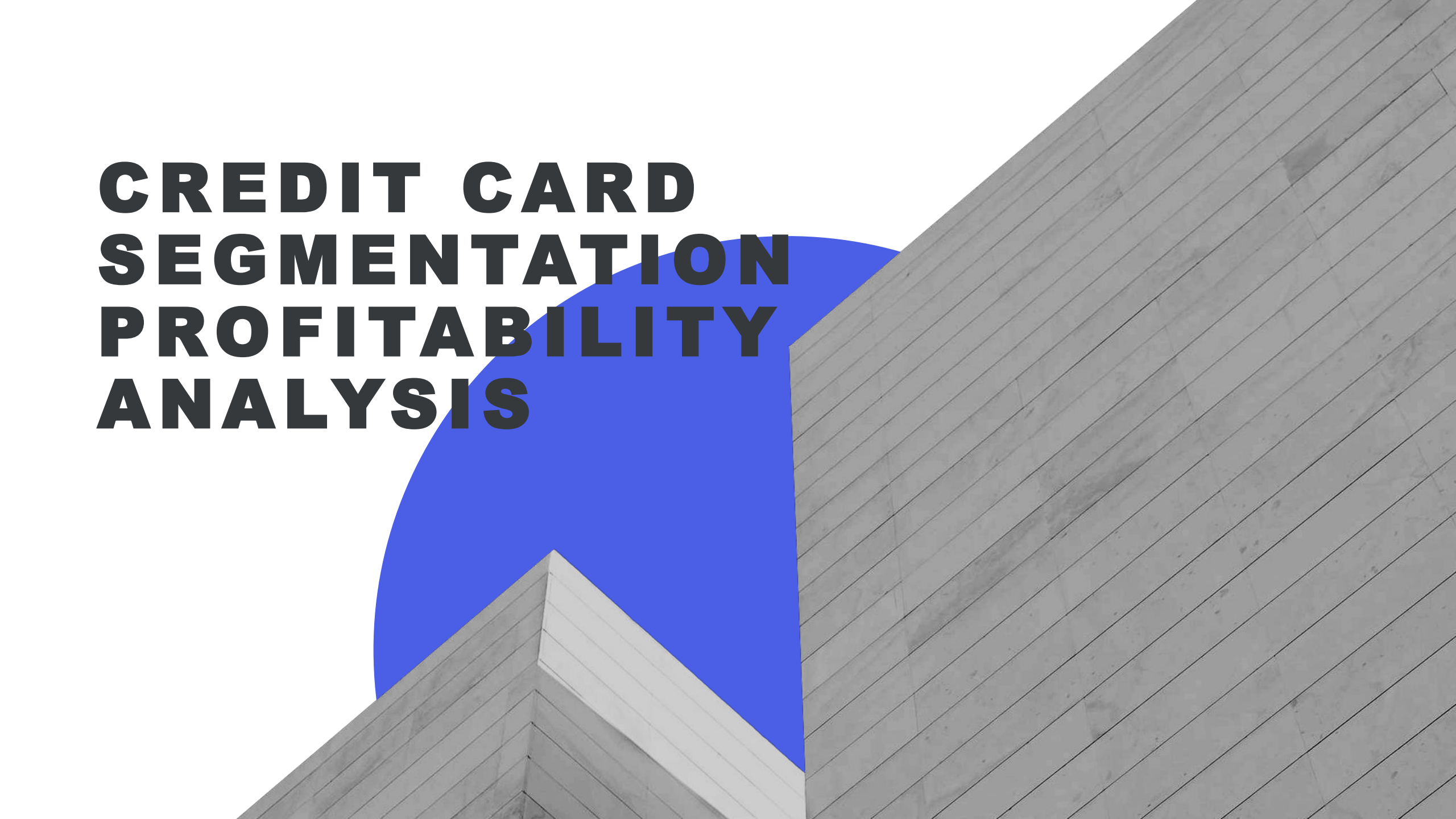
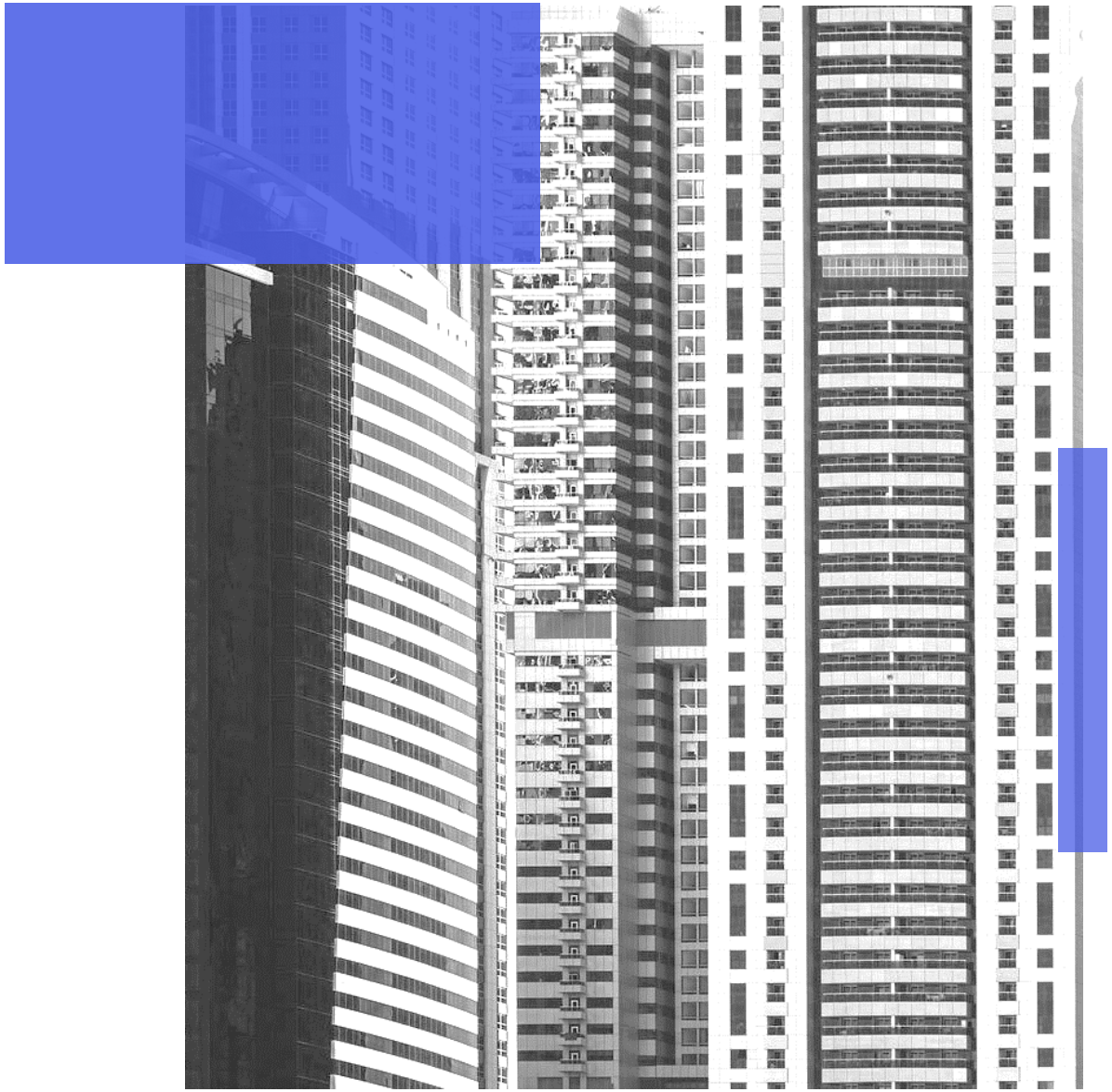
# CREDIT CARD SEGMENTATION PROFITABILITY ANALYSIS





