**Problem Statement:**

This case study aims to identify patterns which indicate if a client has difficulty paying their installments which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc. This will ensure that the consumers capable of repaying the loan are not rejected. Identification of such applicants using EDA is the aim of this case study.

To Achieve the model following methods were followed:

**1. Cleaning The Data:**

**a.insights of data/drop unnecessary columns:**

In the 1st dataset we have 307511 rows and 122 columns.The columns with more than **50%** of null values were dropped.so **41 columns** has dropped, after that The columns with more than **15%** of null values were dropped.so **10 columns** has dropped.

In another dataset we have 1413701 rows and 120 rows.The columns with more than **50%** of null values were dropped.so **4 columns** has dropped, after that The columns with more than **15%** of null values were dropped.so **10 columns** has dropped.

Many variables are having dataframes with single value which is leading to imbalance

so we will drop the following column

WEEKDAY\_APPR\_PROCESS\_START','HOUR\_APPR\_PROCESS\_START','FLAG\_LAST\_APPL\_PER\_CONTRACT','NFLAG\_LAST\_APPL\_IN\_DAY'

**b.replace negative value to positive value**

DAYS\_BIRTH

DAYS\_EMPLOYED

DAYS\_REGISTRATION

DAYS\_ID\_PUBLISH

In dataset we see dates have negative values so we replace that values to positive values.

**2. EDA:**

Exploratory Data Analysis was done to check categorical variables .

We can find that some numerical variables consisted of very high values as

compared to their respective means. That's why we have created charts using

boxplot to understand the patterns. We have observed that the outliers are very high

and we need to treat it. That was the reason we have retained 99% quantile of the

data and removed the max value from it.

**a.checking outlier:**

we observed the outliers which is very high in AMT\_INCOME\_TOTAL,AMT\_CREDIT,AMT\_ANNUITY so we remove that highest outlier and showing in the **boxplot.**

**3.conclusion-**

1)Banks should focus less on income type ‘Working’ as they are having most number of unsuccessful payments.

2)Banks should focus more on contract type ’pensioner’ with housing ‘type other than ‘Co-op apartment’ for successful payments.

3)Also with loan purpose ‘Repair’ is having a higher number of unsuccessful payments on time.

4)Get as much as clients from housing type house/apartment ‘With parents’ as they are having least number of unsuccessful payments.