**Notes**

* **Drop** ID - Done
* We have setup view to check every column. - Done
* To get which column have missing data – df.info () - Done
* Looking each column and making judgements. - Done
* Customer\_ID can be used for filling up lost data in test. - Done
* Predicted class is multiple – Problem is multi-class classification. Also, it is in string need to convert it to numeric. - Done
* Month will be converted to one-hot encoding.

**We can drop Month too – In training set we have Jan-Aug Month however for testing set we have Sept-Dec months. So, feature is not overlapping and no use of this column. - Done**

* Name is not important – Will **drop** this. - **Done**
* Age is wrong. Need to check and rework on it. Age is in string change it to numeric.

Age Logic – Remove \_, extra symbols using regex get only numeric.

Then need to fill right age where deviation is present. - **Done**

* SSN can be **drop.**
* Occupation is important. We have a lot of \_\_\_ need to fill them up too. And convert this to one-hot encoding too. - **Done**
* Annual Income have \_ and symbols. Need to clean it and then convert to numeric - **Done**
* Have to fill Monthy\_Inhand\_Salary – Imputation method - Median Used - **Done**
* Num\_Bank\_Accounts which has less than < 0, should be set to 0. And anomaly in data so replace it with median. Median used - **Done**
* Num\_Credit\_Card has anomaly in data should replace it with median value. **Done**
* Interest\_Rate has anomaly in data should replace it with median value. **Done**
* Num\_loan – String clean and then remove extra characters and converting to int. Then change the data per customer basis. It has anomaly in data should replace it with median value. **Done**
* **Type\_Loan** where NaN change it to None. And Then Extract it and convert to one-hot.
* Delay\_from\_due\_date which is < 0 turn it to 0 - **Done**
* Num\_of\_Delayed\_Payment to be filled with median. If all is NaN then drop - **Mean**
* Change\_credit\_limit– String clean and then remove extra characters and converting to int. Then change the data per customer basis. - **Done**
* Num\_Credit\_Inquiries is wrong need to change it. -
* Credit\_Mix has \_ need to change it to max of column value and it is string need to change to int.
* Outstanding\_Debt is having \_ need to change.
* Credit\_History\_Age needs to convert to months
* Payment\_of\_Min\_Amount has value NM which needs to be No or depending on Max of col
* Payment\_Behaviour Confused with this feature
* Monthy Balance is String