```
In [25]: import pandas_profiling
         import pandas as pd
         import numpy as np
```

```
In [23]: df=pd.read_csv("data (2).csv")
         df
```

Out[23]:

| | SK_ID_CURR | TARGET | NAME_CONTRACT_TYPE | GENDER | Car | House | CNT_CHILDREN | AMT_INCOME_TOTAL | AMT_CREDIT | AMT_GOODS_ |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 100002 | 1 | Cash loans | M | N | Y | 0 | 202500.0 | 406597.5 | 35 |
| 1 | 100003 | 0 | Cash loans | F | N | N | 0 | 270000.0 | 1293502.5 | 112 |
| 2 | 100004 | 0 | Revolving loans | M | Y | Y | 0 | 67500.0 | 135000.0 | 13 |
| 3 | 100006 | 0 | Cash loans | F | N | Y | 0 | 135000.0 | 312682.5 | 29 |
| 4 | 100007 | 0 | Cash loans | M | N | Y | 0 | 121500.0 | 513000.0 | 51 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 99995 | 216086 | 0 | Cash loans | F | N | Y | 1 | 157500.0 | 755190.0 | 67 |
| 99996 | 216087 | 0 | Cash loans | F | N | Y | 1 | 225000.0 | 284400.0 | 22 |
| 99997 | 216088 | 0 | Cash loans | F | Y | Y | 0 | 135000.0 | 1262583.0 | 110 |
| 99998 | 216089 | 0 | Cash loans | F | Y | N | 0 | 135000.0 | 225000.0 | 22 |
| 99999 | 216090 | 0 | Revolving loans | M | Y | Y | 0 | 202500.0 | 337500.0 | 33 |

100000 rows × 24 columns

## Conclusions

After analysing the datasets, there are few attributes of a client with which the bank would be able to identify if they will repay the loan or not. The analysis is consised as below with the contributing factors and categorization:

### A. Decisive Factor whether an applicant will be Repayer:

NAME_EDUCATION_TYPE: Academic degree has less defaults.

NAME_INCOME_TYPE: Student and Businessmen have no defaults.

REGION_RATING_CLIENT: RATING 1 is safer.

ORGANIZATION_TYPE: Clients with Trade Type 4 and 5 and Industry type 8 have defaulted less than 3%

DAYS_BIRTH: People above age of 50 have low probability of defaulting

DAYS_EMPLOYED: Clients with 40+ year experience having less than 1% default rate

AMT_INCOME_TOTAL:Applicant with Income more than 700,000 are less likely to default

NAME_CASH_LOAN_PURPOSE: Loans bought for Hobby, Buying garage are being repayed mostly.

CNT_CHILDREN: People with zero to two children tend to repay the loans.

### B. Decisive Factor whether an applicant will be Defaulter:

CODE_GENDER: Men are at relatively higher default rate

NAME_FAMILY_STATUS : People who have civil marriage or who are single default a lot.

NAME_EDUCATION_TYPE: People with Lower Secondary & Secondary education

NAME_INCOME_TYPE: Clients who are either at Maternity leave OR Unemployed default a lot.

REGION_RATING_CLIENT: People who live in Rating 3 has highest defaults.

OCCUPATION_TYPE: Avoid Low-skill Laborers, Drivers and Waiters/barmen staff, Security staff, Laborers and Cooking staff as their default rate is huge.

ORGANIZATION_TYPE: Organizations with highest percent of loans not repaid are Transport: type 3 (16%), Industry: type 13 (13.5%), Industry: type 8 (12.5%) and Restaurant (less than 12%). Self-employed people have relative high defaulting rate, and thus should be avoided to be approved for loan or provide loan with higher interest rate to mitigate the risk of defaulting.

DAYS_BIRTH: Avoid young people who are in age group of 20-40 as they have higher probability of defaulting

DAYS_EMPLOYED: People who have less than 5 years of employment have high default rate.

