Subjective Questions – Lead Scoring Case Study

Q1. Which are the top three variables in your model which contribute most towards the probability of a lead getting converted?

A1. According to the coefficient values in the model build, following are the top 3 variables that are contributing the most towards the probability of lead getting converted.

1. What is your current occupation_housewife (from What is your current occupation)
2. Last Activity_email marked spam (from Last Activity)
3. Last Activity_email received (from Last Activity)

TotalVisits	1.1360	0.263	4.320	0.000	0.621	1.651
Total Time Spent on Website	4.5859	0.173	26.471	0.000	4.246	4.925
Specialization_banking, investment and insurance	0.4788	0.195	2.455	0.014	0.097	0.861
Specialization_business administration	0.0995	0.189	0.527	0.599	-0.271	0.470
Specialization_e-business	0.2447	0.455	0.537	0.591	-0.648	1.137
Specialization_e-commerce	0.4713	0.315	1.496	0.135	-0.146	1.089
Specialization_finance management	0.2945	0.136	2.158	0.031	0.027	0.562
Specialization_healthcare management	0.0468	0.273	0.171	0.864	-0.488	0.582
Specialization_hospitality management	-0.4082	0.335	-1.217	0.224	-1.066	0.249
Specialization_human resource management	0.1460	0.141	1.036	0.300	-0.130	0.422
Specialization_international business	0.1776	0.257	0.692	0.489	-0.325	0.681
Specialization_it projects management	0.3280	0.200	1.640	0.101	-0.064	0.720
Specialization_marketing management	0.4441	0.141	3.149	0.002	0.168	0.721
Specialization_media and advertising	0.0631	0.239	0.264	0.792	-0.406	0.532
Specialization_operations management	0.2626	0.170	1.541	0.123	-0.071	0.597
Specialization_retail management	-0.4674	0.357	-1.308	0.191	-1.168	0.233
Specialization_rural and agribusiness	0.6460	0.365	1.770	0.077	-0.070	1.362
Specialization_services excellence	0.3680	0.641	0.574	0.566	-0.889	1.625
Specialization_supply chain management	0.1979	0.197	1.004	0.315	-0.188	0.584
Specialization_travel and tourism	0.1118	0.251	0.445	0.656	-0.380	0.604
Lead Source_google	0.4023	0.092	4.353	0.000	0.221	0.583
Lead Source_olark chat	2.0149	0.147	13.706	0.000	1.727	2.303
Lead Source_organic search	0.2118	0.126	1.674	0.094	-0.036	0.460
Lead Source_others	1.8326	0.611	3.000	0.003	0.635	3.030
Lead Source_reference	3.9919	0.243	16.458	0.000	3.517	4.467
Lead Source_referral sites	0.3300	0.371	0.889	0.374	-0.397	1.057
Lead Source_welingak website	5.8821	0.736	7.995	0.000	4.440	7.324
Last Activity_converted to lead	-1.2489	0.231	-5.403	0.000	-1.702	-0.796
Last Activity_email bounced	-1.8830	0.320	-5.877	0.000	-2.511	-1.255
Last Activity_email link clicked	-0.5207	0.221	-2.356	0.018	-0.954	-0.088
Last Activity_email marked spam	20.5999	4.82e+04	0.000	1.000	-9.44e+04	9.45e+04
Last Activity_email received	19.0227	4.82e+04	0.000	1.000	-9.44e+04	9.45e+04
Last Activity_form submitted on website	-0.8466	0.342	-2.478	0.013	-1.516	-0.177
Last Activity_had a phone conversation	1.9766	0.725	2.728	0.006	0.556	3.397
Last Activity_olark chat conversation	-1.4929	0.170	-8.757	0.000	-1.827	-1.159
Last Activity_page visited on website	-0.5999	0.157	-3.818	0.000	-0.908	-0.292
Last Activity_resubscribed to emails	1.794e-10	4.79e-07	0.000	1.000	-9.38e-07	9.38e-07
Last Activity_sms sent	1.0687	0.081	13.223	0.000	0.910	1.227
Last Activity_unreachable	-0.3402	0.342	-0.994	0.320	-1.011	0.330
Last Activity_unsubscribed	-0.4735	0.461	-1.028	0.304	-1.376	0.429
Last Activity_view in browser link clicked	-21.9411	2.32e+04	-0.001	0.999	-4.55e+04	4.55e+04
Last Activity_visited booth in tradeshow	-21.1695	4.82e+04	-0.000	1.000	-9.45e+04	9.44e+04
What is your current occupation_housewife	24.0387	1.48e+04	0.002	0.999	-2.89e+04	2.9e+04

