

PARTH GUPTA
LOAN DEFaulTER
EDA





AGENDA

PROBLEM STATEMENT

DATASETS

ASSUMPTION

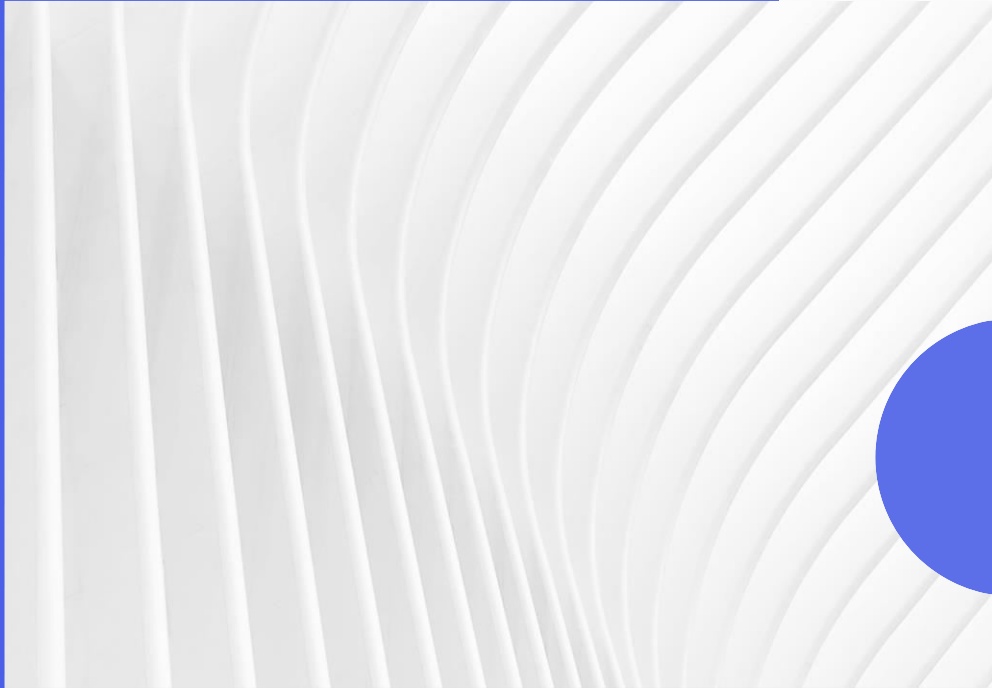
STRUCTURE OF DATA

KEY VARIABLES

INSIGHTS

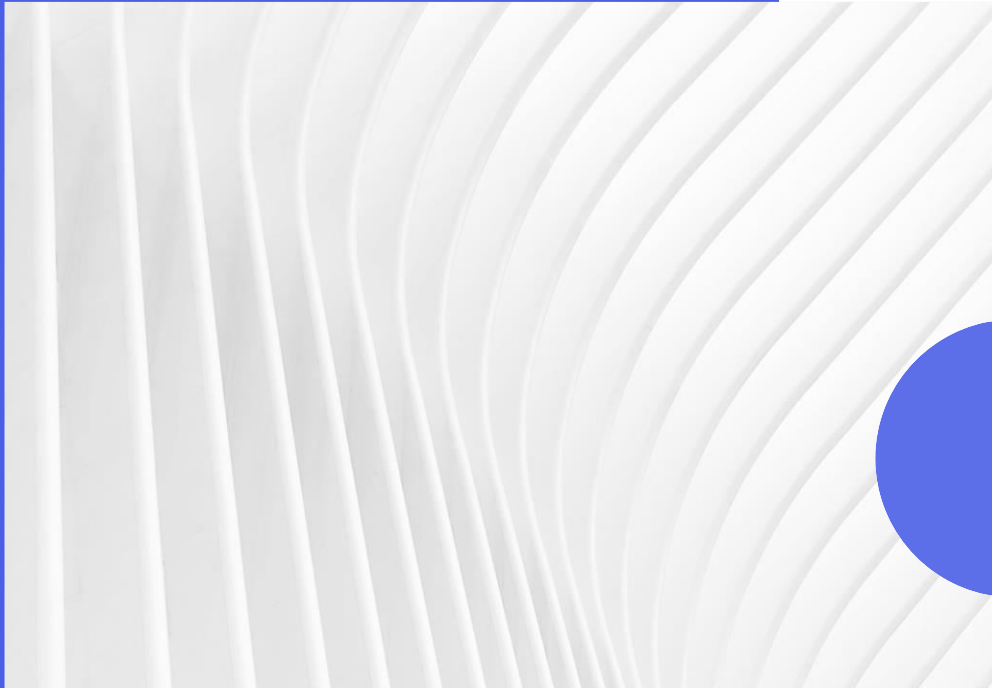
CONCLUSION

PROBLEM STATEMENT



The loan providing companies find it hard to give loans to the people due to their insufficient or non-existent credit history. Because of that, some consumers use it to their advantage by becoming a defaulter. Suppose you work for a consumer finance company which specializes in lending various types of loans to urban customers. You have to use EDA to analyze the patterns present in the data. This will ensure that the applicants capable of repaying the loan are not rejected.

DATASETS



This dataset has 3 files as explained below:

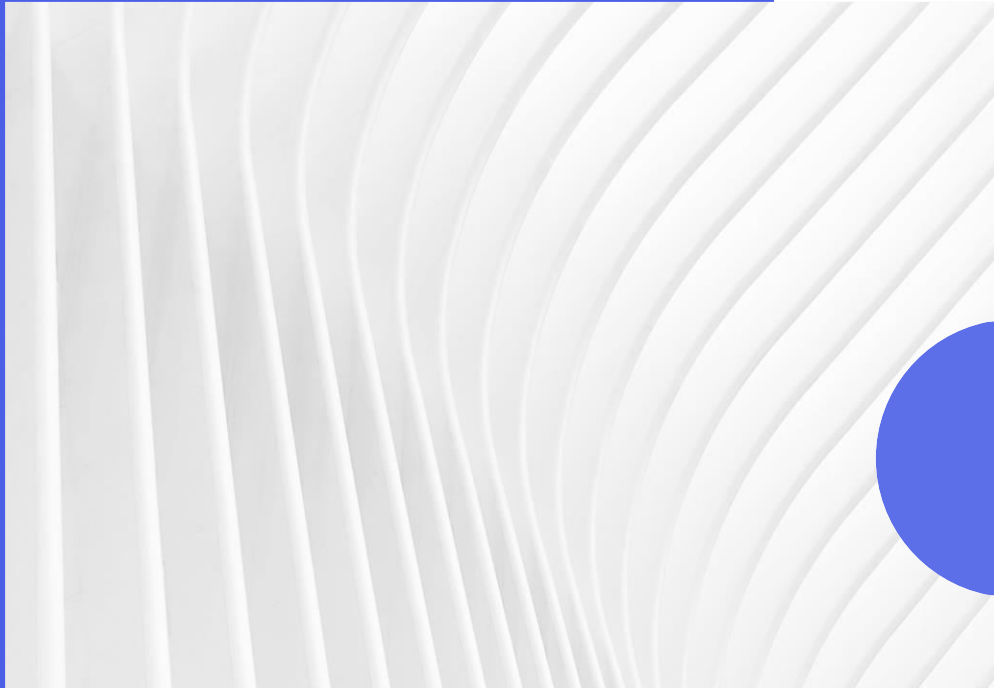
1. *'application_data.csv'* contains all the information of the client at the time of application.

