**Observations From Data:**

Dataset

1. There are 36 variables out of which 32 are Boolean and 4 numerical variables and a total of 36000 observations.
2. Dataset has no missing values.
3. Dataset has 73.8% Duplicated values.
4. F and K are highly skewed.
5. C1 and C2 are highly corelated.
6. Numerical variables D and E have 25% and 83.4% zeros respectively.
7. Dataset is highly biased as majority of the observations falls in specific class.
8. Label is divided equally i.e. both 0,1 has 18,000 observations

Dividing dataset on basis of label i.e. churn or not churn:

Not Churn (label==0):

1. 37 variables with 18000 observations.
2. Numerical variables D and E have 26.6% and 83.6% zeros respectively.
3. J6 has constant value 0.
4. F and K are highly skewed.
5. C1 and C2 are highly corelated.

Churn (Label==1):

1. 37 variables with 18000 observations
2. Numerical variables D and E have 26.5% and 82.9%
3. F is highly skewed
4. C1 and C2 are highly corelated.