

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

Ans. Top three variables are

“Total Time Spent on Website”: 4.53

“Lead Origin_Lead Add Form”: 3.82

“What is your current occupation_Working Professional”: 3.56

| | coef | std err | z | P> z | [0.025 | 0.975] |
|--|---------|---------|---------|-------|--------|--------|
| const | -3.7318 | 0.137 | -27.288 | 0.000 | -4.000 | -3.464 |
| TotalVisits | 0.9757 | 0.174 | 5.612 | 0.000 | 0.635 | 1.316 |
| Total Time Spent on Website | 4.5308 | 0.166 | 27.363 | 0.000 | 4.206 | 4.855 |
| A free copy of Mastering The Interview | -0.4636 | 0.087 | -5.335 | 0.000 | -0.634 | -0.293 |
| Lead Origin_Lead Add Form | 3.8260 | 0.200 | 19.086 | 0.000 | 3.433 | 4.219 |
| Lead Source_Olark Chat | 1.5141 | 0.124 | 12.214 | 0.000 | 1.271 | 1.757 |
| Lead Source_Welingak Website | 2.0669 | 0.743 | 2.780 | 0.005 | 0.610 | 3.524 |
| Specialization_Hospitality Management | -1.0225 | 0.326 | -3.133 | 0.002 | -1.662 | -0.383 |
| What is your current occupation_Other | 1.8294 | 0.519 | 3.524 | 0.000 | 0.812 | 2.847 |
| What is your current occupation_Student | 0.9938 | 0.239 | 4.166 | 0.000 | 0.526 | 1.461 |
| What is your current occupation_Unemployed | 1.0075 | 0.087 | 11.525 | 0.000 | 0.836 | 1.179 |
| What is your current occupation_Working Professional | 3.5612 | 0.198 | 17.990 | 0.000 | 3.173 | 3.949 |
| City_Not Mentioned | -0.3743 | 0.095 | -3.936 | 0.000 | -0.561 | -0.188 |
| Last Notable Activity_Email Opened | 0.7082 | 0.083 | 8.524 | 0.000 | 0.545 | 0.871 |
| Last Notable Activity_Other | 0.9130 | 0.253 | 3.612 | 0.000 | 0.418 | 1.408 |
| Last Notable Activity_SMS Sent | 1.9147 | 0.089 | 21.444 | 0.000 | 1.740 | 2.090 |

- 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?

Ans. Top three Categorical dummy variables are

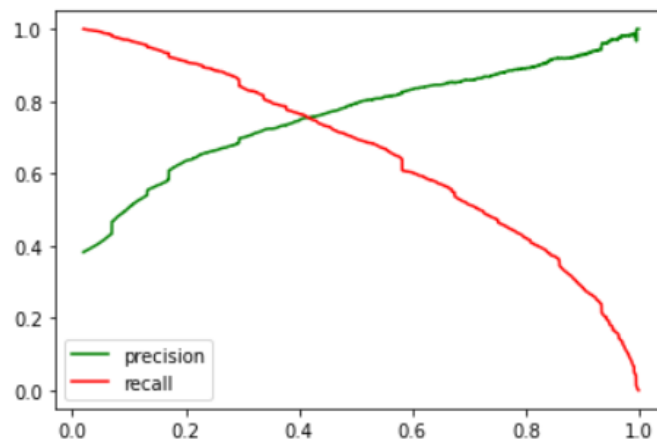
“Lead Origin_Lead Add Form”: 3.82

“What is your current occupation_Working Professional”: 3.56

“Lead Source_Welingak Website”: 2.07

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.

Ans. Looking at the plot below (variation of Precision and Recall with cutoff) it can be seen that: Recall falls with cutoff, but at the same time precision increases with cutoff. To get almost all of the potential leads we should choose a high recall value.



Typical confusion matrix of test set cutoff of .25 is shown below

```
array([[1219, 458],
       [ 120, 975]],
```

Precision for the same is $975/(458+975) = 68\%$

Whereas recall is $975/(120+975)=89\%$

If we want to convert all the potential leads it will be better to keep recall as high as possible, it also means that precision will fall so we will be calling almost 90% of the potential leads at .25 cutoff, but as precision is 68% around 32% of the “hot leads” according to the model will not convert, but as we have 10 interns available, low precision is not a problem.

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.

Ans: To minimize useless calls the precession should be kept as high as possible, so the cutoff should be kept high.

Typical confusion matrix of test set cutoff of .60 is shown below

```
array([[1559, 118],
       [ 466, 629]], dtype=int64)
```

Precession for the same is $629/(118+629) = 84\%$

Whereas recall is $629/(466+629)=57\%$

With high cutoff the precession is high, which means false positives will be low, as shown above with a cutoff of .60 precession is 84%, so according to the model only 16 percent of the total calls will be useless, at a higher cutoff the useless calls will drop further.