CREDIT EDA ASSIGNMENT

Submitted by:
Prachi Garg

BUSINESS OBJECTIVE

This case study aims to identify patterns which indicate if a client has difficulty paying their instalments 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.

In other words, the company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

PROBLEM STATEMENT

1. 1. *'application_data.csv'* contains all the information of the client at the time of application. The data is about whether a **client has payment difficulties.**
2. *'previous_application.csv'* contains information about the client's previous loan data. It contains the data on whether the previous application had been **Approved, Cancelled, Refused or Unused offer.**

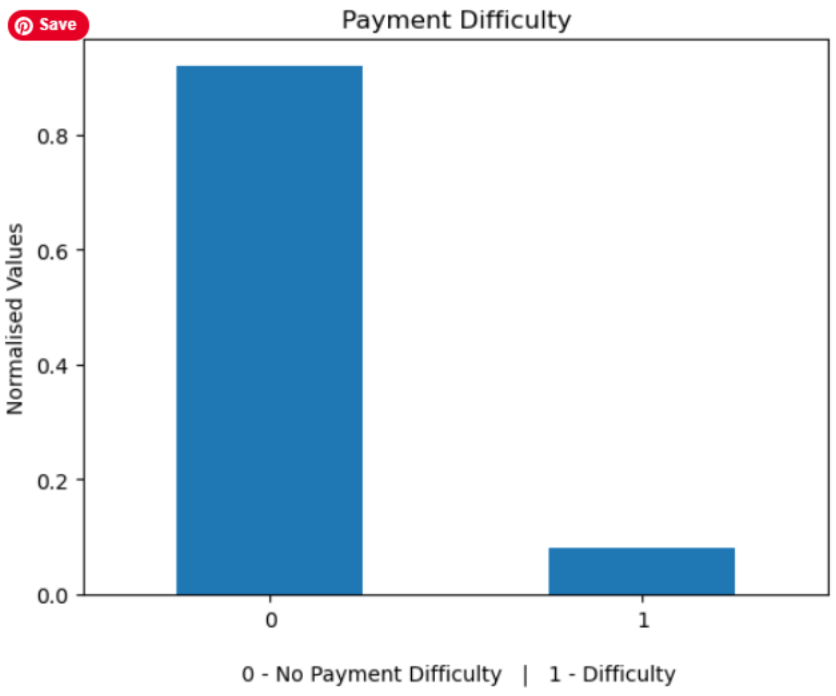
• The client with payment difficulties: he/she had late payment more than X days on at least one of the first Y instalments of the loan in our sample. (in our analysis, it is mentioned as target =1)

• All other cases: All other cases when the payment is paid on time. (in our analysis, it is mentioned as target =0)

METHODOLOGY

1. Data Sourcing _(given in assignment)
2. Data loading and Data Cleaning (details in the workbook):
   I. Fixing the rows and columns
   II. Imputing and Removing Missing columns
   III. Handling Outliers
3. Univariate Analysis (details in the workbook): :
   I. Categorical – Numerical Analysis
   II. Numerical- Numerical Analysis
   III. Categorical-Categorical Analysis
4. Bivariate and Multivariate Analysis (details in the workbook): :
   I. Numeric – Numeric Analysis
   II. Correlation
   III. Numerical – Categorical Analysis
   IV. Categorical – Categorical Analysis

PAYMENT DIFFICULTY



Inference : There is imbalance in the data as the number of people with payment difficulty is much less compared to number of people with no payment difficulty

OUTLIER ANALYSIS

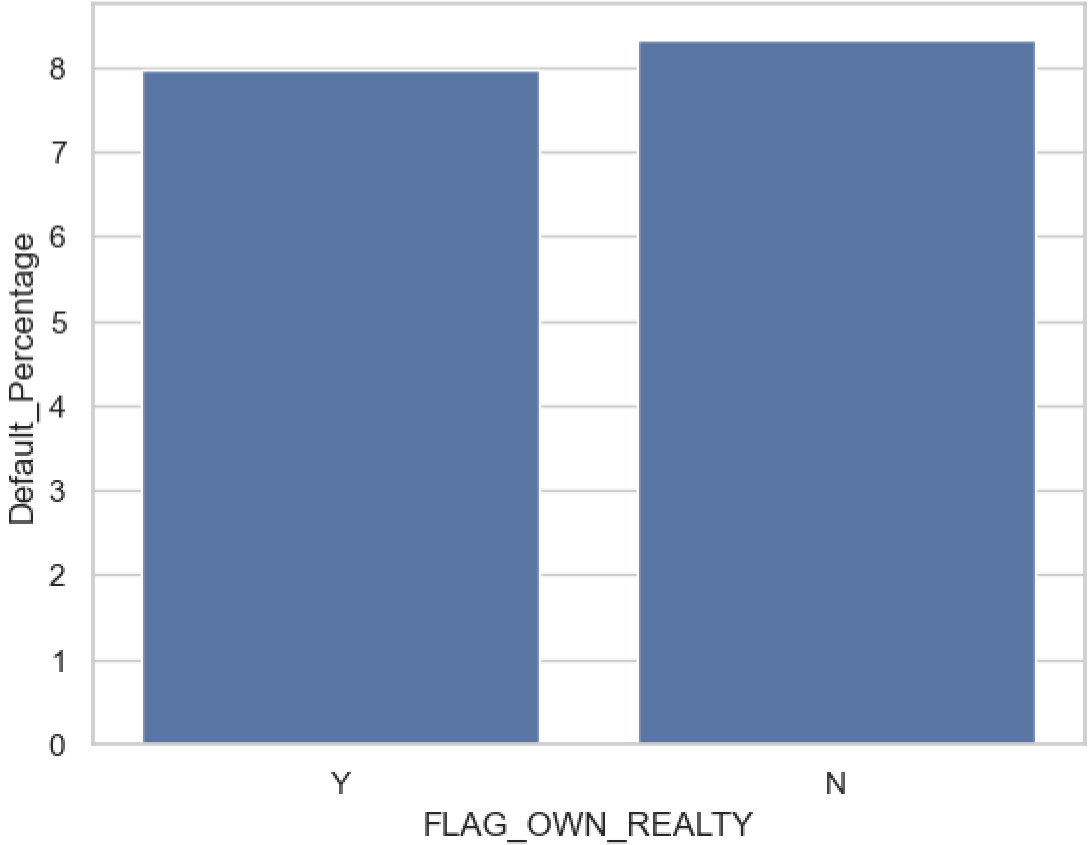
[87]:

	count	mean	std	min	25%	50%	75%	max
CNT_FAM_MEMBERS	307509.0	2.152665	0.910682	1.0	2.0	2.0	3.0	20.0
DAYS_EMPLOYED	307511.0	63815.045904	141275.766519	-17912.0	-2760.0	-1213.0	-289.0	365243.0
AMT_INCOME_TOTAL	307511.0	168797.919297	237123.146279	25650.0	112500.0	147150.0	202500.0	117000000.0

