# Team Name :- Data Nerds

# Members:

Zenil Sanghvi

Abhik Naskar

Data Analytics

**Evaluation of data set:**

After carefully evaluating the given data set we found out that we have to predict binary classified values for the dataset.

**Algorithm for Data set:**

Since it is a classification problem we can use the following algorithms:

1. Logistic Regression
2. Neural Networks
3. Support Vector Machines
4. Random Forest

Out of all these algorithms we applied Logistic Regression according to our dataset because it was given a better accuracy than others in the test case.

**Why Logistic Regression?**

* It is more robust: the independent variables don’t have to be normally distributed, or have equal variance in each group
* It does not assume a linear relationship between the IV and DV
* It may handle nonlinear effects
* You can add explicit interaction and power terms
* The DV need not be normally distributed.
* There is no homogeneity of variance assumption.
* Normally distributed error terms are not assumed.
* It does not require that the independents be an interval.
* It does not require that the independents be unbounded.

**Prevention of overfitting:**

We tested our dataset on different values of alpha i.e. regularisation coefficient, so we created our learning curve which gave us the different values of accuracy for different values of alpha and from that we defined the optimum value of alpha to prevent overfitting.

Thanking You