# AGENDA

PROBLEM DEFINITION

DATASET

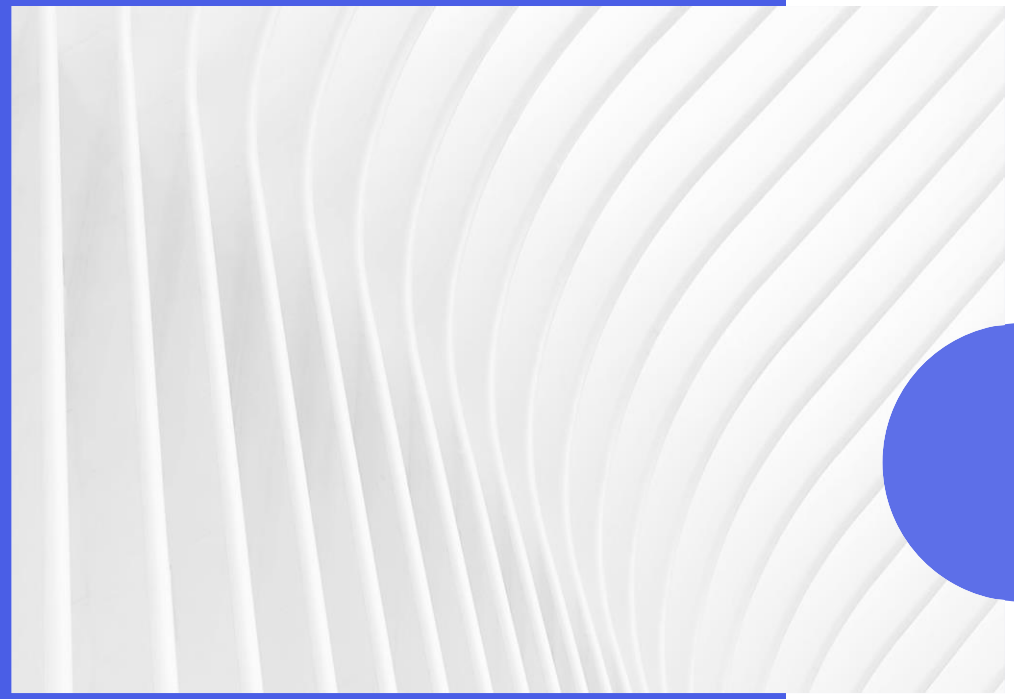
ASSSUMPTIONS

EXPLORATORY ANALYSIS

CHALLENGES

TIMELINE

# PROBLEM DEFINITION



The primary driver of commercial bank failures in recessionary high interest rate environments are credit losses and weak loan demand.

This work aims to provide a risk-driven framework for lenders to issue credit the that maximizes revenue and minimizes risk in the current economic landscape.

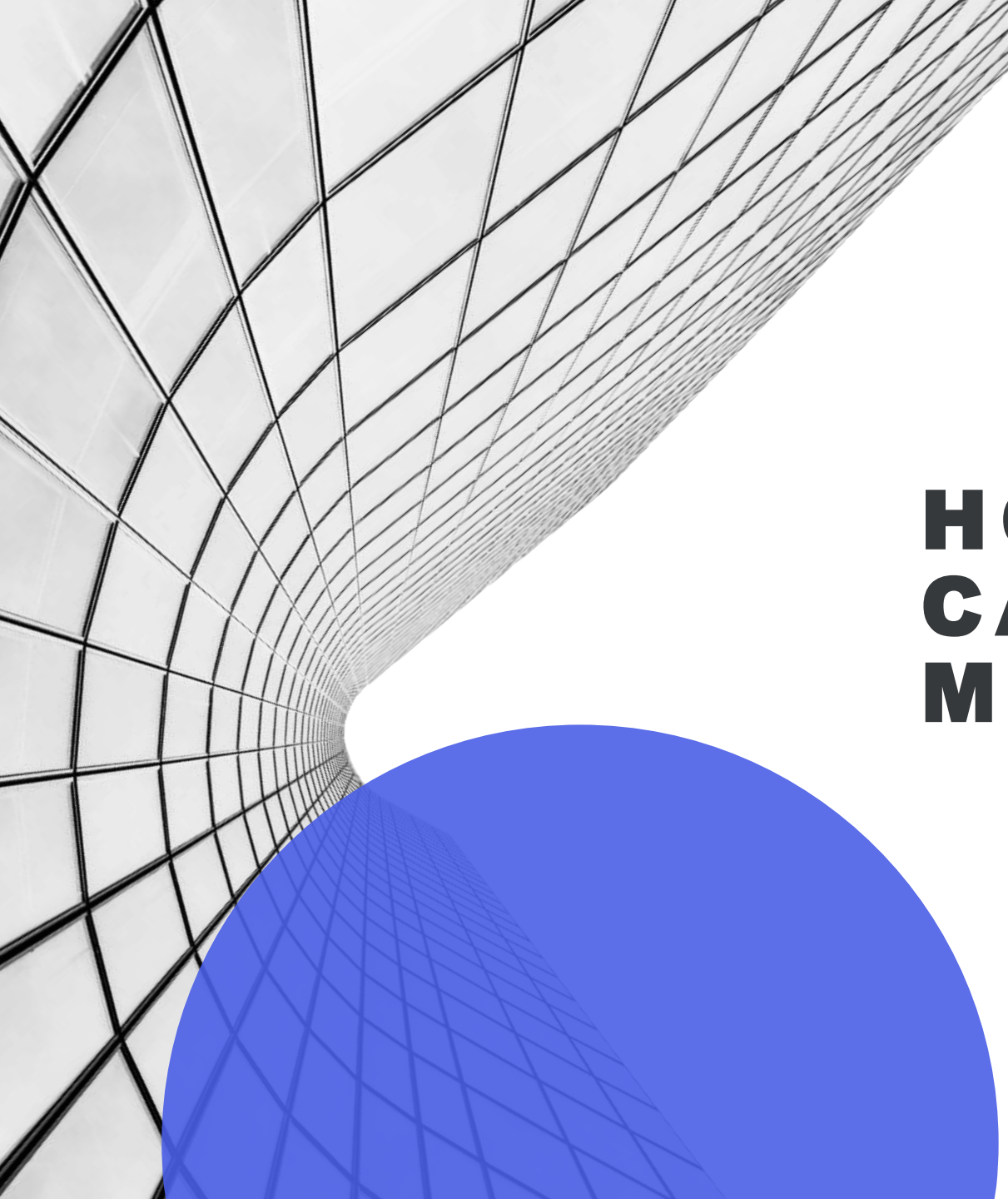
# DATASET

## application_record.csv

| | |
|---|---|
| **ID** | Client Number |
| **CODE_GENDER** | Gender |
| **FLAG_OWN_CAR** | Owns car? |
| **FLAG_OWN_REALITY** | Owns property? |
| **CNT_CHILDREN** | Number of children |
| **AMT_INCOME_TOTAL** | Annual income |
| **NAME_INCOME_TYPE** | Income category/source |
| **NAME_EDUCATION_TYPE** | Education level |
| **NAME_FAMILY_STATUS** | Marital status |
| **NAME_HOUSING_TYPE** | Way of living |
| **DAYS_BIRTH** | Age |
| **DAYS_EMPLOYED** | Days Employed |
| **FLAG_MOBIL** | Mobile phone? |
| **FLAG_WORK_PHONE** | Work phone? |
| **FLAG_PHONE** | Phone? |
| **FLAG_EMAIL** | Email? |
| **OCCUPATION TYPE** | Occupation category |
| **CNT_FAM_MEMBERS** | Family Size |

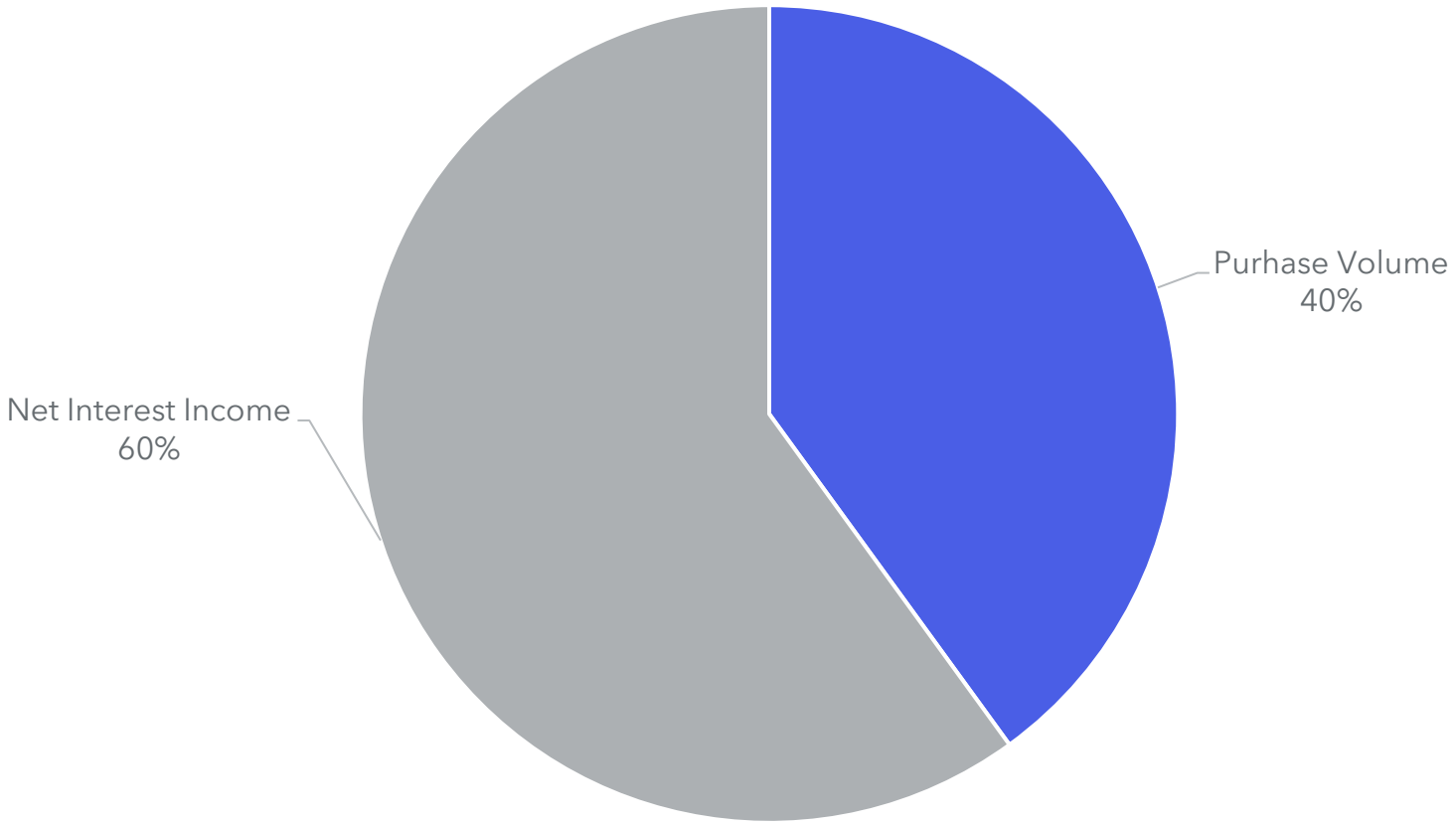
## credit_record.csv

| | |
|---|---|
| **ID** | Client Number |
| **MONTHS_BALANCE** | Number of months on book |
| **STATUS** | No loan / loan paid off / X days overdue |

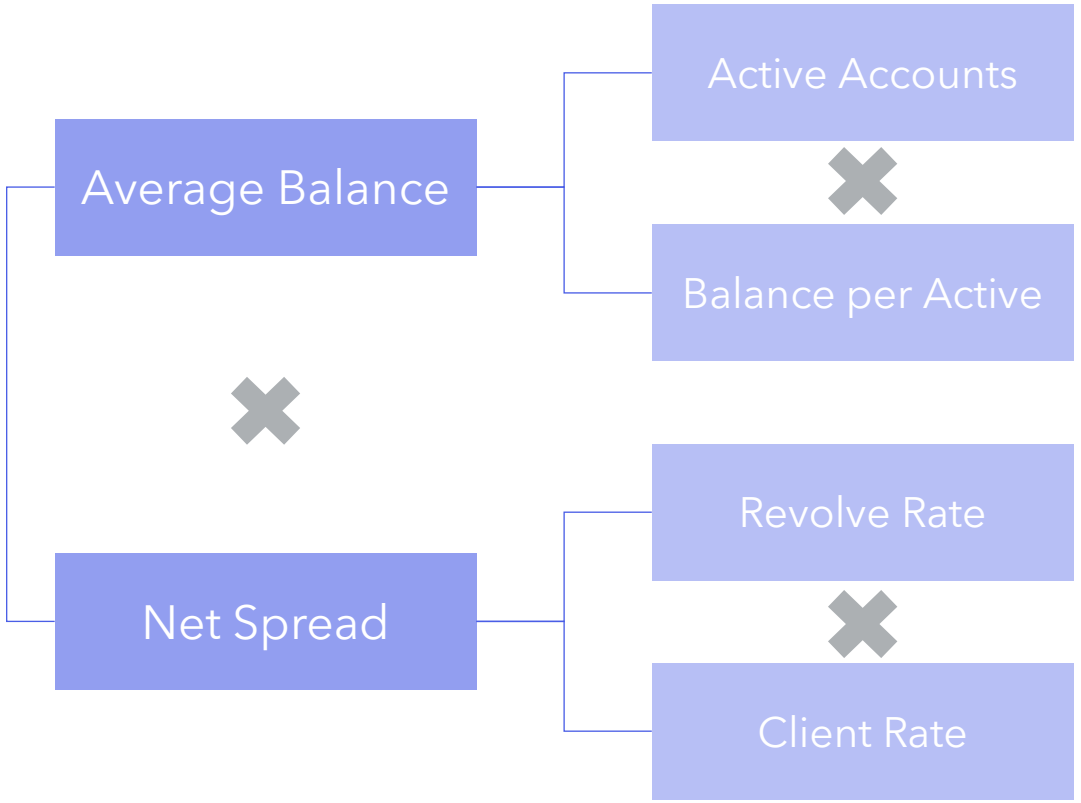


# HOW DO CREDIT CARDS MAKE MONEY

# REVENUE DRIVERS



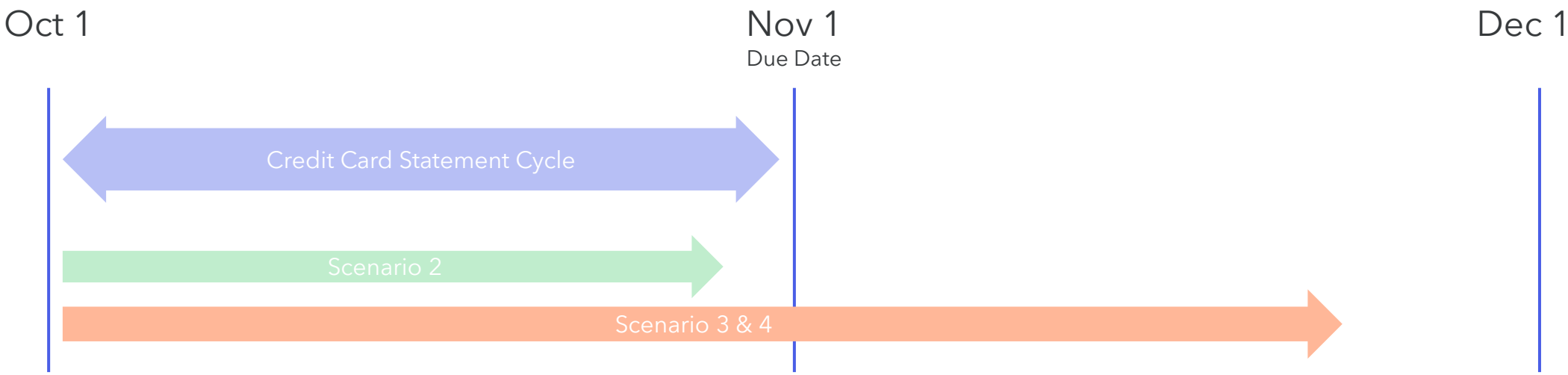
# NET INTEREST INCOME
## (OVERDUE FEES)




**% OF ACTIVE CLIENTS OVERDUE**

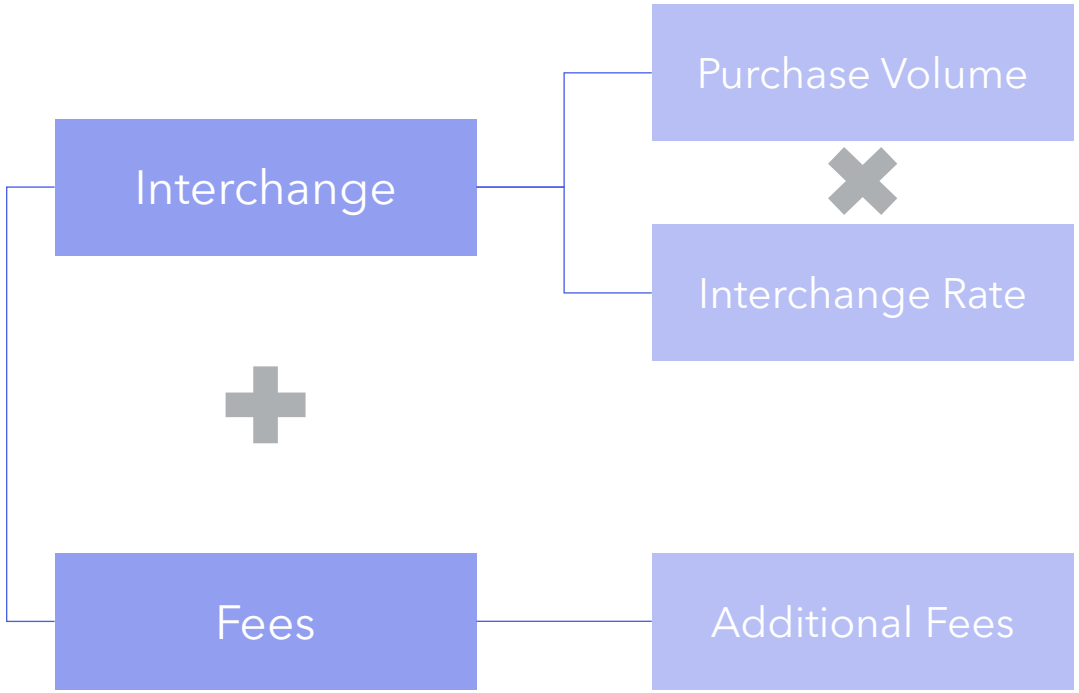
**NOMINAL INTEREST RATE (%19.99)**

# CALCULATING NET INTEREST INCOME




Scenario 1: No transactions made during the statement period. No interest income.
Scenario 2: Full balance paid by due date, hence no interest charged. No interest income.
Scenario 3: Interest assessed on overdue balance and compounds for subsequent cycles until paid off.
Scenario 4: Write off clients as bad debts after many cycles, resulting in a net loss.

