**A short primer of underwriting in the credit industry -**

When an individual/corporation applies for a loan from a bank, their credit history undergoes a rigorous check to ensure their capability of replaying the loan.

The issuers have a set of model/s and rule/s in place which take information regarding their current financial standing, previous credit history and some other variables as input and output a metric which gives a measure of the risk that the issuer will potentially take on issuing the loan. The measure is generally in the form of a probability and is the risk that the person will default on their loan (called the probability of default) in the future.

Based on the amount of risk that the issuer is willing to take (plus some other factors) they decide on a cutoff of that score and use it to take a decision regarding whether to pass the loan or not. This is a way of managing credit risk. The whole process collectively is referred to as underwriting.

**Overview of the problem**

The text files contain complete loan data for all loans issued by the company from 2007-2015. The data contains the indicator of default, payment information, credit history, etc.

You will have use the training data to build models/analytical solution and finally apply it to test data to measure the performance and robustness of the models.

**Data and Problem Details**

Objective : You have to build a data model to predict the probability of default, and choose a cut-off based on what you feel is suitable. Alternatively you can also use a modelling technique which gives binary output.

You have to do the following :

Based on the data that is available during loan application, build a model to predict default in the future. This will help the company in deciding whether or not to pass the loan.

Also note that the data contains defaulters, successful payers and customers who were current during that time. To simplify the problem, customers under 'current' status have been considered as non-defaulters in the dataset.

Steps to be followed (tentative time required) :

1) Understand the problem and objective (1 hour)

2) Understand the data and develop some business sense. (4-5 hours)

3) EDA, segmentation (if you think is required in this case). (5-6 Hours)

4) Data Cleaning (4-5 Hours)

5) Feature engineering (4-5 Hours)

6) Model Building (try various techniques and at the end justify why you chose a particular technique over others) (3-4 hours)

7) Testing and validating(3-4 hours)

8) Final results, recommendations and plots/visualizations. (4-5 hours)

9) BONUS : Any other insights or recommendations that you can give from the data which will help the business. (subjective)