1. The outliers are handled in the analysis with values replaced with quartile ranges

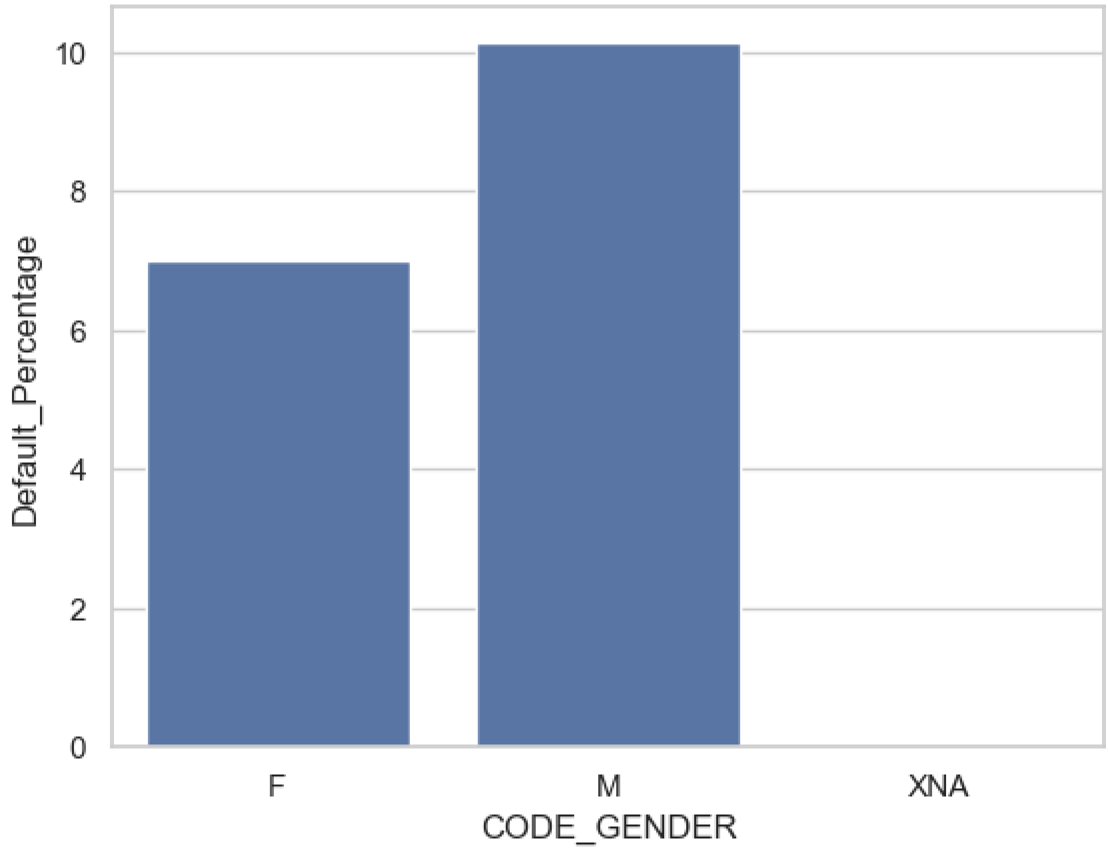
FLAG_OWN_REALTY

#From above graph, we can see that the number of non-payers of loan i.e., defaulters are very close almost equal to 9%. It is difficult to decide a target based on this metric.



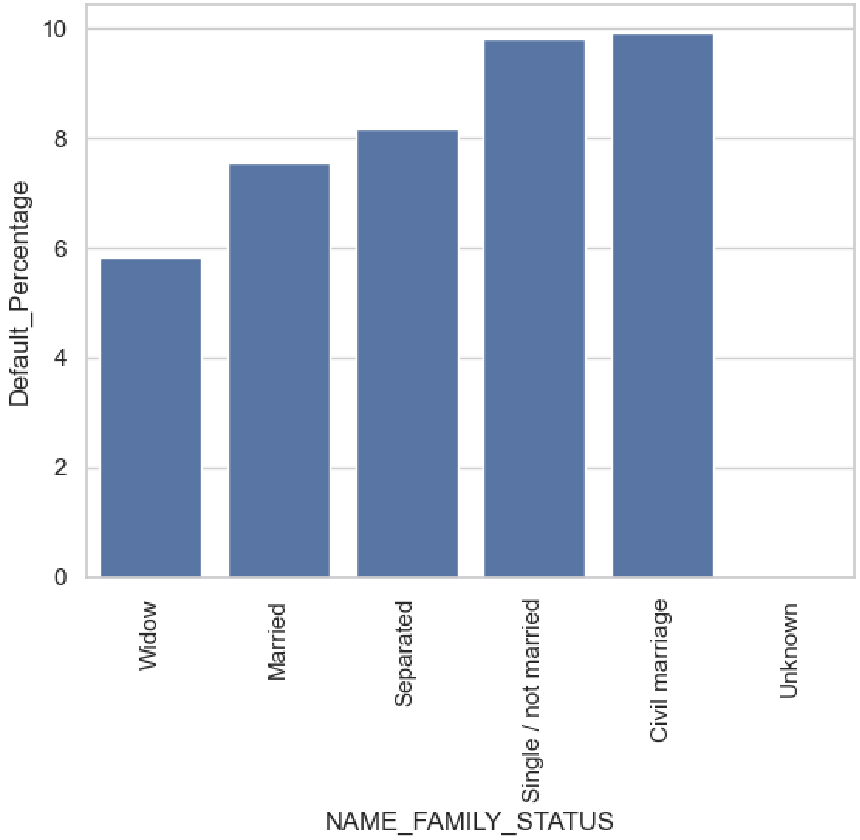
CODE_GENDER

#So, from above plots and data we can cleary see that the Female clients are a better TARGET as compared to the Male clients. Observing the percent of defaulted credits, male client have a higher chance of not returning their loans [10.14%], compared to the female clients [7%].