CNT_CHILDREN & CNT_FAM_MEMBERS: Client who have children equal to or more than 9 default 100% and hence their applications are to be rejected.

AMT_GOODS_PRICE: When the credit amount goes beyond 3lakhs, there is an increase in defaulters.

### C. Factors that Loan can be given on Condition of High Interest rate to mitigate any default risk leading to business loss:

NAME_HOUSING_TYPE:

High number of loan applications are from the category of people who live in Rented apartments & living with parents and hence offering the loan would mitigate the loss if any of those default.

AMT_CREDIT:

People who get loan for 3-6 Lakhs tend to default more than others and hence having higher interest specifically for this credit range would be ideal.

People who got loan for 3-5 Lakhs tend to default more than others and hence having higher interest specifically for this credit range would be ideal.

AMT_INCOME:

Since 90% of the applications have Income total less than 3Lakhs and they have high probability of defaulting, they could be offered loan with higher interest compared to other income category.

CNT_CHILDREN & CNT_FAM_MEMBERS:

Clients who have 4 to 8 children has a very high default rate and hence higher interest should be imposed on their loans.

NAME_CASH_LOAN_PURPOSE:

Loan taken for the purpose of Repairs seems to have highest default rate. A very high number applications have been rejected by bank or refused by client in previous applications as well which has purpose as repair or other. This shows that purpose repair is taken as high risk by bank and either they are rejected, or bank offers very high loan interest rate which is not feasible by the clients, thus they refuse the loan. The same approach could be followed in future as well.

> ### D. Suggestions:

90% of the previously cancelled client have actually repayed the loan. Record the reason for cancellation which might help the bank to determine and negotiate terms with these repaying customers in future for increase business opportunity. 88% of the clients who were refused by bank for loan earlier have now turned into a repaying client. Hence documenting the reason for rejection could mitigate the business loss and these clients could be contacted for further loans.

Insights

> ### NAME_CONTRACT_TYPE :

Most of the clients have applied for Cash Loan while very small proportion have applied for Revolving loan for both Defaulters as well as Non-defaulters.

> ### NAME_TYPE_SUIT :

Most of the clients were accompanied while applying for the loan.And with few clients a family member was accompanying for both Defaulters and Non-Defaulters. But who was accompanying client while applying for the loan doesn't impact on the default.Also both the populations have same proportions. NAME_INCOME_TYPE:

Clients who applied for loans were getting income by Working,Commercial associate and Pensioner are more likely to apply for the loan, highest being the Working class category . Businessman, students and Unemployedless likely to apply for loan . Working category have high risk to default. State Servant is at Minimal risk to default. NAME_EDUCATION_TYPE:

Clients having education Secondary or Secondary Special are more likey to apply for the loan. Clients having education Secondary or Secondary Special have higher risk to default.Other education types have minimal risk. NAME_FAMILY_STATUS :

Married Clients seems to be applied most for the loan compared to others for both Defaulters and Non-Defaulters. In case of Defaulters,Clients having single relationship are less risky In case of Defaulters, Widows shows Minimal risk. NAME_HOUSING_TYPE: From the bar chart, it is clear that Most of the clients own a house or living in a apartment for both Defaulters and Non-Defaulters. OCCUPATION_TYPE:

Pensioners have applied the most for the loan in case of Defaulters and Non-Defaulters. Pensioner being highest followed by laborers have high risk to default. WEEKDAY_APPR_PROCESS_START:

There is no considerable difference in days for both Defaulters and Non-defaulters. AMT_INCOME_TYPE:

Clients having Medium salary range are more likely to apply for the loan for both Defaulters and Non-defaulters. Clients having lowandmedium income are at high risk to default.

AMT_CREDIT_TYPE:

Most of the clients applied for Medium Credit Amount of the loan for both Defaulters and Non-defaulters. Clients applying for high and low credit are at high risk of default.

In [ ]: