**Campaign Analytics Day 2 Workshop**

Prepared by: Norvin Chandra - A0195335Y

1. Which set of exhaustive variables they should consider for developing effective customer targeting strategy? What will be the source of data? What will be frequency of collating data?

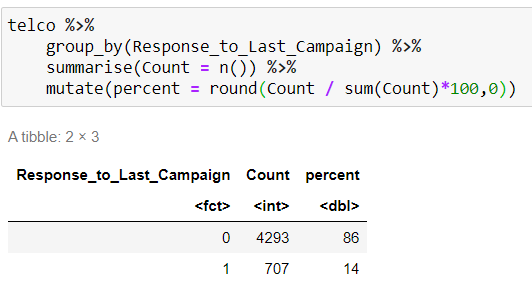
First of all, we need to be clear about the objective of this campaign and feature of the 2 products.

We understand that the marketing department wants to relaunch their value campaign to convince existing customers under “Basic plan” to upgrade to “Value plan”. Before we look at the data, it is important to understand the difference between the 2 plan and what segment of customer each plan targets. With that understanding we can look at the data in informed way. Such information is not provided in this exercise.

Without knowing the difference between “Basic plan” and “Value plan” and assuming that the plan is a postpaid plan, we assume we would need the following information to develop customer targeting strategy.

* 1. **Variables:**
     1. Length of time with company
     2. Time until contract end
     3. Data Usage
     4. Call usage split day and night
     5. Sms usage
     6. Intl call usage
     7. Roaming usage
     8. Current plan
     9. Monthly bill size
     10. Demographic information such as gender, age group, education level
     11. Mode of bill payment
     12. Response to last campaign
  2. **Source of data**
     1. Telco billing system
     2. Telco customer database
  3. **Frequency of collating data**
     1. Monthly

1. Based on last year campaign (Value Offer) data (n =5000) can you identify most impactful variables driving response to campaign.
   1. **Review Data** 
      1. When using past campaign data to target customer, firstly we look at the distribution of the target variable, Response\_to\_Last\_Campaign

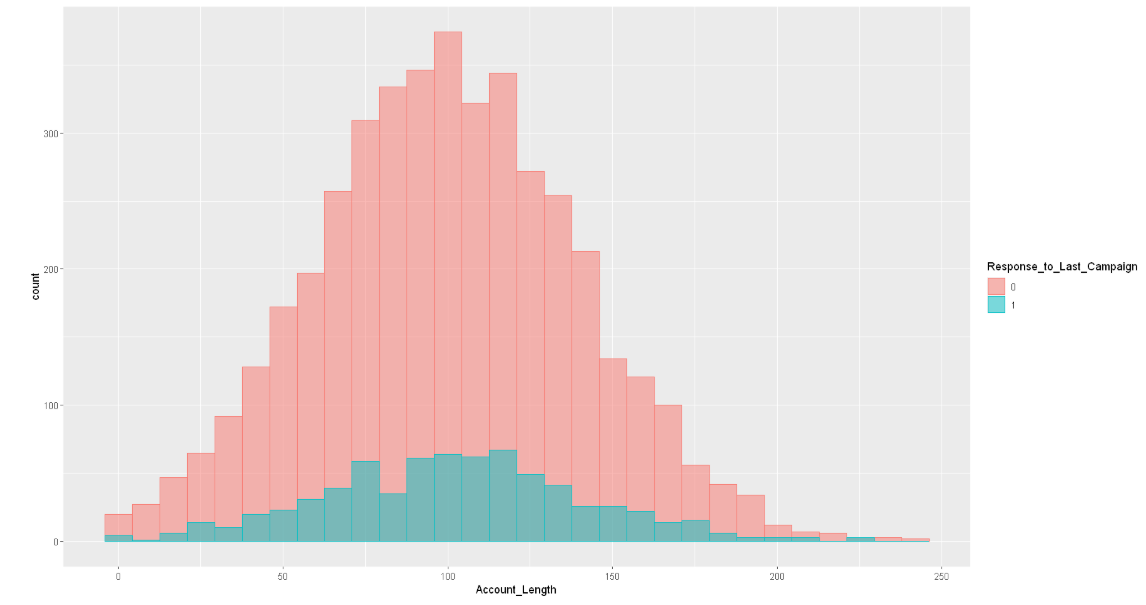


Only 14% of the customer targeted in previous campaign responded positively. Data is skewed towards negative responder which is common. We can deal with this using random over / under sampling to balance the training data set.

* + 1. Variables

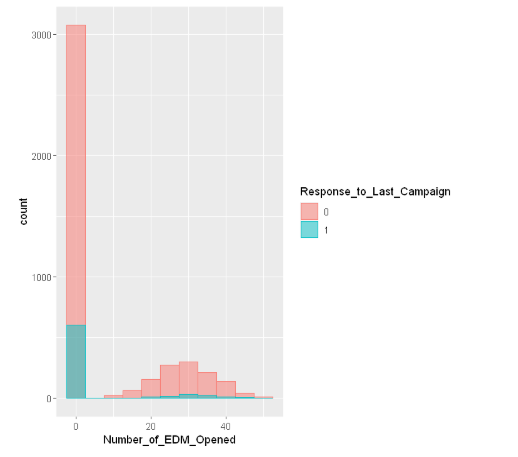
|  |  |
| --- | --- |
| **Variable Names** | **Useful for Model Building** |
| **ID** | **No** |
| **Account\_Length** | **Yes** |
| **Number\_of\_EDM\_Opened** | **Yes** |
| **total\_day\_charge** | **Yes** |
| **total\_eve\_charge** | **Yes** |
| **total\_night\_charge** | **Yes** |
| **total\_intl\_charge** | **Yes** |
| **Promotion\_Coupons\_Used** | **Yes** |
| **Competitor\_use** | **Yes** |
| **Call\_center** | **Yes** |
| **Response\_to\_Last\_Campaign** | **Target Variable (y)** |

* + 1. Exploratory Data Analysis
       - Account\_Length



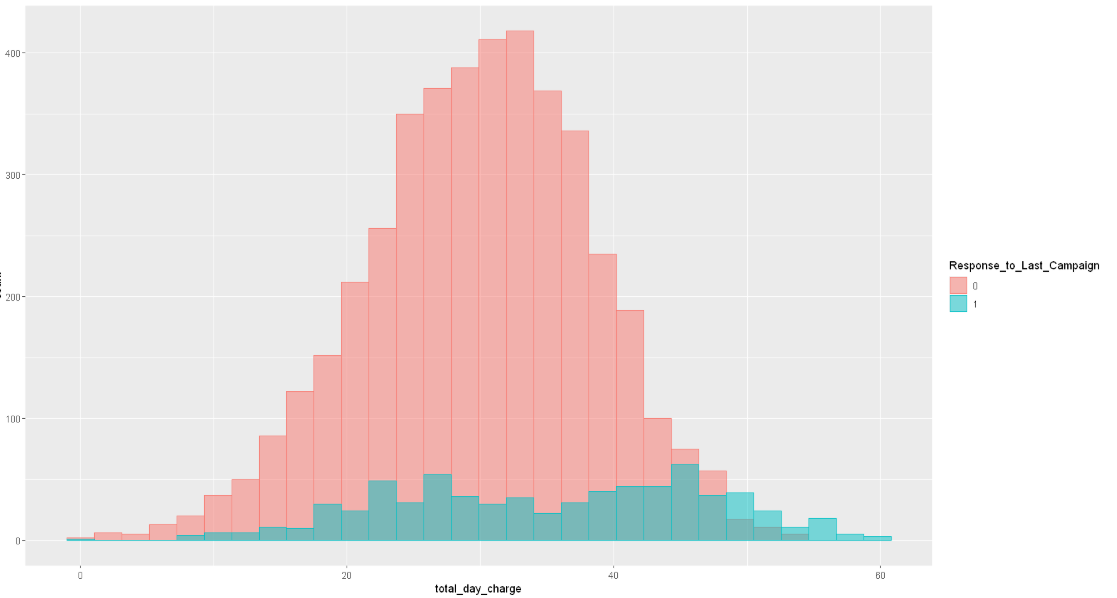
Those who are new to the company are not likely to respond positively to the campaign