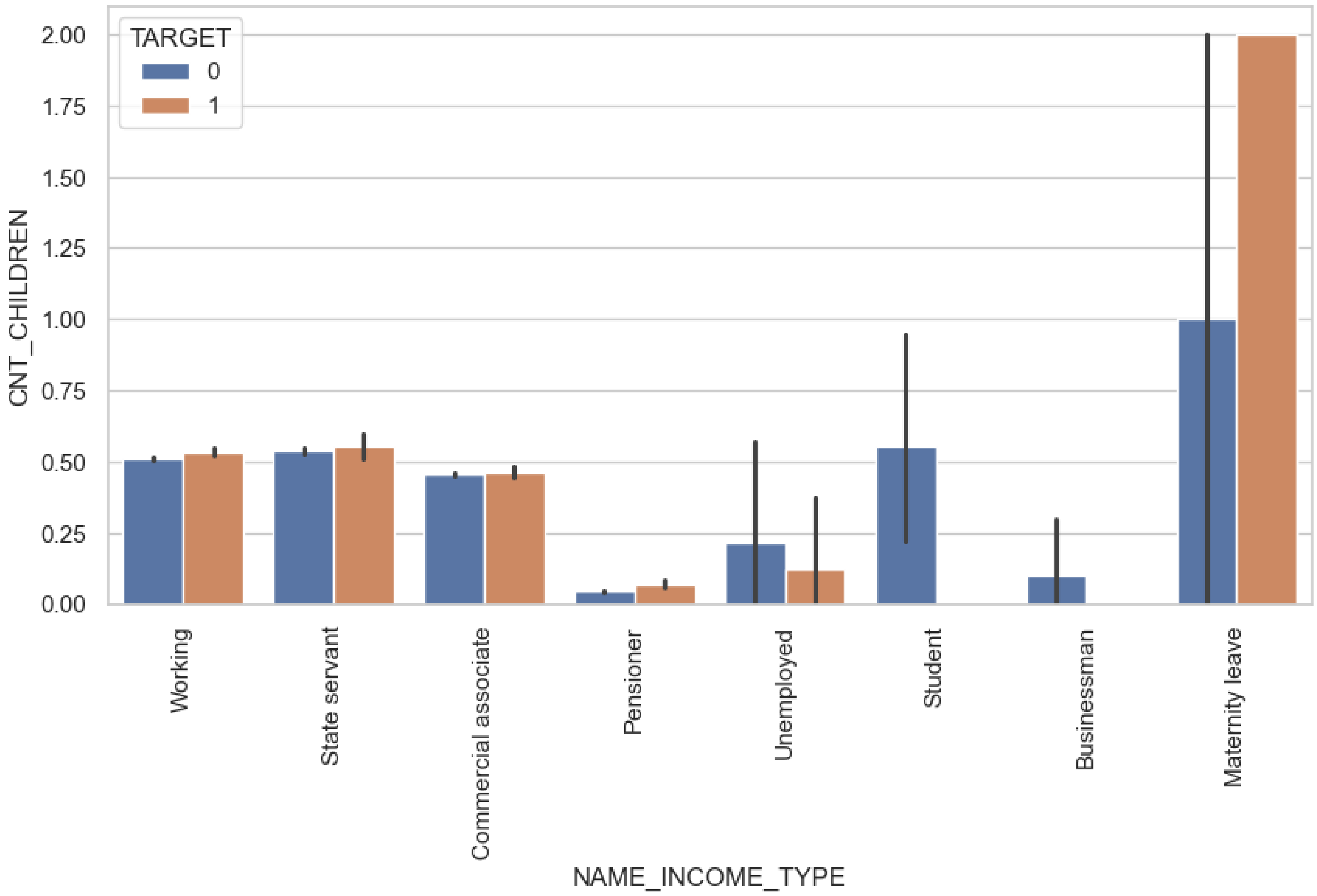
NAME_FAMILY_STATUS

#From above graph we can say that the percentage of non-repayment of loan is at highest for civil marriage and is lowest for widows, which is interesting to see because you expect widows to not payback their loans but it is the opposite here.

