

## Assignment Subjective Questions:

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

**Answer:**

Final model with variables

	coef	std err	z	P> z	[0.025	0.975]
const	0.4006	0.337	1.188	0.235	-0.260	1.061
Do Not Email	-1.0560	0.246	-4.299	0.000	-1.537	-0.575
Total Time Spent on Website	1.1097	0.062	17.970	0.000	0.989	1.231
Direct Traffic	-1.7135	0.172	-9.977	0.000	-2.050	-1.377
Google	-1.3184	0.158	-8.328	0.000	-1.629	-1.008
Organic Search	-1.3802	0.197	-6.991	0.000	-1.767	-0.993
Referral Sites	-1.4254	0.485	-2.938	0.003	-2.376	-0.475
Welingak Website	4.9708	1.024	4.853	0.000	2.963	6.978
Last Activity_SMS Sent	1.9502	0.118	16.571	0.000	1.720	2.181
What is your current occupation_Unemployed	-0.7171	0.316	-2.270	0.023	-1.336	-0.098
Tags_Already a student	-3.3307	1.035	-3.217	0.001	-5.360	-1.301
Tags_Busy	0.9300	0.237	3.932	0.000	0.466	1.394
Tags_Closed by Horizzon	6.8430	0.729	9.389	0.000	5.415	8.271
Tags_Interested in full time MBA	-1.7512	0.788	-2.223	0.026	-3.295	-0.207
Tags_Interested in other courses	-1.6051	0.361	-4.443	0.000	-2.313	-0.897
Tags_Lost to EINS	7.0167	0.816	8.595	0.000	5.417	8.617
Tags_Ringing	-3.2726	0.241	-13.604	0.000	-3.744	-2.801
Tags_Will revert after reading the email	4.6329	0.197	23.498	0.000	4.246	5.019
Tags_switched off	-3.6637	0.612	-5.991	0.000	-4.862	-2.465
Lead Quality_Worst	-2.2515	0.674	-3.340	0.001	-3.573	-0.930
Last Notable Activity_Email Link Clicked	-1.2281	0.440	-2.788	0.005	-2.091	-0.365
Last Notable Activity_Modified	-1.7171	0.128	-13.372	0.000	-1.969	-1.465
Last Notable Activity_Olark Chat Conversation	-1.2557	0.372	-3.373	0.001	-1.985	-0.526

$Y=mx+c$

1 unit change in x results in m units of change in y

If m is positive it results in increase of y as x increases

Top 3 variables that improve probability of lead getting converted are

Tags

- 1) Tags\_Lost to EINS
- 2) Tags\_Closed by Horizzon
- 3) Welingak Website

Next 3 are

- 4) Tags\_Will revert after reading the email
- 5) Last Activity\_SMS Sent
- 6) Total Time Spent on Website

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?

**Answer:**

Assuming probability of lead conversion is converted leads/ total leads

To improve the conversion rate, we have focus on the categorical variables which are having high leads but with low conversion.

Dummy variables	Lead Count	Converted	Conversion %
Google	8358	3331	39.85
Direct Traffic	8159	2716	33.29
Tags_Will revert after reading the email	3267	680	20.81
Organic Search	2873	1147	39.92
Tags_Already a student	2716	1705	62.78
Lead Quality_Worst	2543	818	32.17
Last Notable Activity_Modified	2016	1957	97.07
Tags_Ringing	1187	34	2.86
What is your current occupation_Unemployed	1154	436	37.78
Last Activity_SMS Sent	601	12	2
Tags_Interested in other courses	509	13	2.55
Tags_Closed by Horizon	465	3	0.65
Last Notable Activity_Olark Chat Conversation	301	299	99.34
Do Not Email (No)	240	4	1.67
Welingak Website	185	105	56.76
Referral Sites	183	25	13.66
Tags_switched off	173	45	26.01
Tags_Busy	172	168	97.67
Tags_Lost to EINS	129	127	98.45
Tags_Interested in full time MBA	125	31	24.8
Last Notable Activity_Email Link Clicked	116	3	2.59

From the above table

We have to focus more on the below variables

Google, Direct traffic, Will revert after reading mail because they have high lead count but low conversion rate. So, by increase its conversion we can increase overall probability of conversion (lead conversion rate).

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.

**Answer:**

Here we choose the cutoff as 0.28 so that we have good values of sensitivity, specificity and accuracy. By choosing 0.28 as cut off we include most of the users and hence we will be able to make them convert.

Conversion_prob	accuracy	sensitivity	specificity	
0.1	0.1	0.856243	0.971778	0.787277
0.2	0.2	0.894347	0.952822	0.859442
0.3	0.3	0.920170	0.911542	0.925321
0.4	0.4	0.924421	0.893008	0.943173
0.5	0.5	0.927098	0.874895	0.958260
0.6	0.6	0.927098	0.855097	0.970078
0.7	0.7	0.915446	0.804971	0.981393
0.8	0.8	0.909778	0.781382	0.986422
0.9	0.9	0.891986	0.728307	0.989691

Since we want to make as much conversions as possible, we can concentrate on the persons who has lead score greater than or equal to 28 (as model predicted 1 for score greater than or equal to 28 lead score) and having a phone conversation is one of the best way to convert leads.

Concentrate more on the leads from the following since they have low conversion rate, we must increase its conversion, to reach the goal of 80% conversion rate

- Last activities like Email opened and SMS sent have more lead but less conversion. So, try to have phone conversion with such leads.
- Most of the leads are unemployed and they have low conversion rate. To increase conversion rate, provide some scholarship schemes or paid internships to them to cut down their fee expenses.
- Lead who spend more time in website have good conversion rate. So, ask leads to visit site and allow to see them some demo lectures and how internship program is designed videos
- Concentrate on leads from API, landing page submissions who has low conversion rates. Try to increase the leads from lead add forms since they have high conversion rates.
- Also leads from lead sources like Direct Traffic, Google, Olark Chat, Organic Search

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.

**Answer:**

In this case we will contact the persons whose lead score is of medium values or high as they have high chance of probability to convert with limited calls.

- Leads who choose do not call option as they have very high conversion rate and not need to spend staff in calling them.
- Leads who approached upfront
- Housewives, Business and working professionals
- 4) Referred leads
- 5) Welingak website sourced leads
- As these leads can be converted with limited number or no calls.

Since we want to make conversions with limited staff and calls, we can concentrate on those leads whose lead score  $>70$ .