* + - * Number\_of\_EDM\_opened



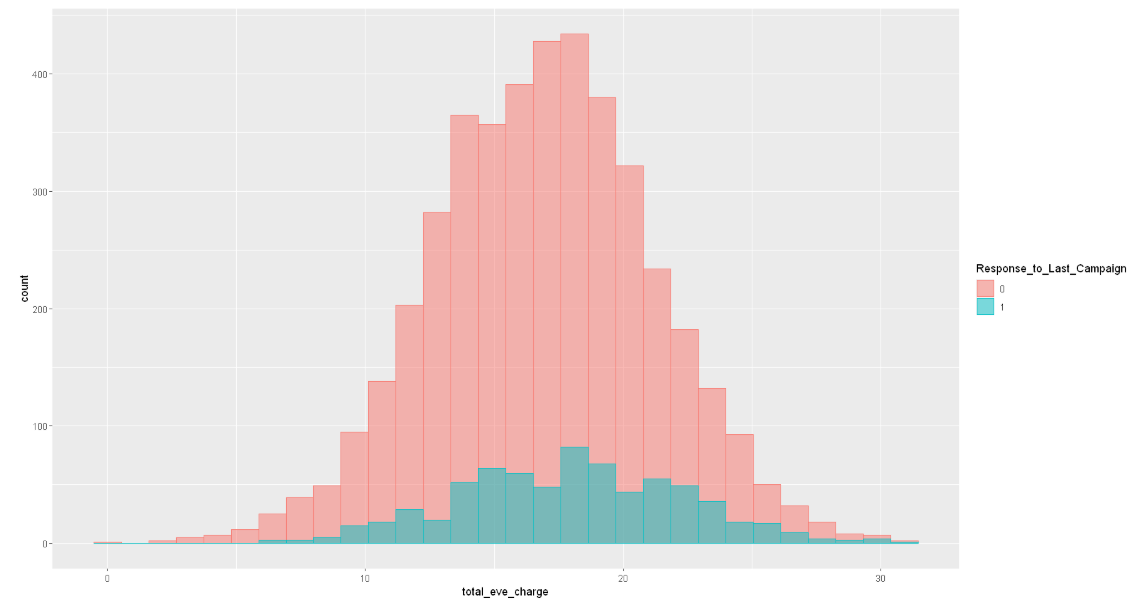
Majority of the customers do not open EDM and among those who responded positively to campaign, most also do not open EDM.