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**.

3. *'columns_description.csv'* is data dictionary which describes the meaning of the variables.

ASSUMPTIONS



- The insights are provided after cleaning the data.
Mostly all the missing values are treated by imputing with mean, median and mode OR dropping them.
- The outliers are treated accordingly based on the understanding of the column.
- The TARGET column is the one we are focusing on and deriving insights using it.

STRUCTURE OF DATA

❖ Application Data

- Shape `(307511, 122)` , consists of 122 columns and roughly 3 lacs of rows.
- Info

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 307511 entries, 0 to 307510
Columns: 122 entries, SK_ID_CURR to AMT_REQ_CREDIT_BUREAU_YEAR
dtypes: float64(65), int64(41), object(16)
memory usage: 286.2+ MB
```

❖ Previous Application Data

- Shape `(1122312, 37)` , consists of 37 columns and roughly 11 lacs of rows.
- Info

```
dtypes: float64(15), int64(6), object(16)
memory usage: 325.4+ MB
```


KEY VARIABLES

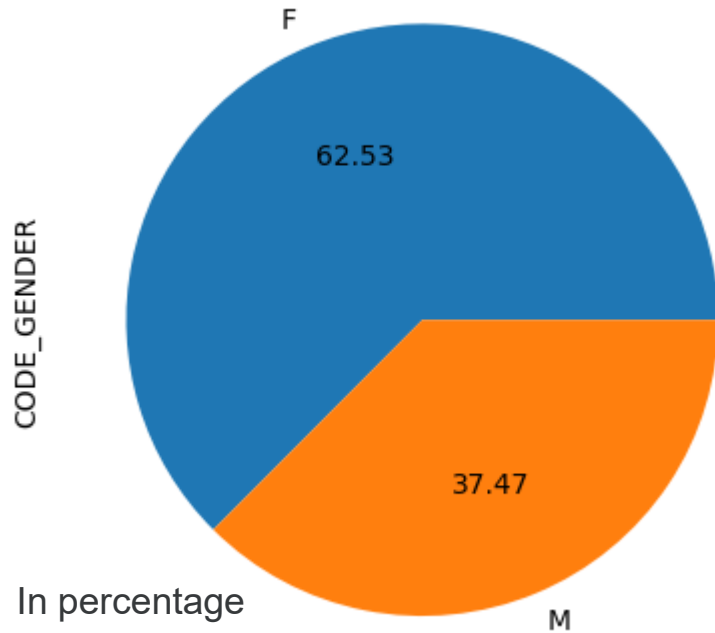
❖ Application Data

```
Index(['SK_ID_CURR', 'TARGET', 'NAME_CONTRACT_TYPE', 'CODE_GENDER',  
      'FLAG_OWN_CAR', 'FLAG_OWN_REALTY', 'CNT_CHILDREN', 'AMT_INCOME_TOTAL',  
      'AMT_CREDIT', 'AMT_ANNUITY', 'AMT_GOODS_PRICE', 'NAME_TYPE_SUITE',  
      'NAME_INCOME_TYPE', 'NAME_EDUCATION_TYPE', 'NAME_FAMILY_STATUS',  
      'NAME_HOUSING_TYPE', 'REGION_POPULATION_RELATIVE', 'DAYS_BIRTH',  
      'DAYS_EMPLOYED', 'DAYS_REGISTRATION', 'DAYS_ID_PUBLISH', 'FLAG_MOBILE',  
      'FLAG_EMP_PHONE', 'FLAG_WORK_PHONE', 'FLAG_CONT_MOBILE', 'FLAG_PHONE',  
      'FLAG_EMAIL', 'OCCUPATION_TYPE', 'CNT_FAM_MEMBERS',  
      'REGION_RATING_CLIENT', 'REGION_RATING_CLIENT_W_CITY',  
      'WEEKDAY_APPR_PROCESS_START', 'HOUR_APPR_PROCESS_START',  
      'REG_REGION_NOT_LIVE_REGION', 'REG_REGION_NOT_WORK_REGION',  
      'LIVE_REGION_NOT_WORK_REGION', 'REG_CITY_NOT_LIVE_CITY',  
      'REG_CITY_NOT_WORK_CITY', 'LIVE_CITY_NOT_WORK_CITY',  
      'ORGANIZATION_TYPE', 'OBS_30_CNT_SOCIAL_CIRCLE',  
      'DEF_30_CNT_SOCIAL_CIRCLE', 'OBS_60_CNT_SOCIAL_CIRCLE',  
      'DEF_60_CNT_SOCIAL_CIRCLE', 'DAYS_LAST_PHONE_CHANGE', 'YEARS_BIRTH',  
      'YEARS_EMPLOYED', 'YEARS_REGISTRATION', 'YEARS_ID_PUBLISH'],  
      dtype='object')
```

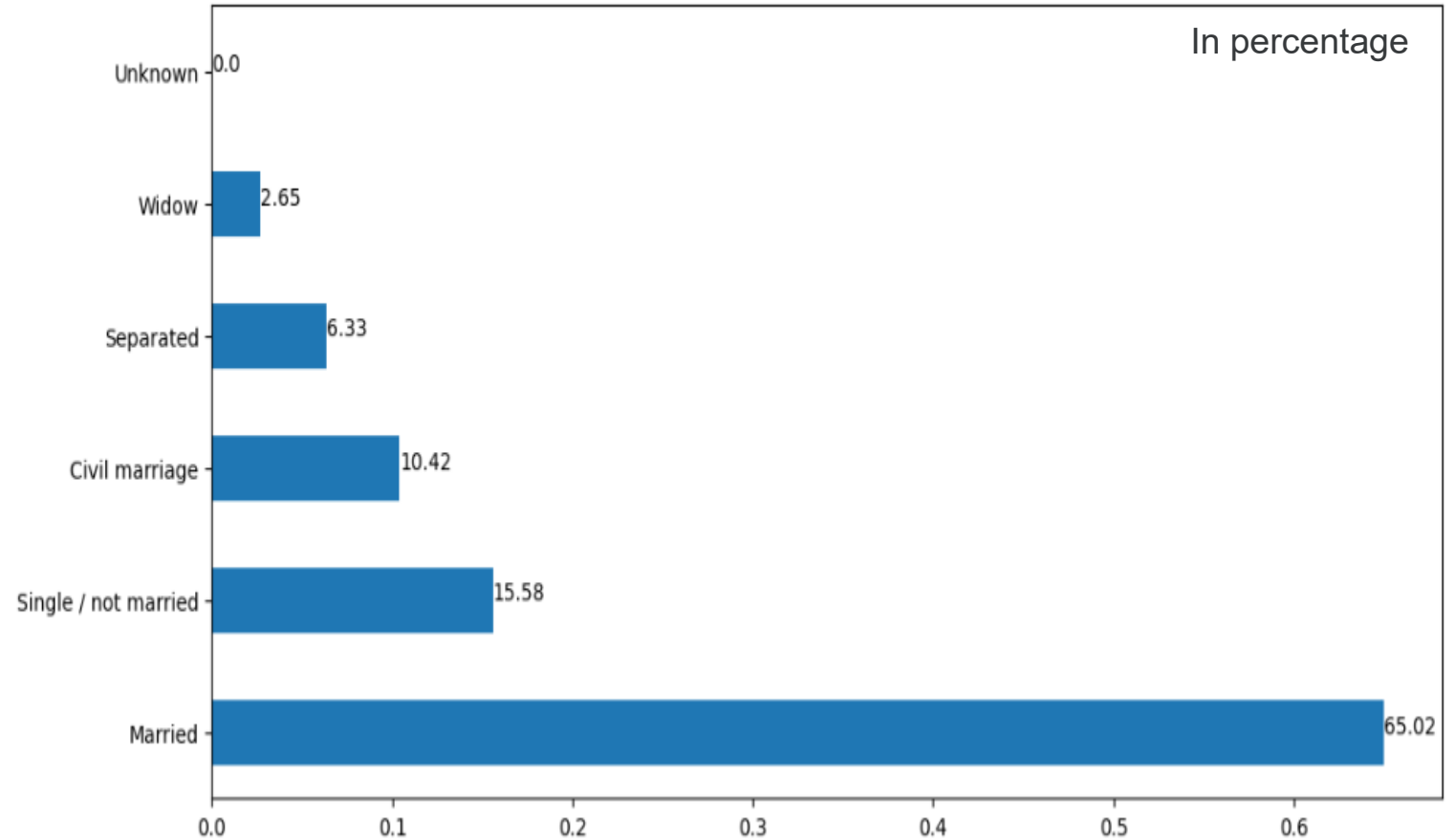
❖ Previous Application Data

