In this project the dataset we have is from a marketing campaign implemented by a major banking institution.

The outcome variable is y which is basically telling us that whether or not a bank salesperson was able to get a client to sign up for a term deposit (and is labeled 0 for no, and 1 for yes).

So, we would use classification algorithms in order to help the bank management understand how can they maximize the clients for signing up for the term deposit.

**1st part:**

We have calculated the probability for any contacted client to set up the term deposit using the “intercept only logistic regression” model.

After which we use the intercept from the above results to calculate the y=1 values using the formula .

Finally, we will confirm if these values are correct by constructing cross table for y.

**2nd part:**

This part we will see assess the term deposit campaign against the other variables like the literacy of the clients (education). So, we created a logistics model in order to answer this question.

Next, we compare the Odds Ratio for the highest group of education (education =2) compared to the lowest group of education (education =0)

Finally, we will calculate the probability of the lowest education group (education = 0) signing up for the term deposit (y=1) in comparison to the probability of highest education group (education =2) signing up for the term deposit (y =1).

**3rd part:**

This part is very similar to the one done above we just need to change the variable from education to day. So, basically we are trying to figure out which day of the week yields the highest probability of the clients signing the term deposit (y =1).

**4th part:**

In this section we will try and build predictive model in such a way that we get a client with the highest probability of signing up the term deposit using the logistics model and all the variables against the outcome variable y.

**5th part:**

In this section we would like to increasing the prediction accuracy of our model using optimalcutoff() in logistics we would also use the classification tree method and see if the accuracy of the model is increased.

Understanding the dataset: (Copied from Kaggle :p)

Education: 0=lowest, 1=middle, 2=highest

Occupation: 0=unemployed/students/retired, 1=blue-collar, 2=white-collar

Day: 1=Monday, 2=Tuesday, 3=Wednesday, 4=Thursday, 5=Friday

Quarter: 0=Spring, 1=Summer, 2=Fall/Winter

housing: has housing loan? (categorical: 'no','yes','unknown')

contact: contact communication type (categorical: 'cellular','telephone')

duration: last contact duration, in seconds (numeric).

campaign: number of contacts performed during this campaign and for this client (numeric, includes last contact)

pdays: number of days that passed by after the client was last contacted from a previous campaign (numeric; 999 means client was not previously contacted)

previous: number of contacts performed before this campaign and for this client (numeric)

poutcome: outcome of the previous marketing campaign (categorical: 'failure','nonexistent','success')

y - has the client subscribed a term deposit? (binary: 'yes','no')