* **Delinquency:** The meaning of delinquency is a condition that arises when an activity does not occur at its expected date.
* For the given dataset, a consumer is believed to be delinquent if he deviates from the path of paying back the loaned amount within next 5 days.
* The shape of dataset is (209593, 37)
* Quick data exploration was done to find what different data types were present.
* Checked for missing values. No missing data was found.
* The column **"Unnamed: 0"** and "**msisdn**" have no significance in final prediction, so droped it.
* **Feature selection:** Now the next step is feature selection, collinear features are removed.
* “pcircle” and “pdate” is not contributing in the final prediction, these two features are droped.
* Now dataset is split into training and test set (80:20)
* Feature scaling is done on X\_train and X\_test sets using **StandardScaler()**.
* Logistic Regression and Random Forest models are chosen and these models are trained on training data and prediction is done on test data.
* Obtained test accuracy -
  1. Logistic Regression 87.49%
  2. Randon Forest 87.58

With the help of **classification\_report() ,** precision, recall, and f1-score are calucalated but not got descent value.

I feel that to get a descent classification report, we need to do class balancing (smote method can be used for this).