C2-T3 Project

1. Which attributes in the data can we deem to be statistically significant to the problem at hand?

The attributes that are important to solve the problem are the Credit Limit ‘**LIMIT\_BAL'** and the **‘default payment next month’**

1. What concrete information can we derive from the data we have?

* As it can be seen from the graph the Credit limit balance for the default is up to $140,000 which will put the bank in to risk as number of Borrowers increases. (Graph is taken from C2-T2 Lab)

Chart

Description automatically generated

1. How do you ensure that customers can/will pay their loans?

* Counting how many times (by using ID) the customer has defaulted will help to ensure that if they will pay their loans. (Graph is taken from C2-T2 Lab)

Graphical user interface

Description automatically generated with low confidence

1. Can we approve customers with high certainty?

* It is possible to approve customers with high certainty after evaluating their credit default history.

1. After building the three models the three Regression models ('Random Forest Regressor', 'Linear Regression', 'Support Vector Regression') and evaluating the accuracy for the prediction of Balance Limit. I recommend that Random Forest Regressor will be the best to use.

Graphical user interface, text, application, email

Description automatically generated