* + - * Total\_day\_charge

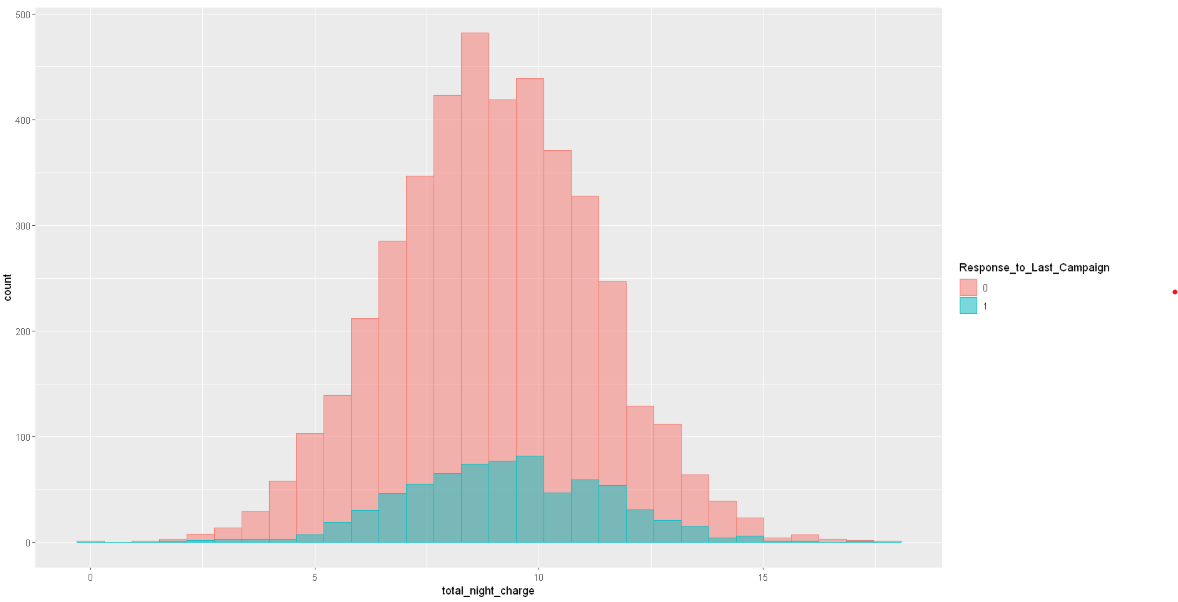


Those who have high total Day Charges are more likely to respond positively to the campaign. Probably the Value Plan offers more day call quota