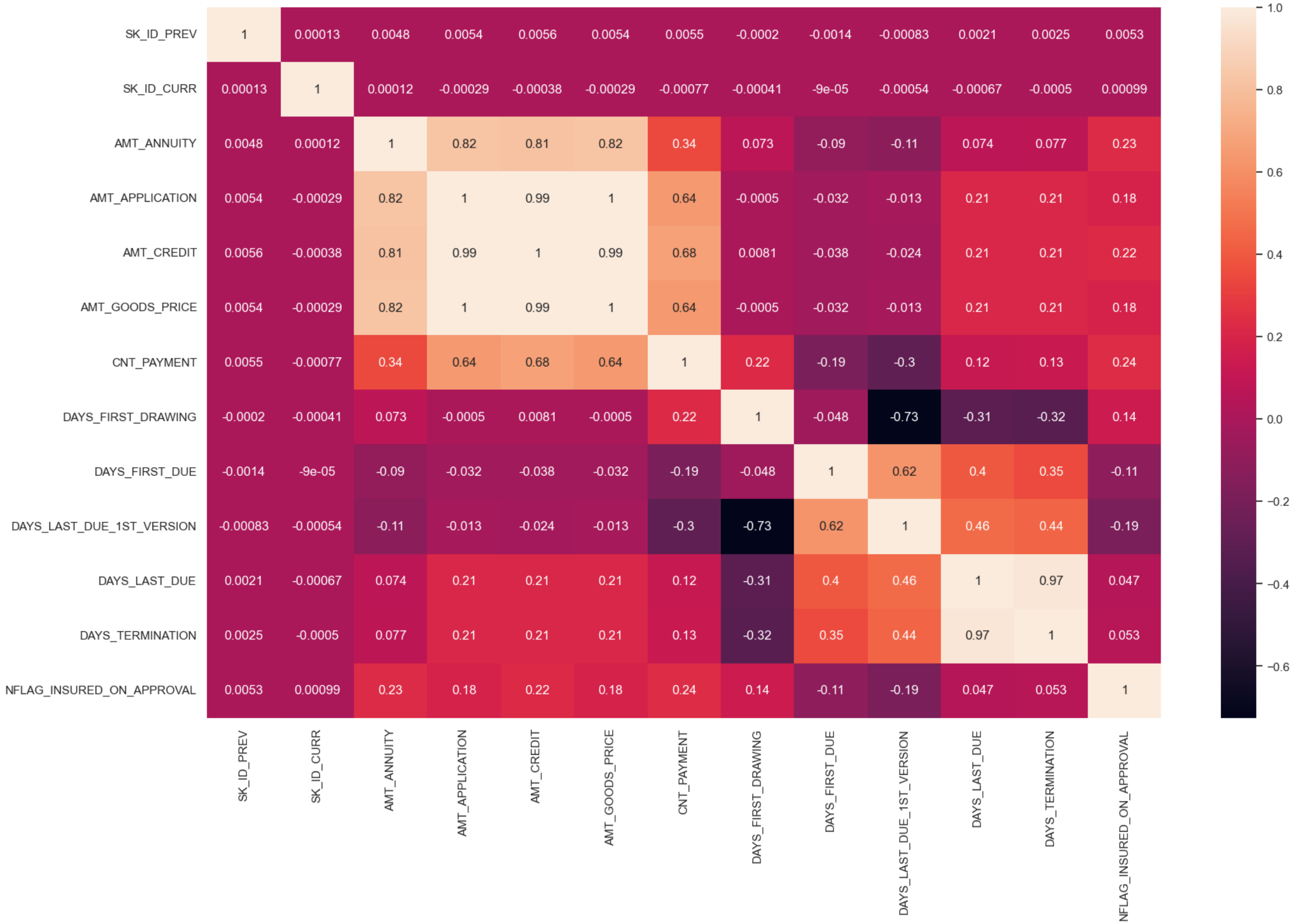


MATERNITY_LEAVE

#People who geting income via Maternity Leave tends to be more Defaulter when they have more children



HEATMAP FOR PREVIOUS_APPLICATION



#AMT_ANNUITY,AMT_CREDIT,AMT_APPLICATION and AMT_GOODS_PRICE have high correlation

MERGED FILES ANALYSIS – HIGHEST CORRELATION AND LEAST CORRELATED FACTORS

```
Correlation.head(10)["TARGET"][1:]
```

```
SK_ID_CURR                   -0.000344
AMT_ANNUITY_x                -0.026347
AMT_APPLICATION              -0.011185
AMT_CREDIT_x                 -0.007283
AMT_GOODS_PRICE_x            -0.011185
CNT_PAYMENT                   0.019037
DAYS_FIRST_DRAWING           -0.047421
DAYS_FIRST_DUE               -0.005020
DAYS_LAST_DUE_1ST_VERSION     0.026592
Name: TARGET, dtype: float64
```

```
Correlation.tail(10)["TARGET"][1:]
```

```
FLAG_DOCUMENT_20              -0.000924
FLAG_DOCUMENT_21              -0.000238
AMT_REQ_CREDIT_BUREAU_HOUR     0.001883
AMT_REQ_CREDIT_BUREAU_DAY      0.003382
AMT_REQ_CREDIT_BUREAU_WEEK    -0.002866
AMT_REQ_CREDIT_BUREAU_MON     -0.010276
AMT_REQ_CREDIT_BUREAU_QRT     -0.005820
AMT_REQ_CREDIT_BUREAU_YEAR     0.013803
AGE                           -0.068925
Name: TARGET, dtype: float64
```