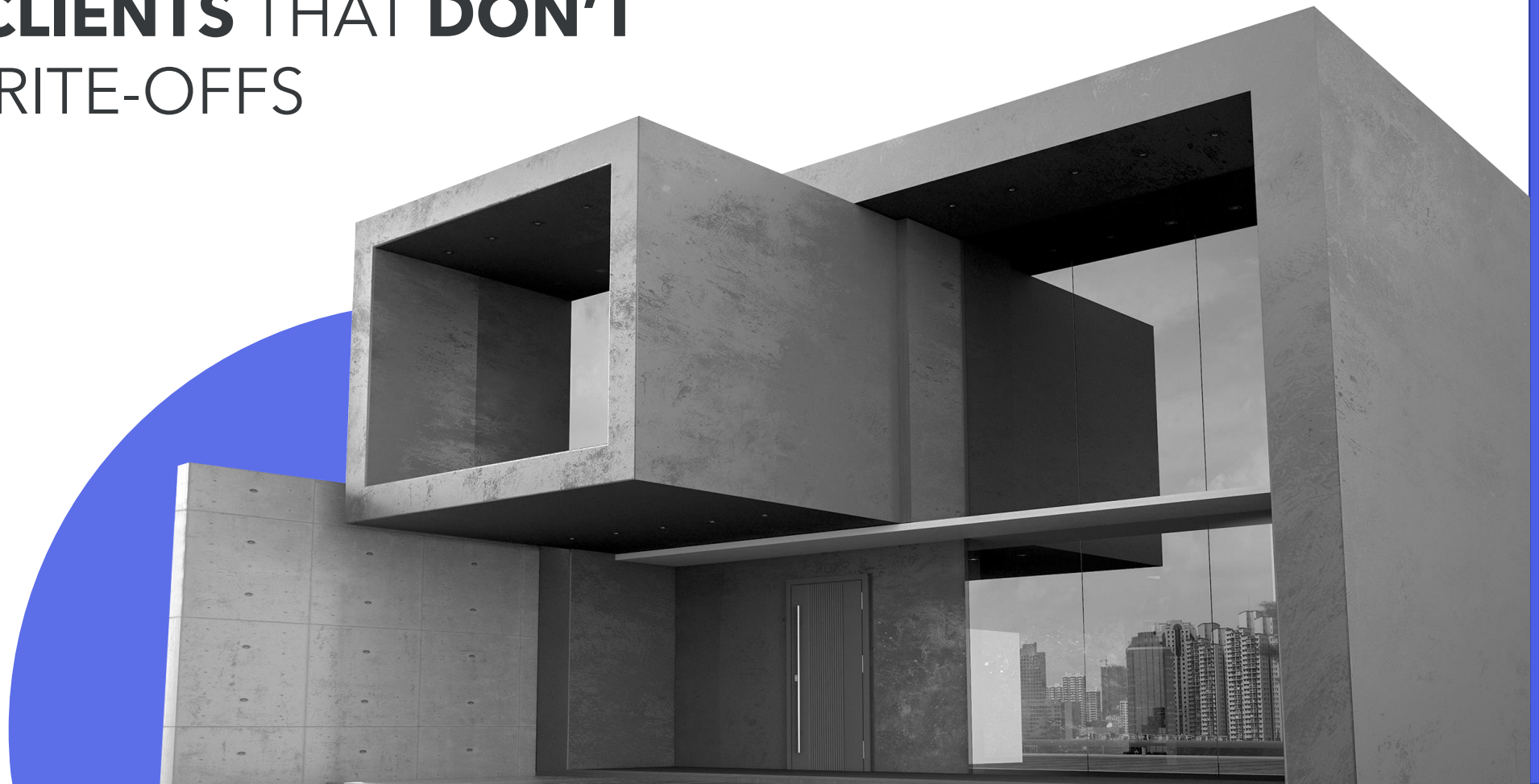
# PURCHASE VOLUME




**PAID FOR ISSUING THE SERVICE ACROSS THE NETWORK**

*the objective*

# **MAXIMIZE** THE NUMBER OF **OVERDUE CLIENTS** THAT **DON'T** BECOME WRITE-OFFS



# CLIENT PROFILING

**Scenario 1**

No purchases made

$0 revenue

**Scenario 2**

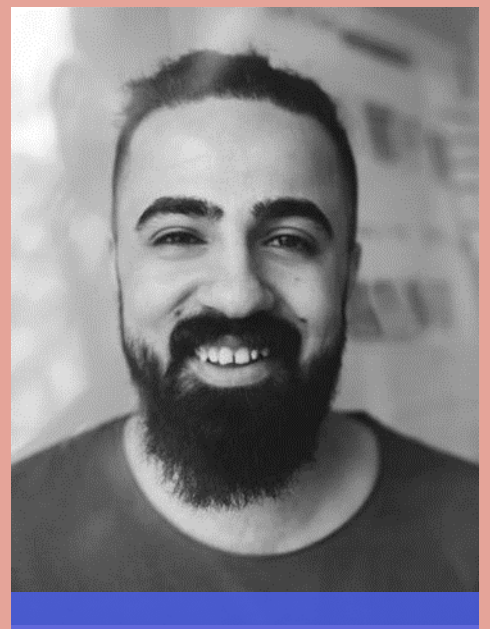
Balance fully paid off

Interchange

**Scenario 3**

Late payment – overdue

Interchange + net interest



**Scenario 4**

Late payment – write off

Net loss

# ASSUMPTIONS

| ADDITIONAL REVENUE | COSTS | SPEND CONTROL | LOSSES |
|---|---|---|---|
| Finance Charges | Cost of Funds | Balance Utilization (Interest Income) | Unit Write Off Rate |
| **Earnings on Capital** | **Management Fees** | **Limit Increases** | **Balance Control Ratio** |
| Balance Transfers | Marginal Operating Expenses | Purchase Utilization (Interchange Income) | Loss Assumption Factors |
| **Transaction Fees** | **Cost of Rewards** | | |
| Over-limit Fees | Insurance | | |
| **Annual Fees** | | | |

# EXPLORATORY ANALYSIS

## UNIVARIATE ANALYSIS

Single variable characteristics and attributes

## BIVARIATE ANALYSIS

Correlative observations involving 2 variables.

## MULTIVARIATE ANALYSIS

Correlative observations involving 2+ variables.

# UNIVARIATE ANALYSIS

**438, 557**

**IDs with an Application Record**

**45,985**

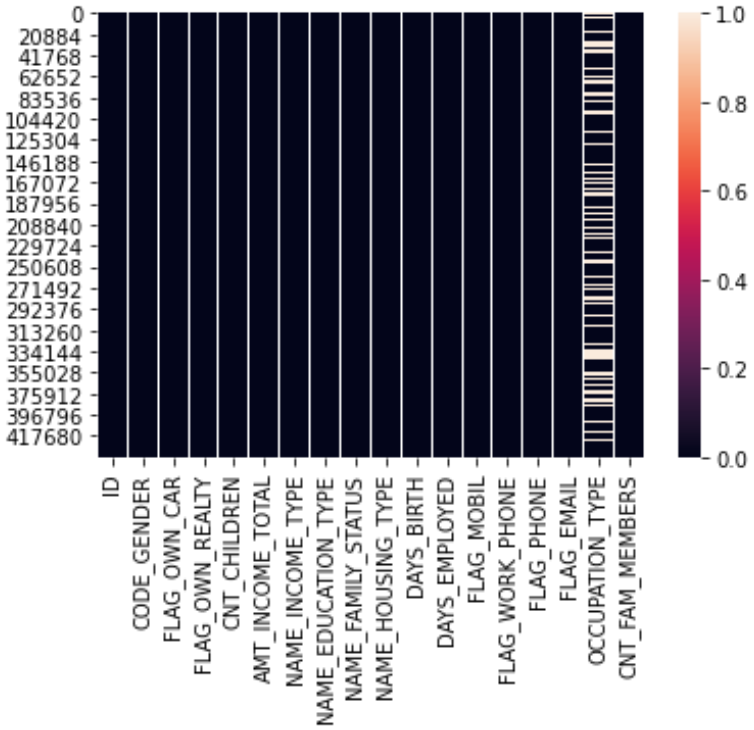
**IDs with a Credit Record**

**36,457**

**IDs with a Credit & Application Record**

Less than 10% of IDs have a credit & application record which limits the set sizes. Although, a high accuracy predictive model will provide unvalidatable approval recommendations to for all applicants.