```
Index(['SK_ID_PREV', 'SK_ID_CURR', 'NAME_CONTRACT_TYPE', 'AMT_ANNUITY',  
      'AMT_APPLICATION', 'AMT_CREDIT', 'AMT_DOWN_PAYMENT', 'AMT_GOODS_PRICE',  
      'WEEKDAY_APPR_PROCESS_START', 'HOUR_APPR_PROCESS_START',  
      'FLAG_LAST_APPL_PER_CONTRACT', 'NFLAG_LAST_APPL_IN_DAY',  
      'RATE_DOWN_PAYMENT', 'RATE_INTEREST_PRIMARY',  
      'RATE_INTEREST_PRIVILEGED', 'NAME_CASH_LOAN_PURPOSE',  
      'NAME_CONTRACT_STATUS', 'DAYS_DECISION', 'NAME_PAYMENT_TYPE',  
      'CODE_REJECT_REASON', 'NAME_TYPE_SUITE', 'NAME_CLIENT_TYPE',  
      'NAME_GOODS_CATEGORY', 'NAME_PORTFOLIO', 'NAME_PRODUCT_TYPE',  
      'CHANNEL_TYPE', 'SELLERPLACE_AREA', 'NAME_SELLER_INDUSTRY',  
      'CNT_PAYMENT', 'NAME_YIELD_GROUP', 'PRODUCT_COMBINATION',  
      'DAYS_FIRST_DRAWING', 'DAYS_FIRST_DUE', 'DAYS_LAST_DUE_1ST_VERSION',  
      'DAYS_LAST_DUE', 'DAYS_TERMINATION', 'NFLAG_INSURED_ON_APPROVAL'],  
      dtype='object')
```

INSIGHTS

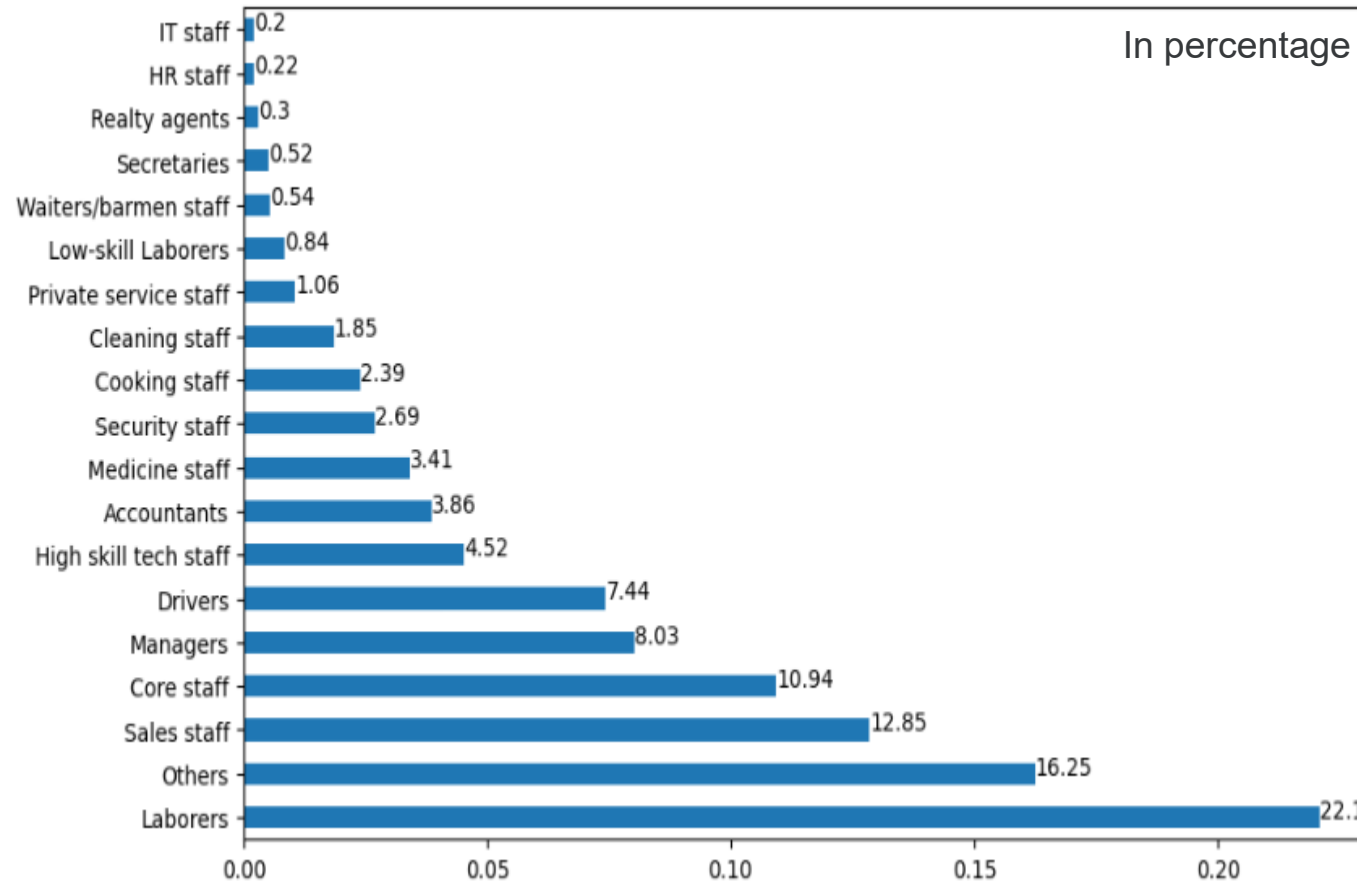


Females are more for loan application than Male