* + - * Total\_eve\_charge

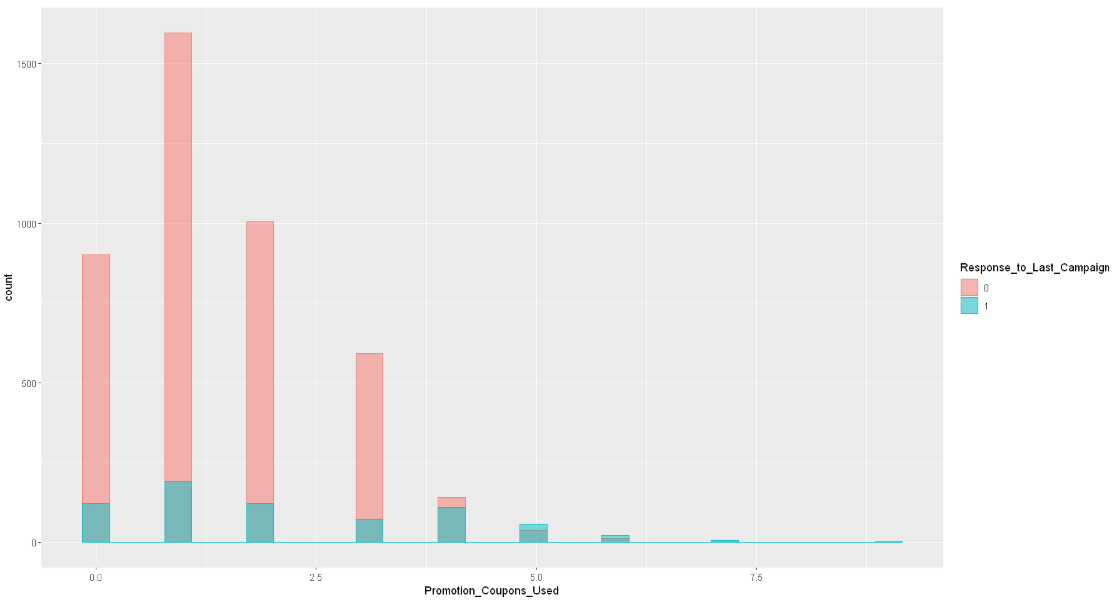


* + - * Total\_night\_charge



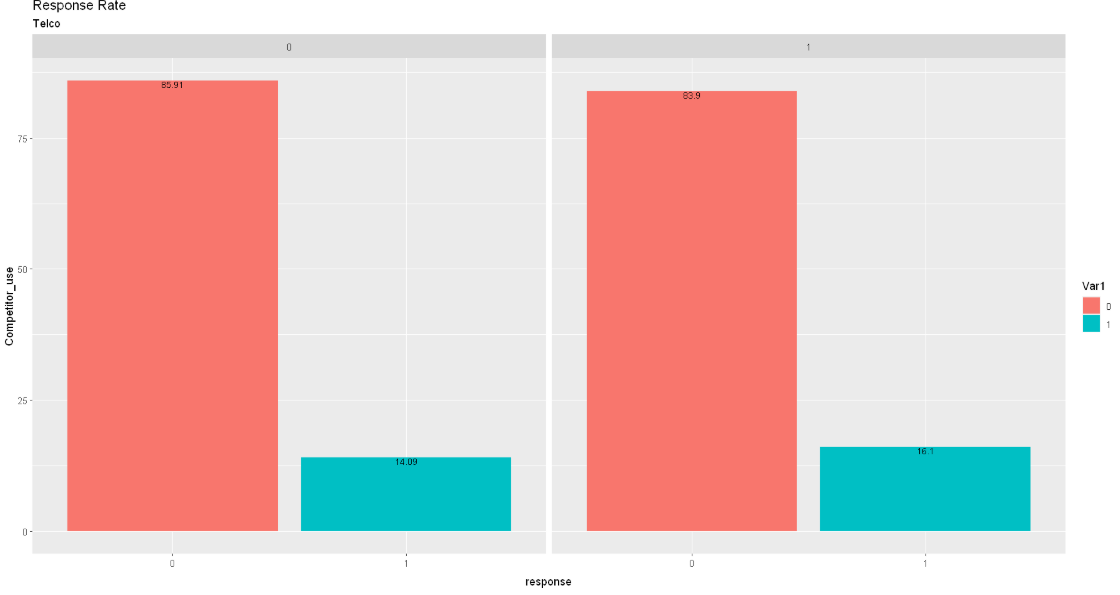
Those who have mid range total\_night\_charge are likely to respond positively to campaign.

* + - * Promotion\_Coupons\_used



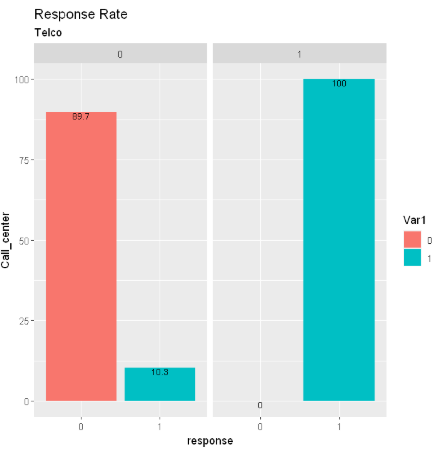
Those who use more promotion coupon are likely to respond positively to campaign