# UNIVARIATE ANALYSIS



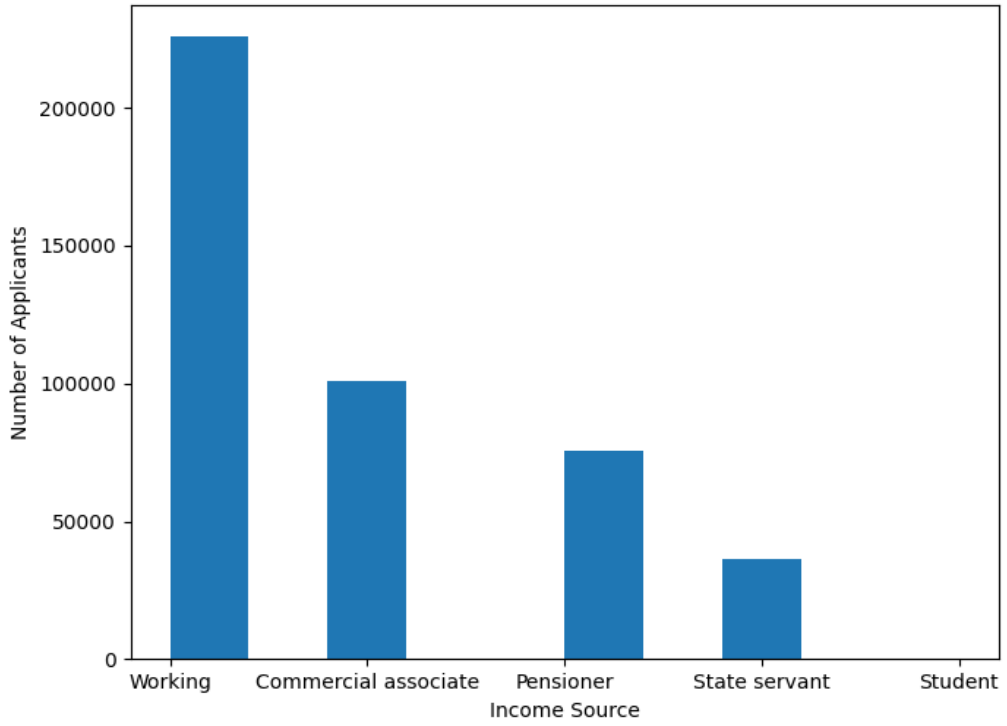


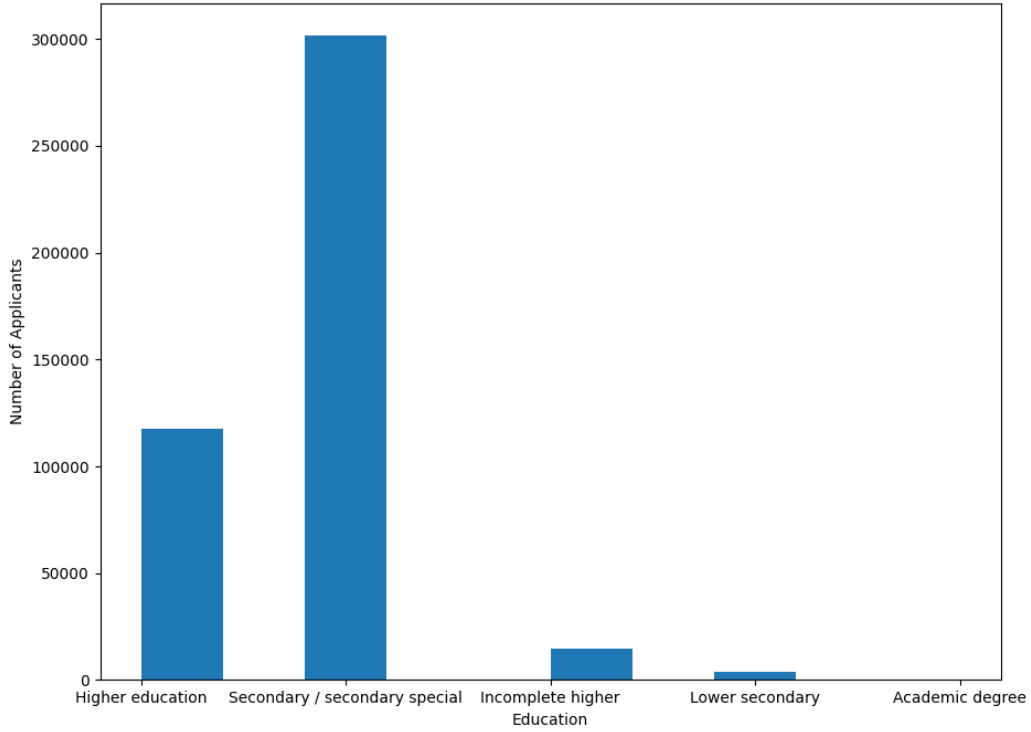
**Applicant NULL Check**

**Credit Record NULL CHECK**

Occupation Type only feature missing values across both datasets. Can be dropped due to information (career/line of work) being captured in other features (Income Type, Education Type, etc.) – see following slides