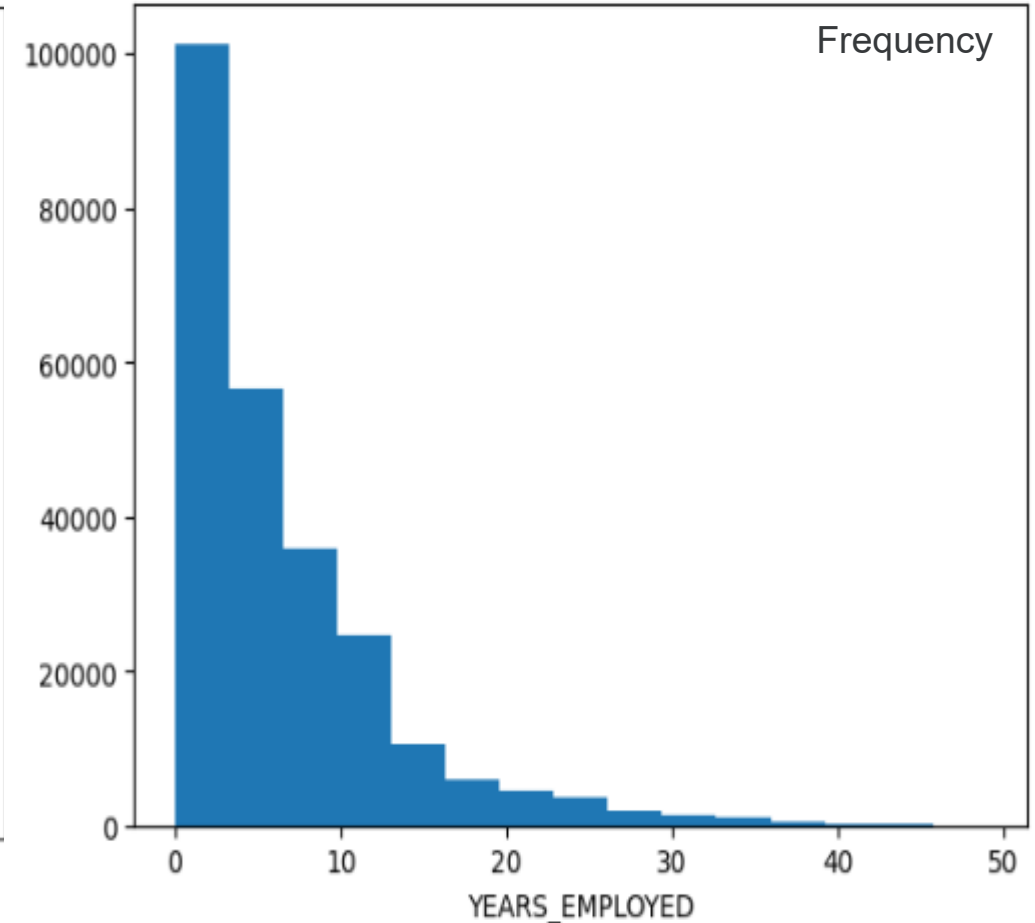


Married people are in majority in applying for a loan.

INSIGHTS

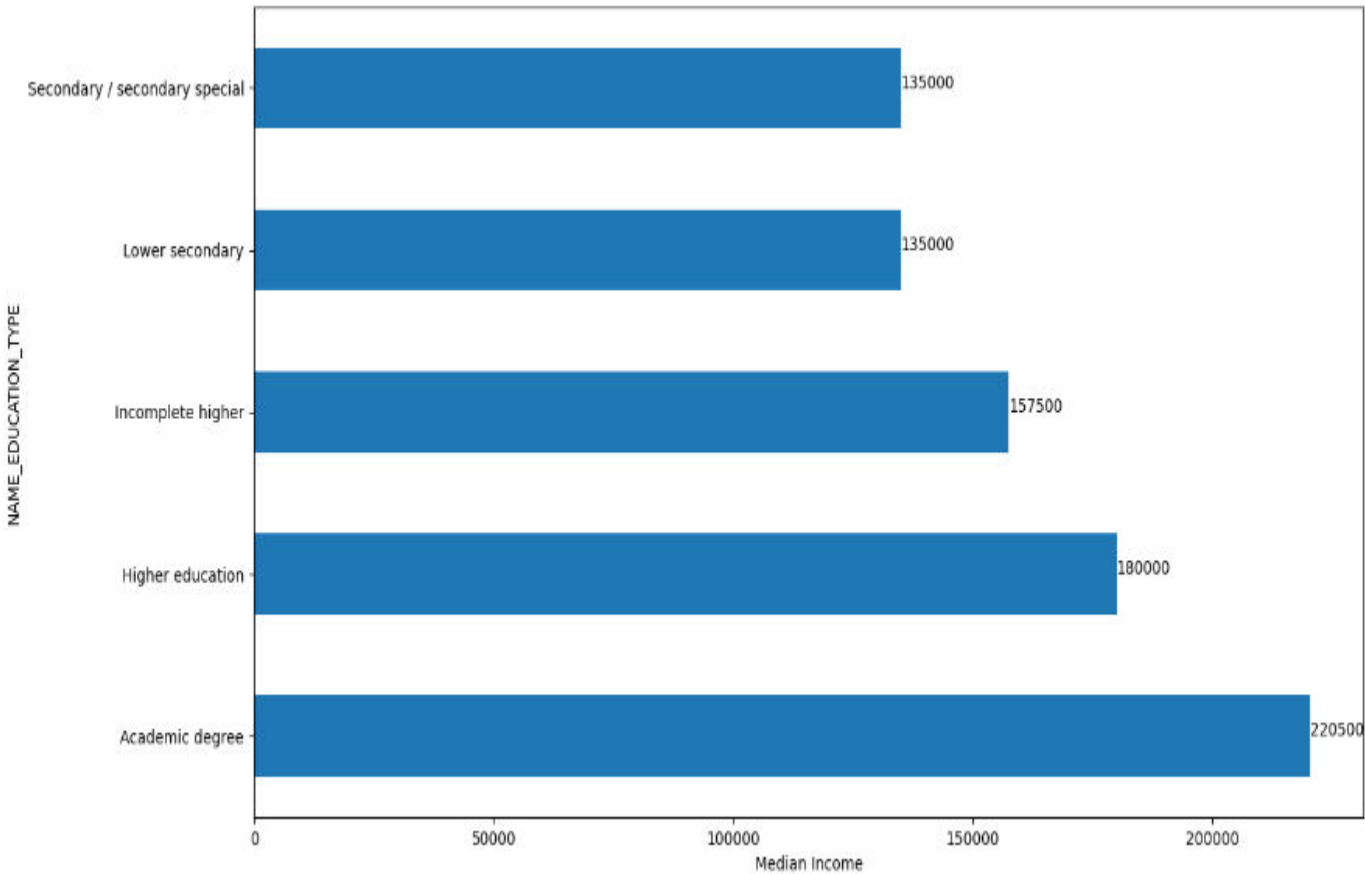


Laborers are in majority for loan applications

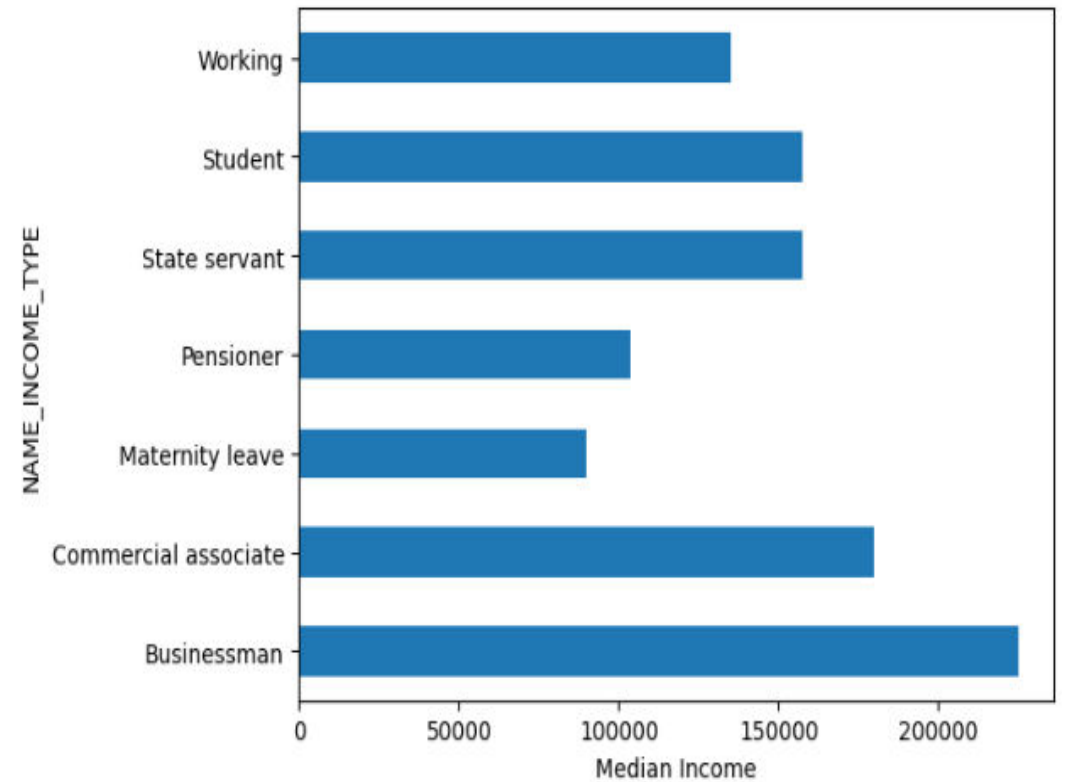


People in their initial phase of employment are more likely to apply for loans

INSIGHTS

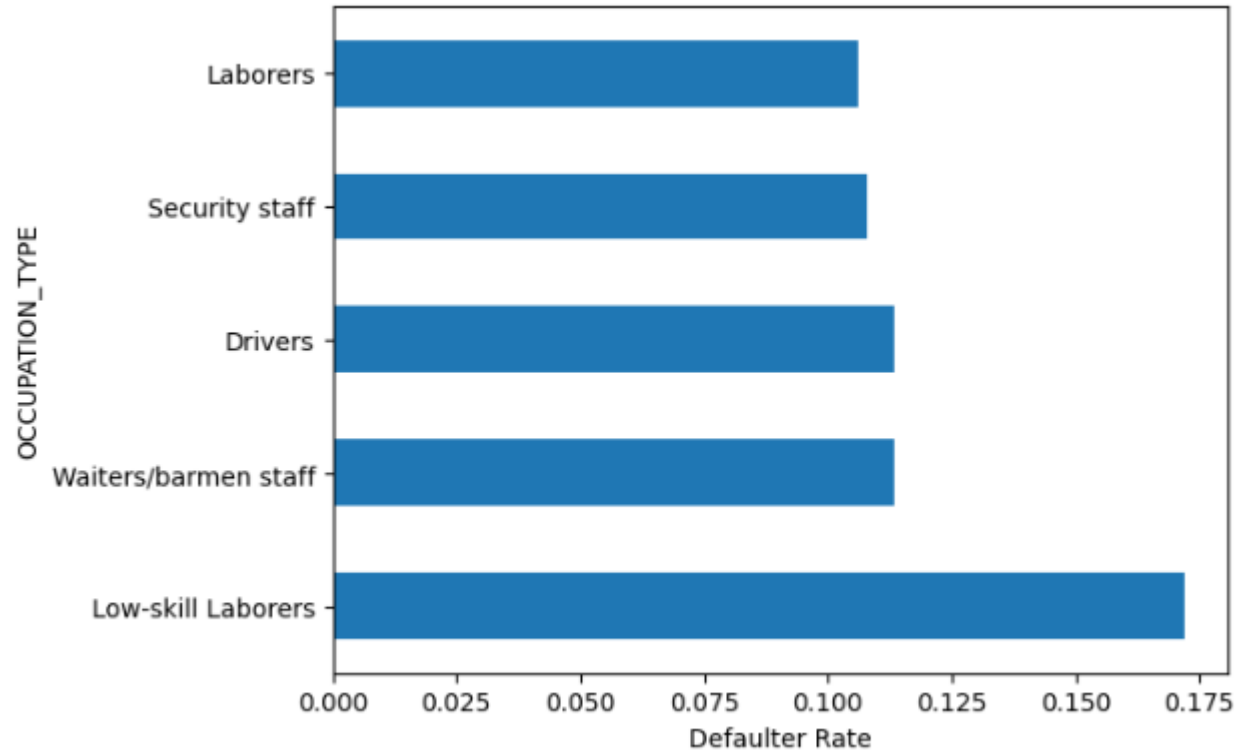


People with an Academic Degree have higher income

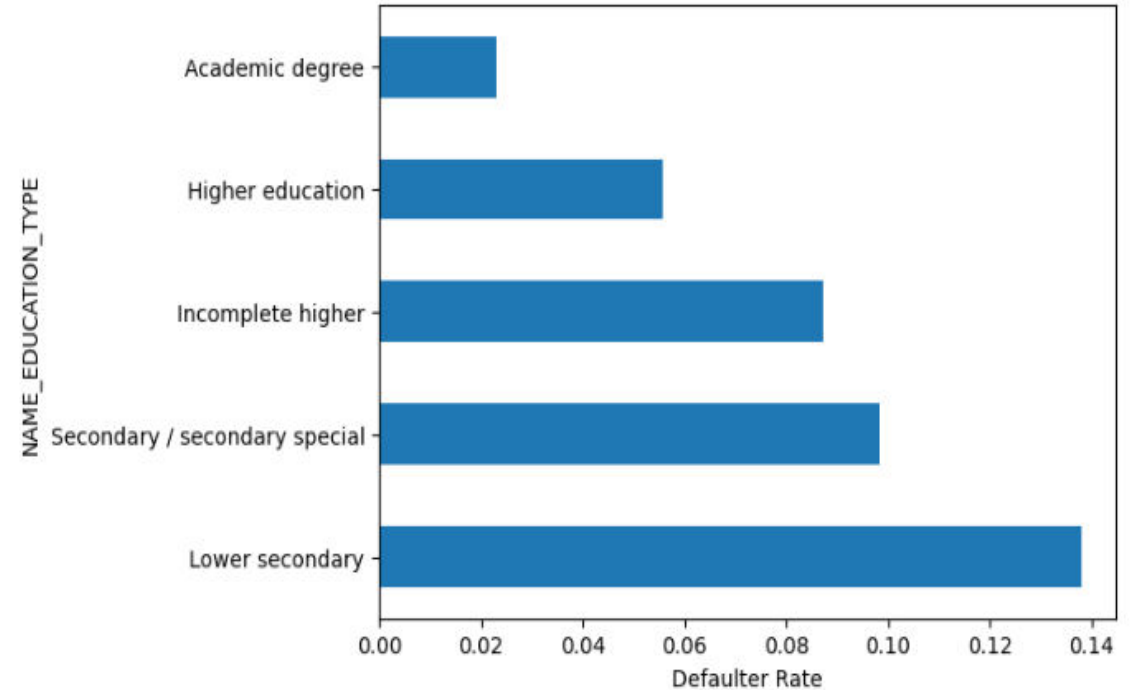


Businessman have higher income

INSIGHTS