Q2. 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?

A2. From the above snapshot, it is seen from the coefficient values, that top 3 categorical/dummy variables in the model which should be focused on in order to increase the probability of lead conversion are:

1. What is your current occupation_housewife (from What is your current occupation)
2. Last Activity_email marked spam (from Last Activity)
3. Last Activity_email received (from Last Activity)

Q3. 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.

A3. The final predicted value is calculated on the basis of the optimal cut-off value of 3.5.
In order to make the sales, the company can contact the leads that have a conversion probability of value 1 under the cut-off value of 0.3.

	Converted	Conversion_Probability	Predicted	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	Final_Predicted
0	1	0.992210	1	1	1	1	1	1	1	1	1	1	1	1
1	1	0.653288	1	1	1	1	1	1	1	1	0	0	0	1
2	1	0.819048	1	1	1	1	1	1	1	1	1	1	0	1
3	0	0.868918	1	1	1	1	1	1	1	1	1	1	0	1
4	0	0.340530	0	1	1	1	1	0	0	0	0	0	0	0
5	0	0.045453	0	1	0	0	0	0	0	0	0	0	0	0
6	1	0.828854	1	1	1	1	1	1	1	1	1	1	0	1
7	0	0.050620	0	1	0	0	0	0	0	0	0	0	0	0
8	0	0.245582	0	1	1	1	0	0	0	0	0	0	0	0
9	0	0.314055	0	1	1	1	1	0	0	0	0	0	0	0

Q4. 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.

A4. In order to make only necessary calls and minimize the rate if useless phone calls, the company should contact the leads that have conversion probability of value 1 and are under the 0.7 column in the below snapshot.

If this case is taken up, then we might miss some of the leads that are actually converted but wrongly predicted as unconverted by the model. The **blue circles** in the snapshot represent the leads that we might miss.

	Converted	Conversion_Probability	Predicted	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	Final_Predicted
0	1	0.992210	1	1	1	1	1	1	1	1	1	1	1	1
1	1	0.653288	1	1	1	1	1	1	1	1	0	0	0	1
2	1	0.819048	1	1	1	1	1	1	1	1	1	1	0	1
3	0	0.868918	1	1	1	1	1	1	1	1	1	1	0	1
4	0	0.340530	0	1	1	1	1	0	0	0	0	0	0	0
5	0	0.045453	0	1	0	0	0	0	0	0	0	0	0	0
6	1	0.828854	1	1	1	1	1	1	1	1	1	1	0	1
7	0	0.050620	0	1	0	0	0	0	0	0	0	0	0	0
8	0	0.245582	0	1	1	1	0	0	0	0	0	0	0	0
9	0	0.314055	0	1	1	1	1	0	0	0	0	0	0	0
10	0	0.108039	0	1	1	0	0	0	0	0	0	0	0	0
11	0	0.072092	0	1	0	0	0	0	0	0	0	0	0	0
12	1	0.186058	0	1	1	0	0	0	0	0	0	0	0	0
13	0	0.117709	0	1	1	0	0	0	0	0	0	0	0	0
14	0	0.324953	0	1	1	1	1	0	0	0	0	0	0	0
15	0	0.031655	0	1	0	0	0	0	0	0	0	0	0	0
16	0	0.244848	0	1	1	1	0	0	0	0	0	0	0	0
17	0	0.112136	0	1	1	0	0	0	0	0	0	0	0	0