# UNIVARIATE ANALYSIS



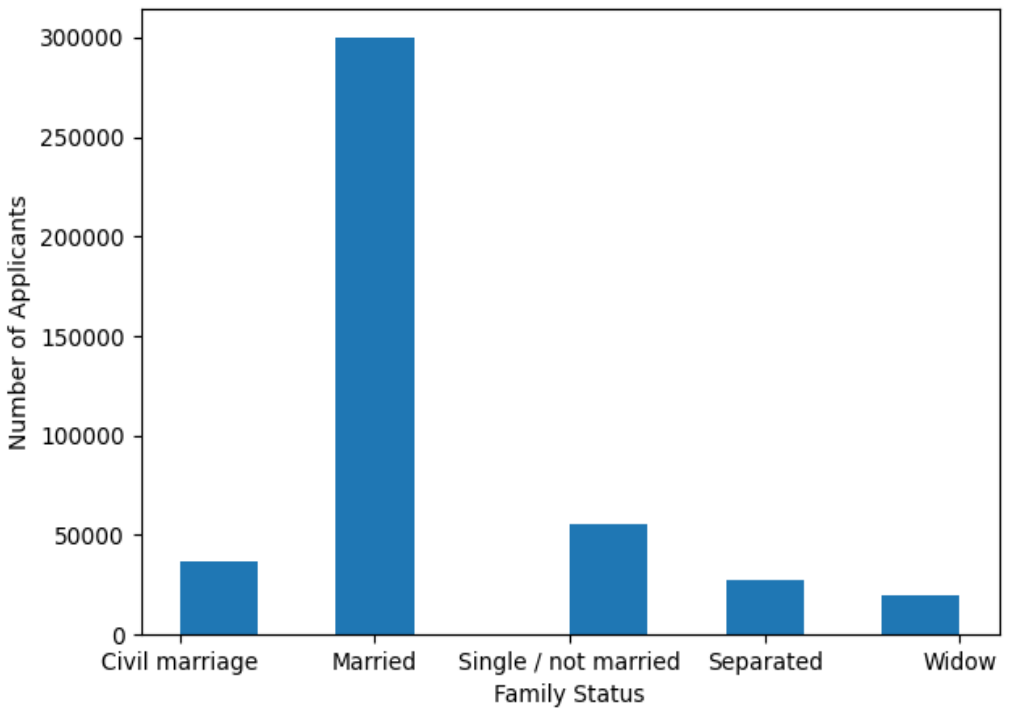




**Income Source**

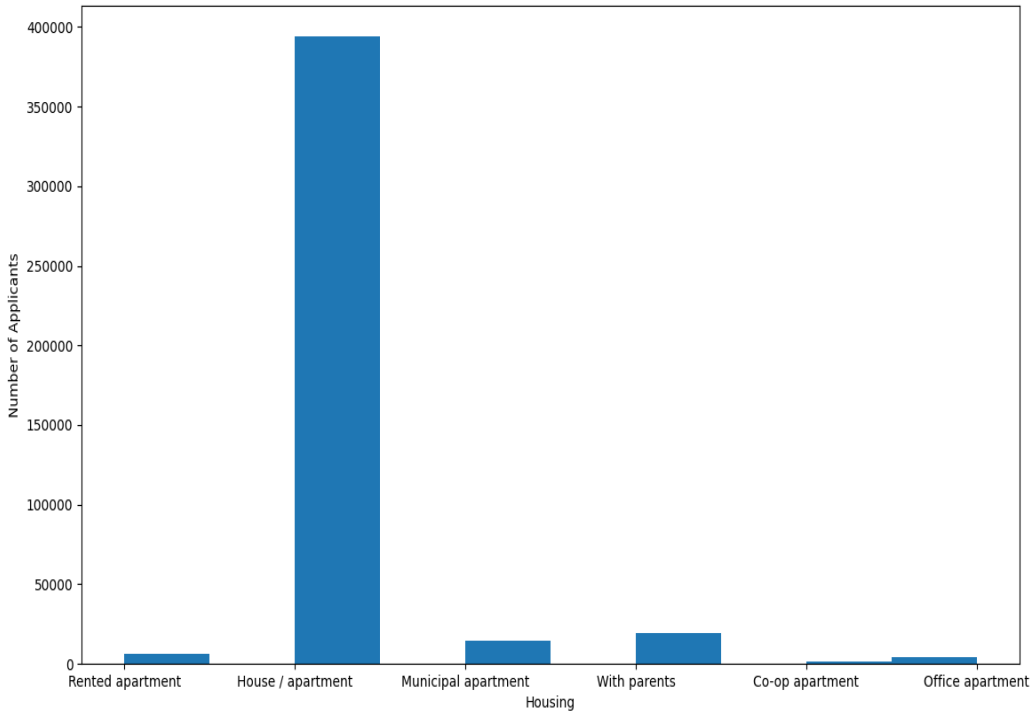
**Level of Education**

Most applicants are secondary education working class.

# UNIVARIATE ANALYSIS



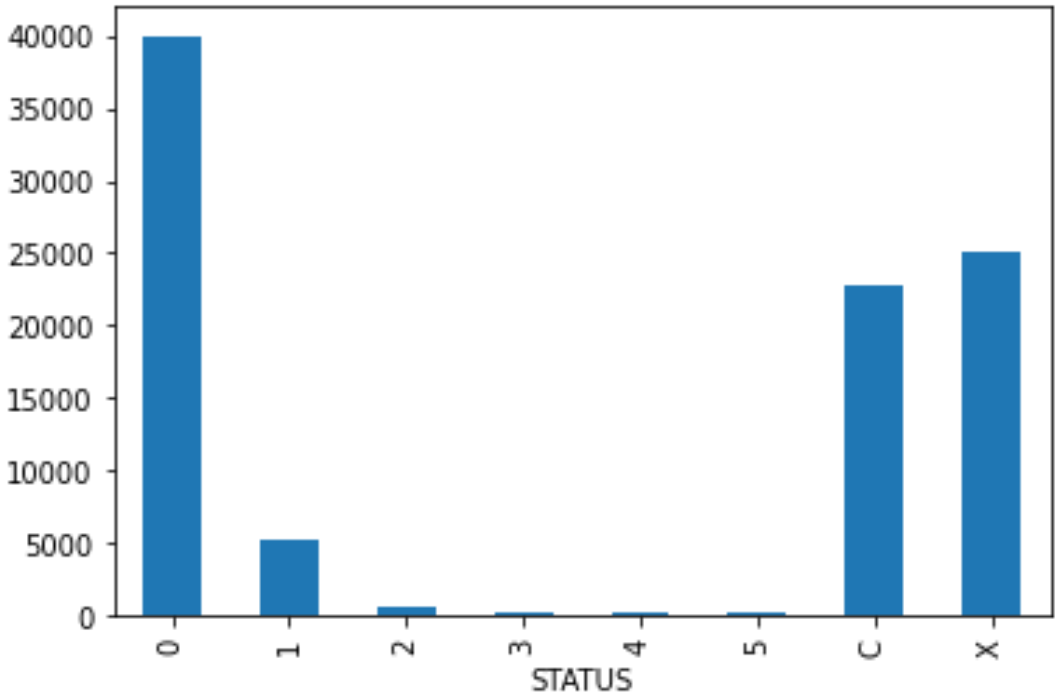

**Family Status**




**Housing Status**

Most applicants are married and own house/apartment.

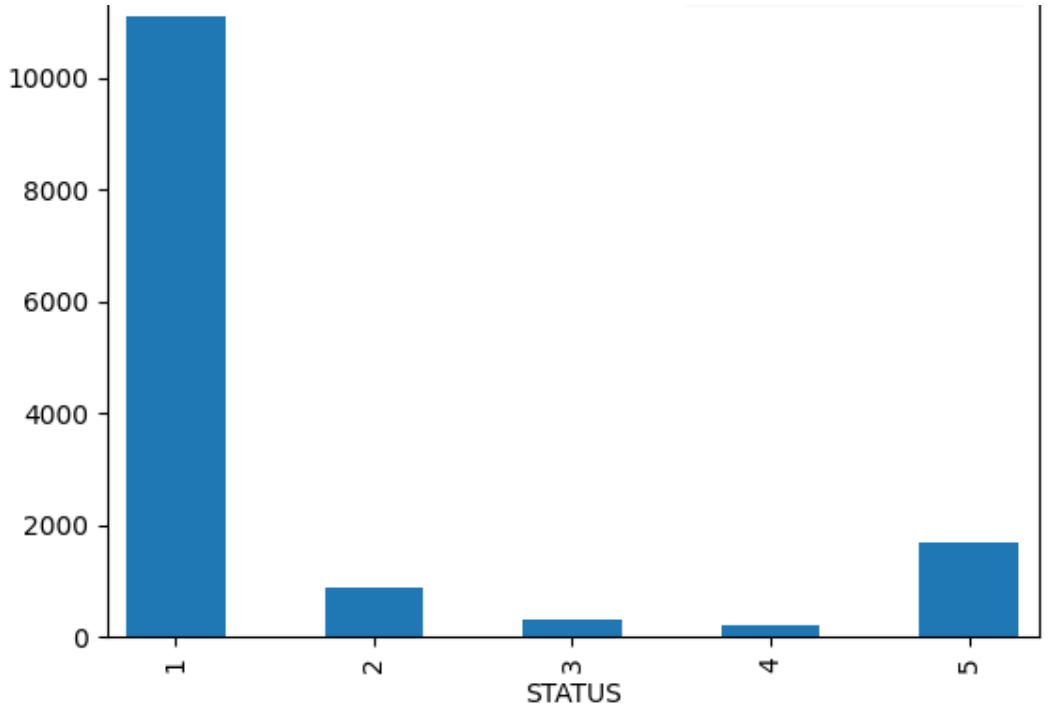
# UNIVARIATE ANALYSIS




**Payback window distribution**

Approx. equal number of months where clients make no purchases as those who pay back within the credit cycle (affects interchange income). Most overdue credit card loans are paid off within 30 days after the statement cycle.

