employ at this stage.

After evaluating the model and EDA, we have come up with the following strategies: all customers who are predicted 1 get a call from the sales team.

Lead Source-Welligak Website

Interns should call customers coming from Welligak Website as they have a very high lead conversion rate.

Occupation-Unemployed and Working Professionals

Interns should call customers who are Unemployed and Working Professional as they have a very high lead conversion rate. Unemployed customers will be more keen on upskilling themselves to get better job opportunities. Working Professionals can afford the course in 'X' education institute.

Total Time Spent on Website

Interns should call customers spending more time on the website (approx 1300 minutes) as it is statistically significant and hence higher chances of conversion.

Last Activity -Email Opened

Interns should call customers coming and have checked their email i.e. their last activity is email seen as very hot leads.

Last Activity- SMS sent

Interns should call customers whose last activity is 'SMS sent'. We do not know as of now what information this activity is conveying to the customer however, this is a very important strategy to keep in mind as this is generating almost 50% lead conversion rate.

Leads who do not opt for Email

Interns should call customers who mark 'No' in the Email option on the website.

Last Activity-Olark chat conversation and Last Activity Email Bounced

Interns should call customers whose Last activity has been Olark chat conversation should not be contacted, very poor leads. This can be researched further and if required investment in this activity can be removed. There is a strong negative relation to this variable against lead conversion both from EDA and ML insights.

Customers who give fake emails need to be avoided as well.

Lead Origin-Landing Page Submission

Interns should call not prioritize calling customers having lead origin from Landing Page Submission

Lead Source-Reference

Interns should call customers who come from reference as they have very high lead conversion rate.

Last Acitvity-Olark Chat Conversion and Email Bounced

Leads whose last activity was 'Olark Chat Conversion' and 'Email Bounced' are cold leads. Usually customers who have email bounced might have given incorrect/fake email ids. Olark customer interaction is obviously a problem since they have few less conversion rate

4. Similarly, at times, the company reaches its target for a quarter before the deadline. During this time, the company wants the sales team to focus on some new work as well. So during this time, the company's aim is to not make phone calls unless it's extremely necessary, i.e. they want to minimize the rate of useless phone calls. Suggest a strategy they should employ at this stage.

These leads can be called as very hot leads
Please refer to the python coding file for graphs and more details.

1. First target those customers who are coming from variables resulting in 100% conversion rate Eg Lead sources such as WeLearn and Live Chat Look.
2. Second, target the next best variable i.e. customers coming from referrals. Customers acquired through referrals or word of mouth publicity are the most reliable
3. First two strategies are good but they might not work on large volumes of data. In order to address the volume, start filtering out the cold leads i.e. people with bounced email, people who do not spend much time on the website etc.
4. With whatever is left behind the strategies adopted in the previous question can be applied.