Step1 -> Create a target -> churn or non-churn -> remove 9 month columns

Step2 -> Handle missing values (

Step3 ->High value customers (removing lot of rows)

Step4 -> EDA

Step5 -> Modeling

Good performance model -> PCA-> Train-test split -> SMOTE -> RF

Good interpretation model -> Train-test split -> SMOTE -> Logistic regression/RF

total_ic_mou_9 -> incoming and outgoing usage

total_og_mou_9

vol_2g_mb_9

vol_3g_mb_9

Total_usage = total_ic_mou_9 + total_og_mou_9 + vol_2g_mb_9 + vol_3g_mb_9

Total_usage=0 -> churn=1, else churn =0

Target = 1/ 0 -> sum of these four columns

For every customer:

total_value_6 = data_recharge + call_recharge(not given)

avg_rechaeg_amt*count

total_value_7 = data_recharge + call_recharge

total_value = avg(total_value_6, total_value_7)

Take >= 70th percentile customers

HV_June = **Total_amt_recharge_data_6(Data)** + **Total_amt_recharge_data_6**

HV_June = **Total_amt_recharge_data_7(Data)** + **Total_amt_recharge_data_7**

Average = avg(HV_june, HV_July) -> 70 percentile

total_recharge_data_6 = number of data recharges = 5

avg_rech_amt_data_6 = 200

Total_amt_recharge_data_6 = total_recharge_data_6 * avg_rech_amt_data_6

Total_amt_recharge_data_6(Data) = 5* 200 = 1000