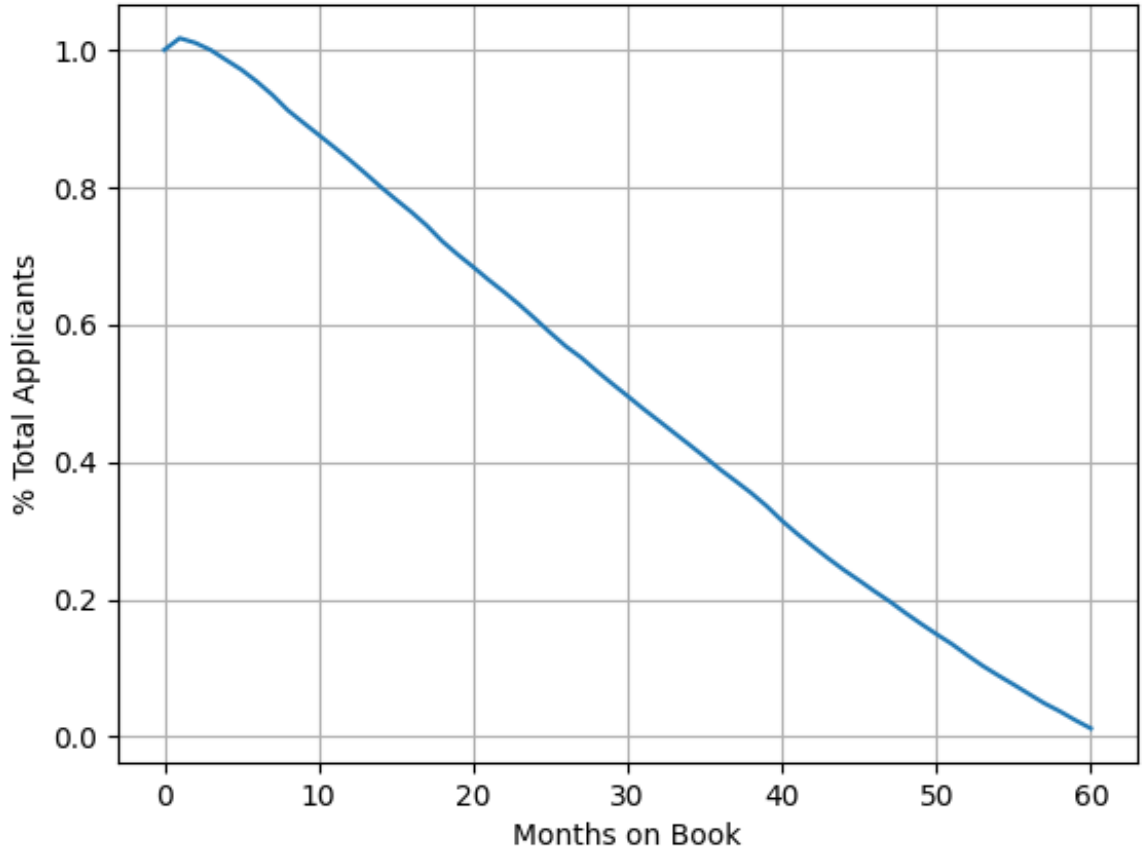
# UNIVARIATE ANALYSIS




**Overdue payback window distribution**

Number of statement cycles past due decrease exponentially with time. People seldom wait up to 150 days (Status 4) to repay their loan. Excess of 150+ days (Status 5) due to write offs being kept in the dataset for 60 months on book.

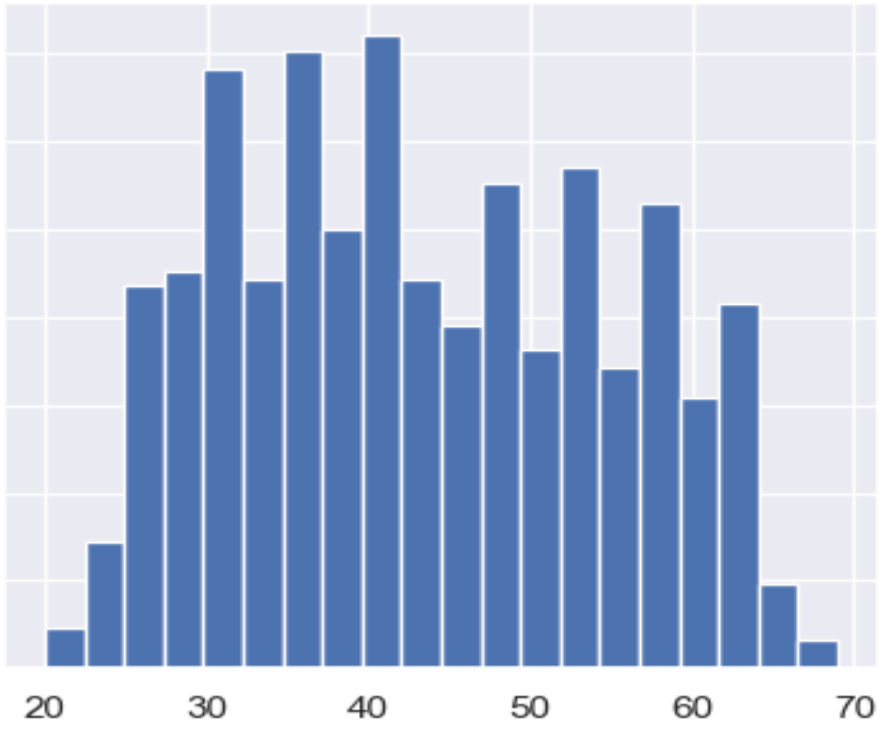
# UNIVARIATE ANALYSIS



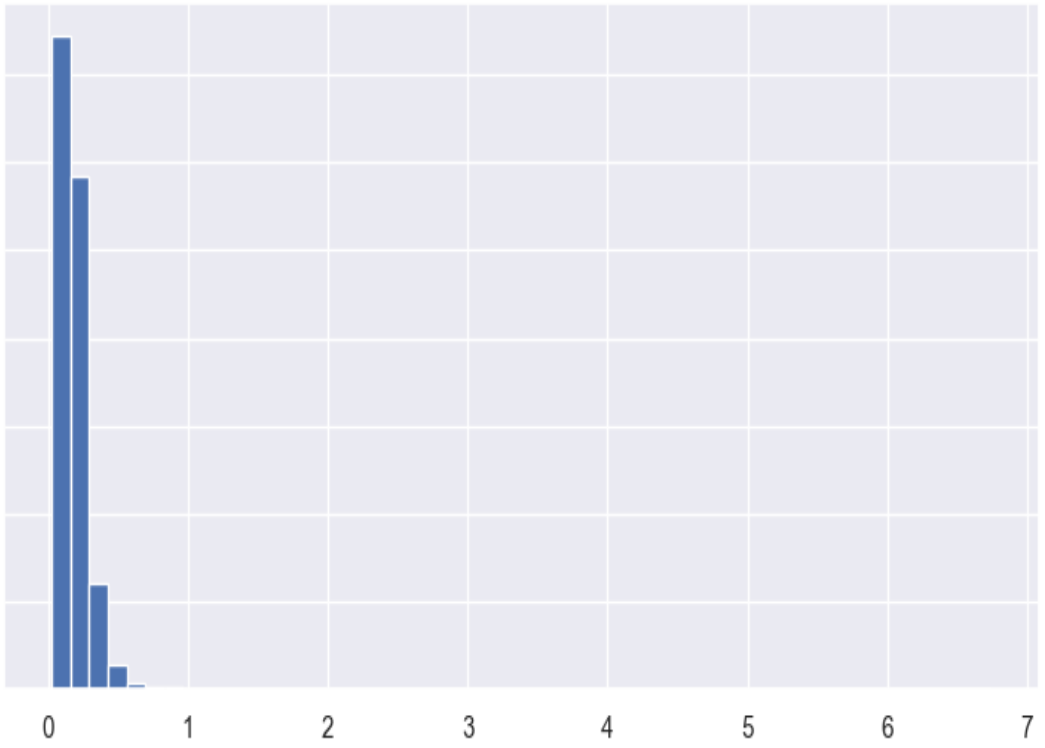

**Credit Record History Duration**

Linear decrease in credit history and number of applicants. 50% of applicants have < 2.5 years on book.
+100% account ratio at MOB1 caused by varying card activation times (real world occurrence).