* + - * Connection with Competitor



Having connection with competitor does not seem to affect campaign response

* + - * Call\_center



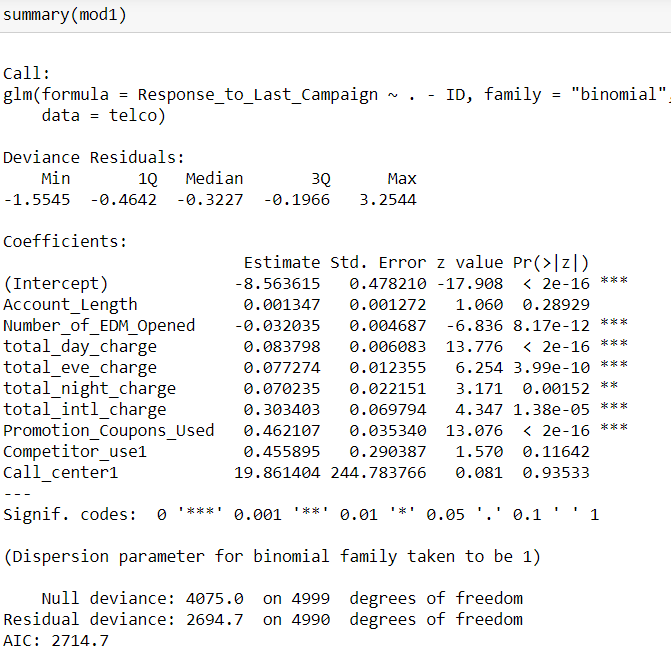
Call\_center (Called Customer Service Call Center in last 3 months) seem to make a big difference in the outcome of the campaign. We are not sure what was the purpose of call or what the customer service officer said but All those who have taken up the Value plan have called Customer Service Call Center in last 3 months. The result of this variable looks unusual and worth further investigation.

* + 1. Based on EDA above, the following variables seem to have most impact in identifying positive responder:
       - Call\_center
       - Promotion\_Coupon\_Used
       - Amount of Total\_day\_charge
  1. **Develop Analytical framework (Explain)**

Assuming the customers are on contract, we would recommend setting the observation period as x number of months before contract end date. In this case, would propose to set observation period to begin 5 months before contract end date for period of 4 months inclusive of 1 scoring month to run the scoring and 3 months for performance period to see if the targeted customer will take up the campaign offer.

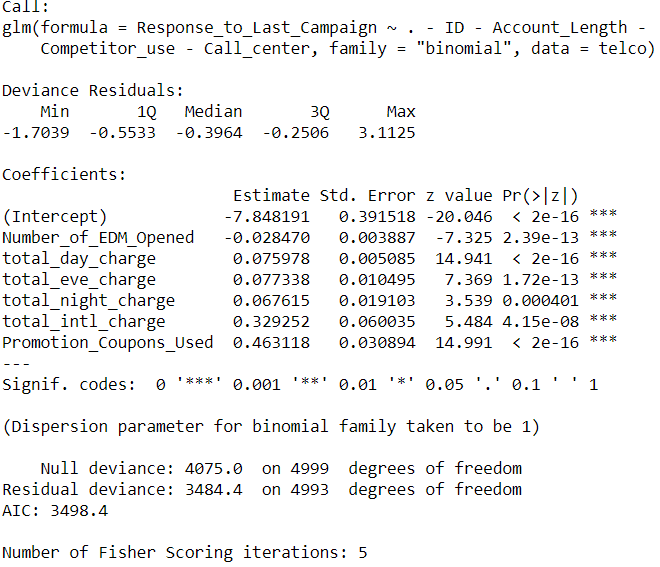
We would propose using Logistic Regression to identify which customers are likely to take up the Value plan and sort the likelihood in descending order. We will then based on that list target customers with high likelihood while balancing the cost incurred. If the campaign cost data per customer is available we will identify the right threshold to maximize benefit gained from True Positive and minimize the cost incurred on False Positive group of customers.

* 1. **Share the Results and Recommendation**
     1. First run of Logistic Regression



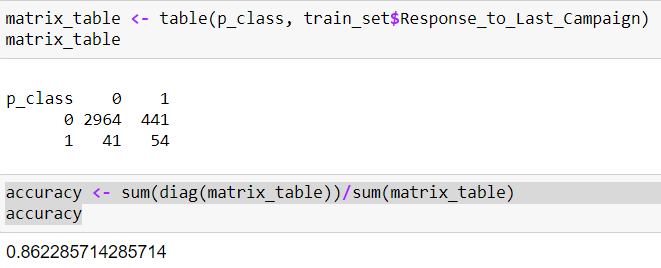
Account\_Length which measures the length of customer relationship with the company is not statistically significant. Similarly Competitor\_use and Call\_center are also not statistically significant.