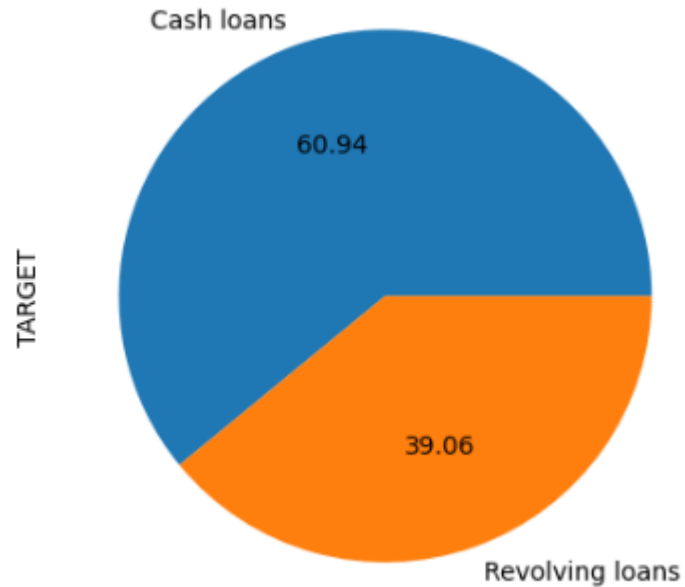


Low-skill Laborers have a higher rate of being a defaulter in cases of loans.

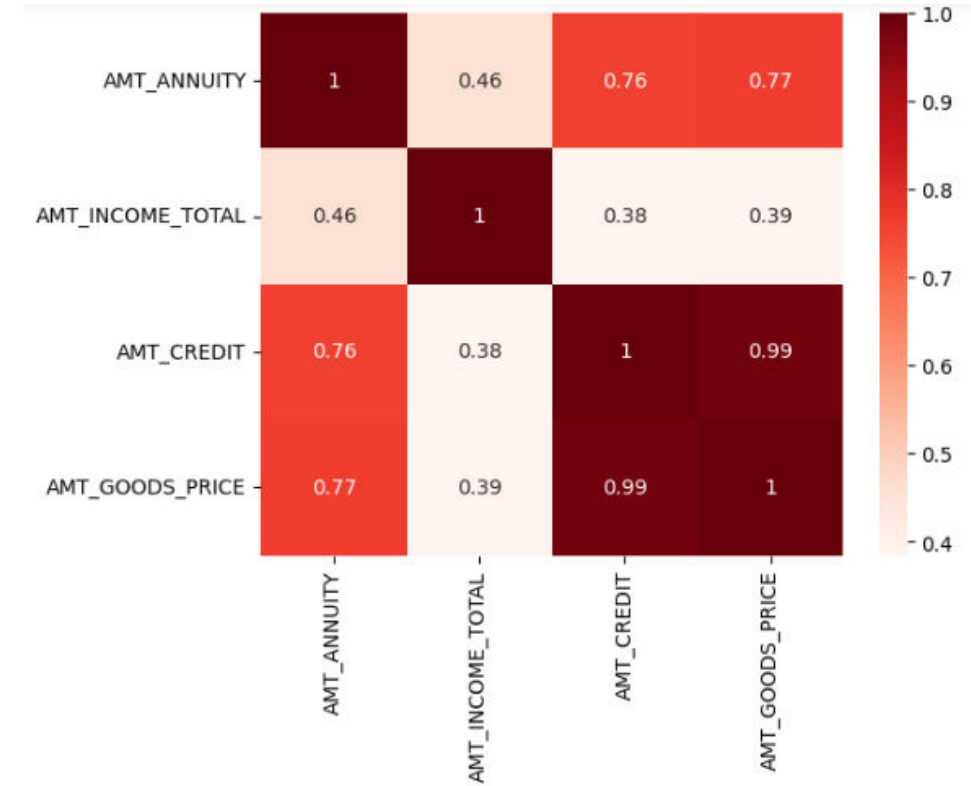


It seems , level of education is proportional of person being a defaulter in cases of loans

INSIGHTS

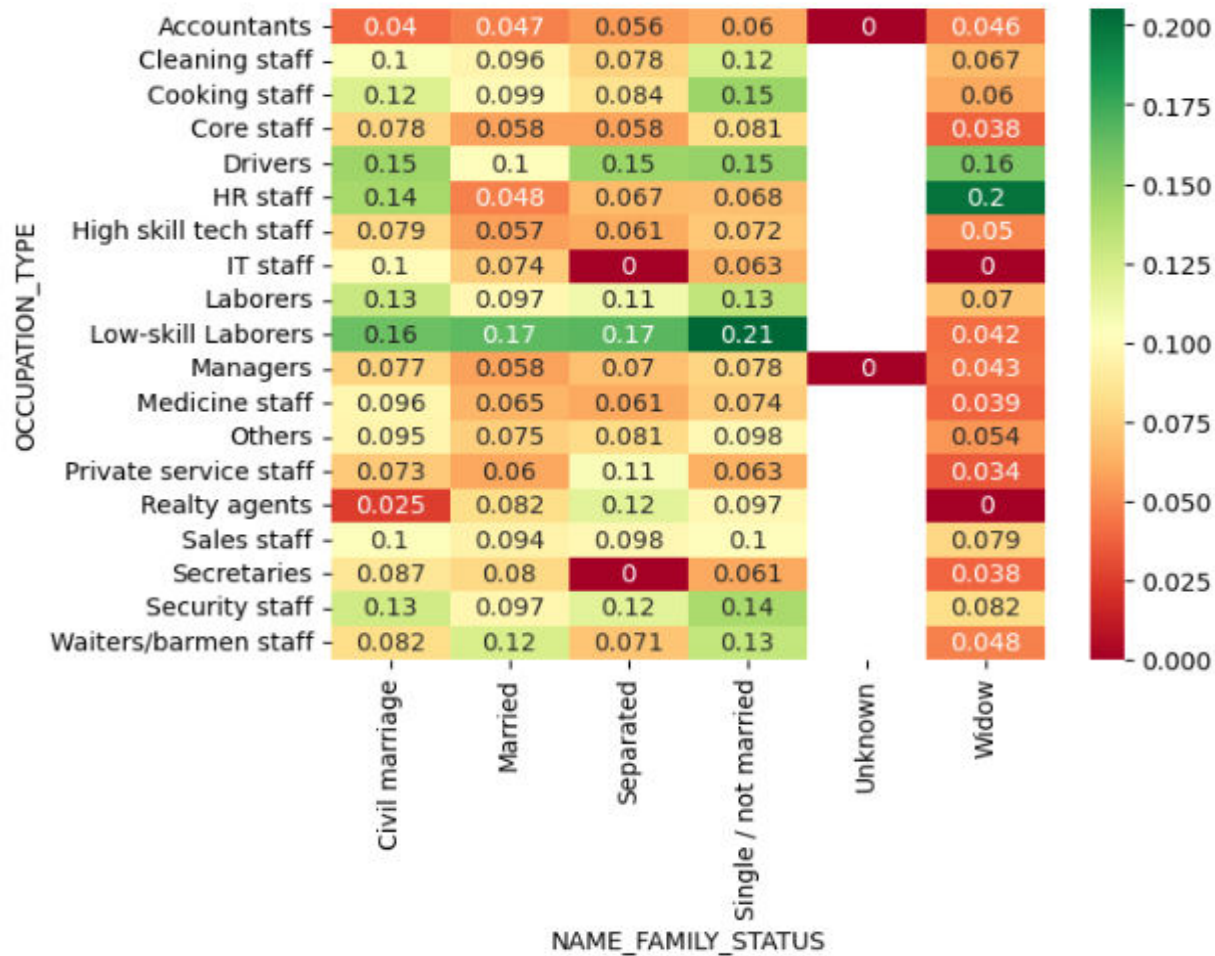


There are more defaulters in case of Cash Loans in comparison to Revolving Loans

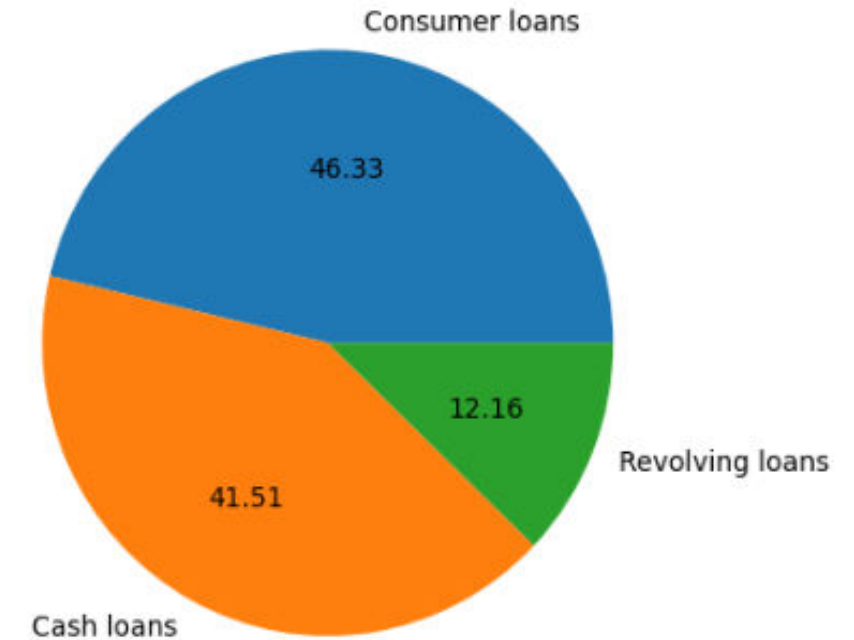


There is a high correlation among the Amount variables.

INSIGHTS



