Machine Learning 1

Lead Scoring Case Study

**For better interpretability of the model we have computed the lead score (which is the probability percentage) and also the odds and log odds (section 9 of the Python Notebook) and this will help to answer the subjective questions.**

**Note:** likelihood of conversion as a probability is difficult to interpret, the logistic regression equation can be converted into a more suitable form for interpretaion by linearising it. For this we can compute the log odds and odds. log odds = beta0 + beta1 x1 + beta2 x2 + ....

## Question 1

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

The variables that contribute most towards the probability of lead getting converted are the ones with the highest coefficients (beta value).

* Lead Origin\_Lead Add Form (3.764)
* Occupation\_Working Professional (2.8031)
* Lead Source\_Welingak Website (1.9529)

## Question 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?**

* **Occupation:** Focus on working professionals, business men, house wife (than focusing on unemployed and students).
* **Lead Source:** Focus on leads sourced from Welingak Website and Olark Chats
* **Lead Origin:** Focus on leads received from the lead add form in the website.

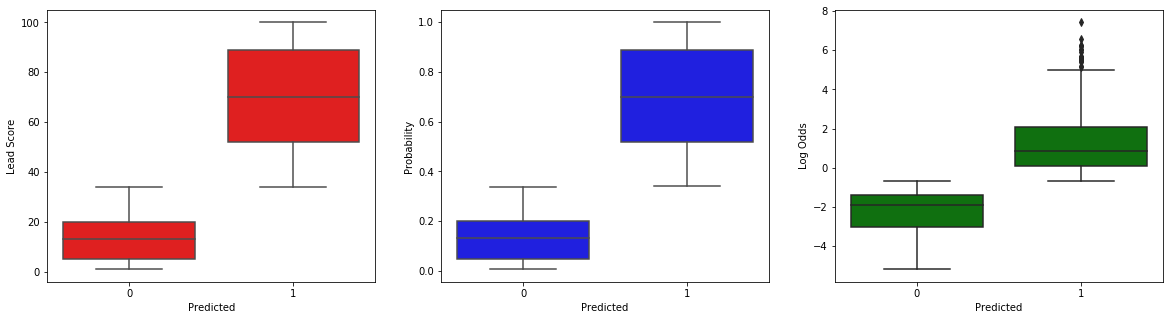
## Question 3

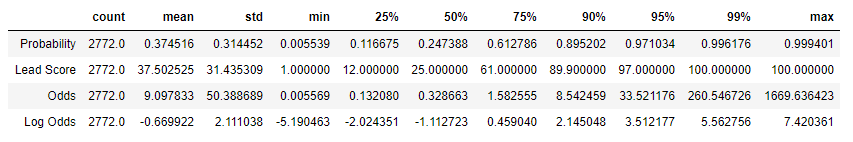
Scenario: X Education has a period of 2 months every year during which they hire some interns. The sales team, in particular, has around 10 interns allotted to them. So during this phase, they wish to make the lead conversion more aggressive. So they want almost all of the potential leads (i.e. the customers 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.**

We have computed the lead score, log odds and odds.

The distribution is shown below for the test data:





Since there are additional resources (10 interns) we can take the leads above the 50th percentile (note: below 50th percentile, the odds of conversion i.e. how likely they are to get converted) becomes less than 0). The lead score for these people is 50 and above. The percentile cut off can be thus adjusted to ensure that the interns are engaged as well as the conversion rate is higher.

## Question 4

Scenario. 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.**

Since there are limitation of resources we can take the leads above the 80th percentile (note: the odds of conversion here is more than 2 times than the other leads) or making calls. The lead score for these people is 73 and above. The percentile cut off can be thus adjusted to ensure that the phone calls are made only if it is exteremely necessary.