* + 1. We drop the 3 variables and run another Logistic Regression Model. The result is

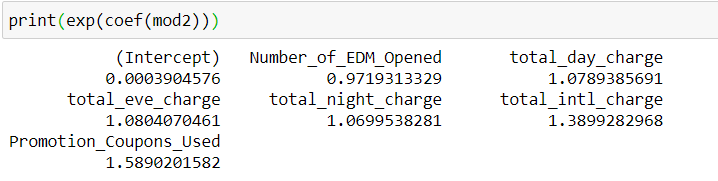


All the variables are statistically significant and we can see that 1 unit increase in total\_intl\_charge has 0.33 increase in likelihood of taking up the Value Plan and 1 unit increase in Promotion\_Coupons\_Used result in 0.46 increase in likelihood of taking up Value Plan.

Accuracy:



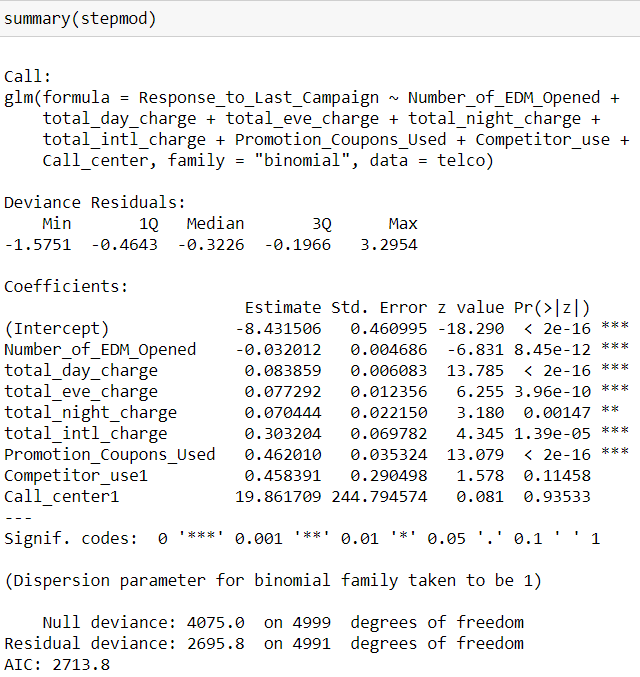
Accuracy is approximately 86% based on threshold of 0.5.



The coefficients of the equation shows that all the variables have positive impact to positive campaign response except Number\_of\_EDM\_Opened.

1 unit increase in Promotion\_Coupons\_Used will increase the likelihood of positive response to campaign by 1.59.

* + 1. Logistic Regression using stepAIC



It seems that Call\_center variable is deemed important in stepAIC although statistically not significant. Based on our exploration above, we noticed that those who responded positively to past campaign all of them have called customer service center in last 3 months which is rather unusual. Given the peculiarity and no significance from statistical point of view, we will omit this variable in model. During our data exploration, Competitor\_use does not show any contribution to positive response to past campaign. Based on the result, we recommend using logistic regression model in item ii above where all variables are statistically significant to predict the target customer for next campaign.

* 1. **Limitations with respect to data available and analysis (if any)**
     1. While the accuracy of our model based on threshold of 0.5 is quite high at 86%, the main metrics that is important for this campaign is not just accuracy but True Positive Rate or the ability of the model to correctly identify those customers who are going to respond positively to the campaign. As the original data shows low positive response to campaign, we will need to adjust the model threshold to maximize the benefit from capturing as many True Positive as possible while minimizing the number of False Positive (customers who are not likely to respond but targeted resulting in unnecessary cost).
     2. As we do not have the campaign cost per customer, we will not be able to calculate the payoff between True Positive vs False Positive and find optimal threshold.