NAME_CONTRACT_TYPE

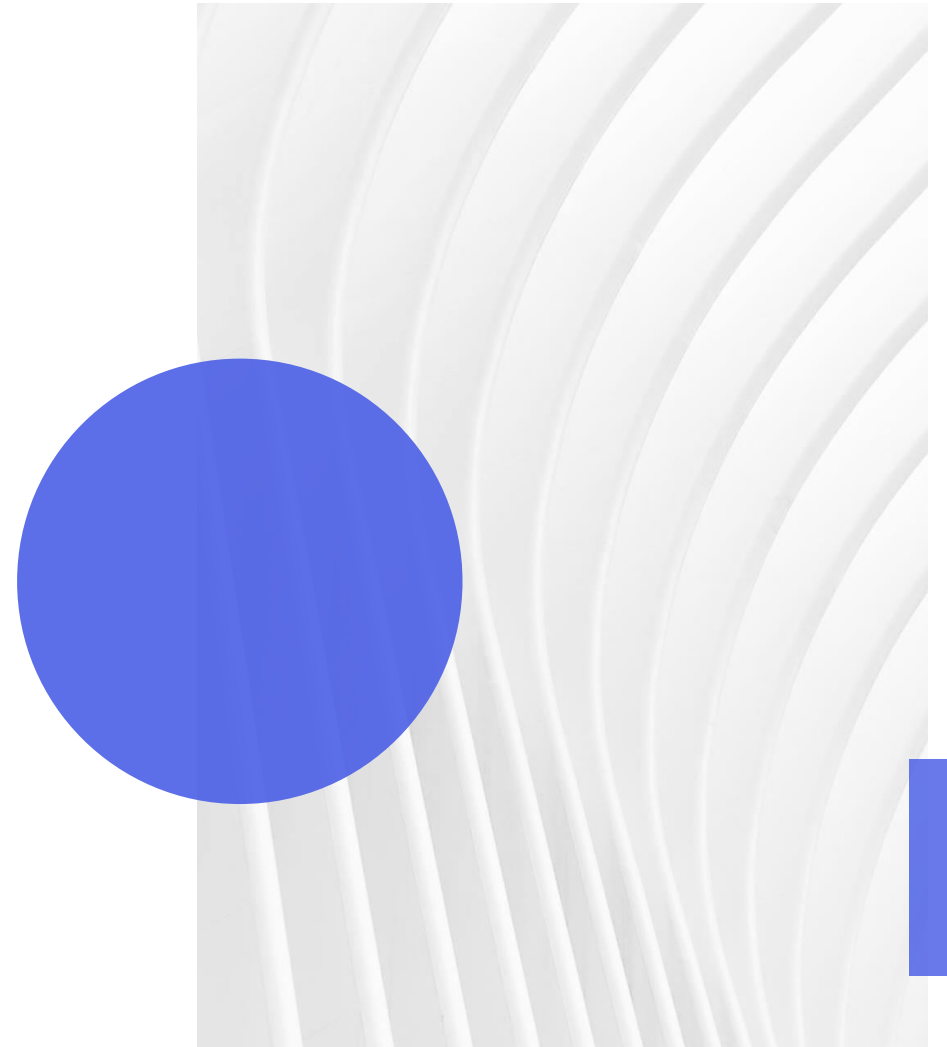


People generally apply for Consumer and Cash Loans

Low-skilled Laborer who is single is more likely to be a defaulter

SUMMARY

- ✓ People with higher education level and good income are very less likely to be a defaulter.
- ✓ People usually apply for Cash Loans.
- ✓ People in initial phase of employment are more likely to apply for a loan.
- ✓ Level of Education is proportional to being a defaulter.





THANK YOU

PARTH GUPTA

parthgupta7267@gmail.com