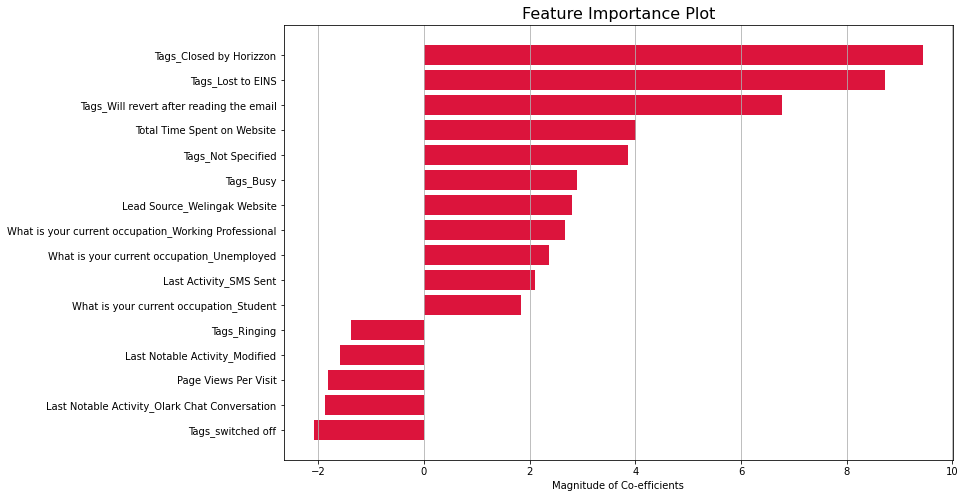
Lead Scoring Assignment

# Subjective Questions and Answers

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

**Ans. As per the feature importance graph plotted below:**

The three most important variables contributing to the model are as follows.

**Tags\_Closed by Horizzon:**

* Positive Co-efficient
* The Sales team should focus on leads which were tagged as Closed by Horizzon as those leads have higher probability of maturing into a customer

**Tags\_Lost to EINS:**

* Positive Co-efficient
* Similarly, the Sales team should focus on leads which were tagged as Lost to EINS as they also have a higher probability of maturing into a customer

**Tags\_Will revert after reading the email:**

* Positive Co-efficient
* Also, they must focus their call backs on leads tagged as Will revert after reading the email.

1. 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:** As per the above plot which shows the feature importance of all features involved in the model, the top 3 categorical variables which should be focussed are:

1. **Tags\_Closed by Horizzon**
2. **Tags\_Lost to EINS**
3. **Tags\_Will revert after reading the email**
4. 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:** Because they now have a larger workforce (inclusion of interns) at their disposal it is suggested that they call as much potential customers as possible.

Here the **Sensitivity - Specificity** trade off comes into play.

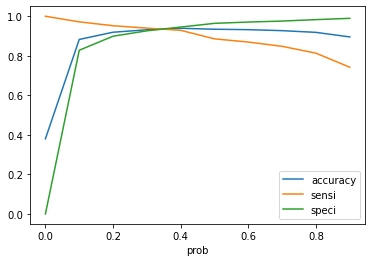
**Sensitivity:** The ratio of Total number of Actual Conversions correctly predicted to the Total no of Actual Conversions.

**Specificity:** The ratio of Total no of Actual non-Conversions correctly predicted to the Total number of Actual non-Conversions.

With variation in threshold values which determine the cutoff for probability we can make trade-off between Sensitivity and Specificity. By keeping the threshold as low as possible we end up in having higher Sensitivity – which means that the model will correctly identify almost all leads who are likely to convert.

This in-turn will help the X Education employees to call as much potential customers as possible.

**Accuracy – Sensitivity – Specificity Curve is given for reference.**

****

From Business perspective, to target all potential leads we should focus on:

* + Unemployed, Students & Management Specialization students who have high conversion rate
  + Leads sourced through References, Google, Direct Traffic have higher conversion rate
  + Leads originating from Lead Add Form, API, Landing page submission have higher conversion

1. 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.** As explained above, here we can make use of **Specificity**. By maintaining a **high threshold value for cut-off probability**, the Specificity will maximize (the model will now correctly identify almost all leads who are not likely to convert). This will be done at the risk of losing some low-conversion rate risky customers.

Since X Education has already reached its target and now wants to focus on some new work, the **agents can make calls only to those most important leads** who have a very high probability of converting into a customer and avoid useless phone calls on risky row conversion rate leads.