# UNIVARIATE ANALYSIS



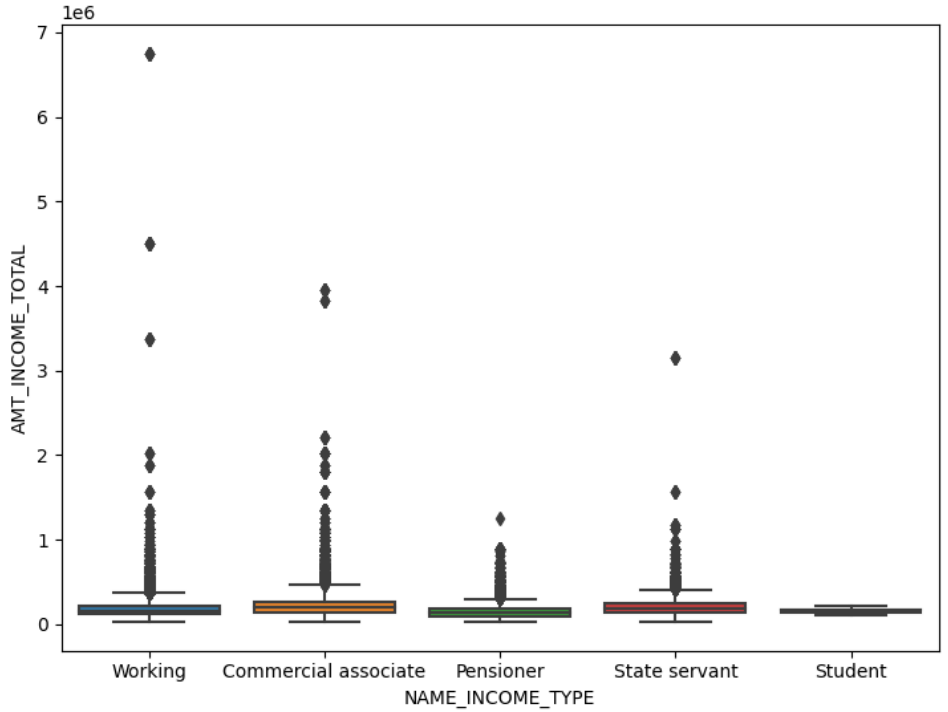

**Age Distribution**

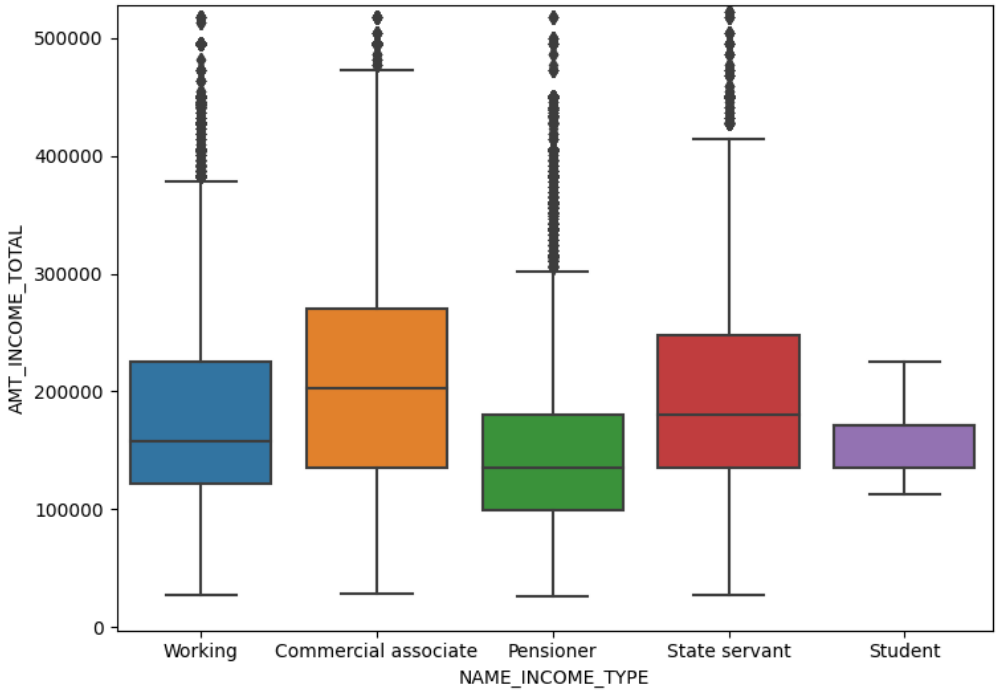



**Income Distribution**

Income represented in 100,000 intervals

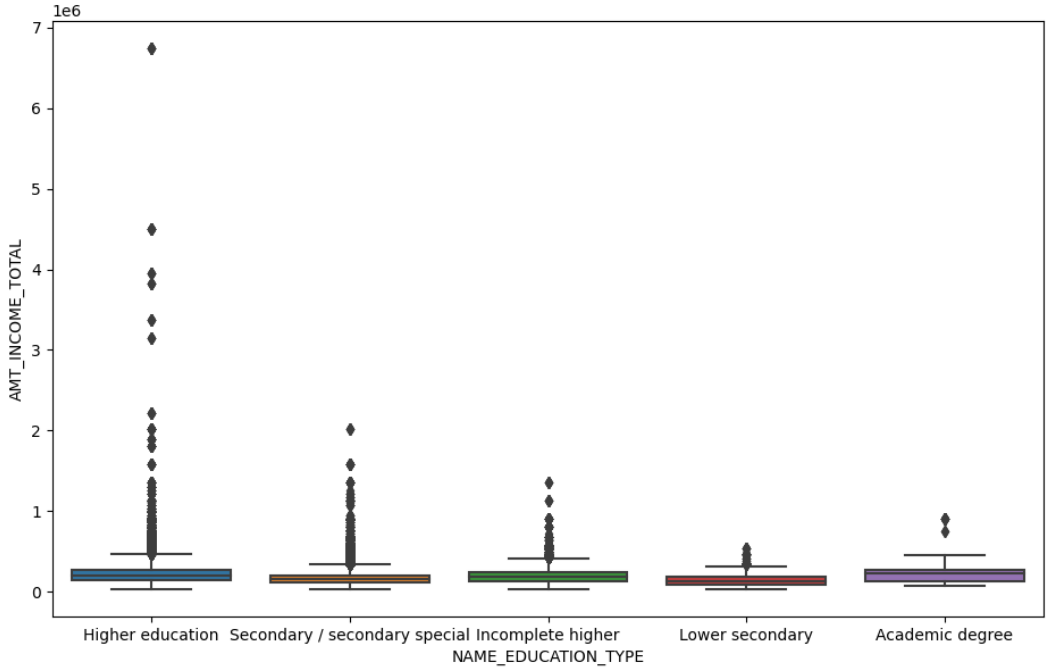
# BIVARIATE ANALYSIS

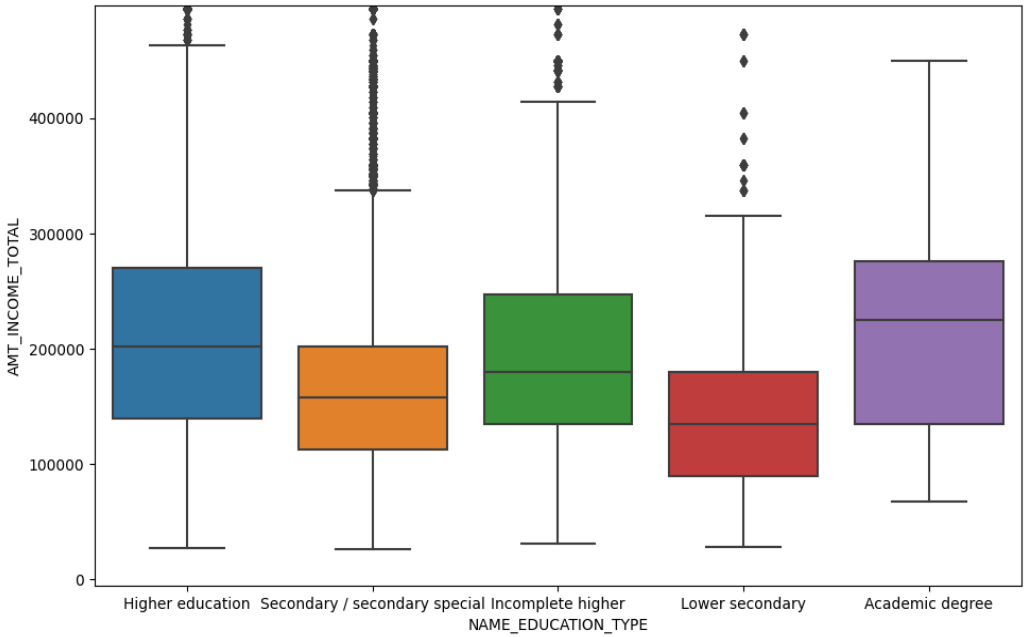



**Income Distribution by Income Type**

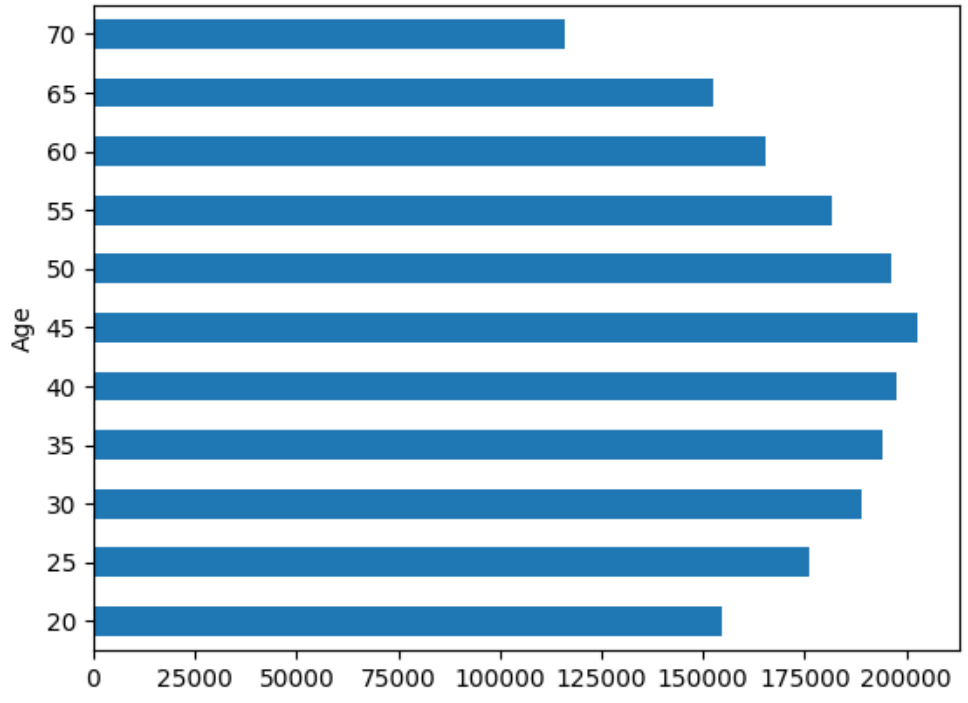
# BIVARIATE ANALYSIS



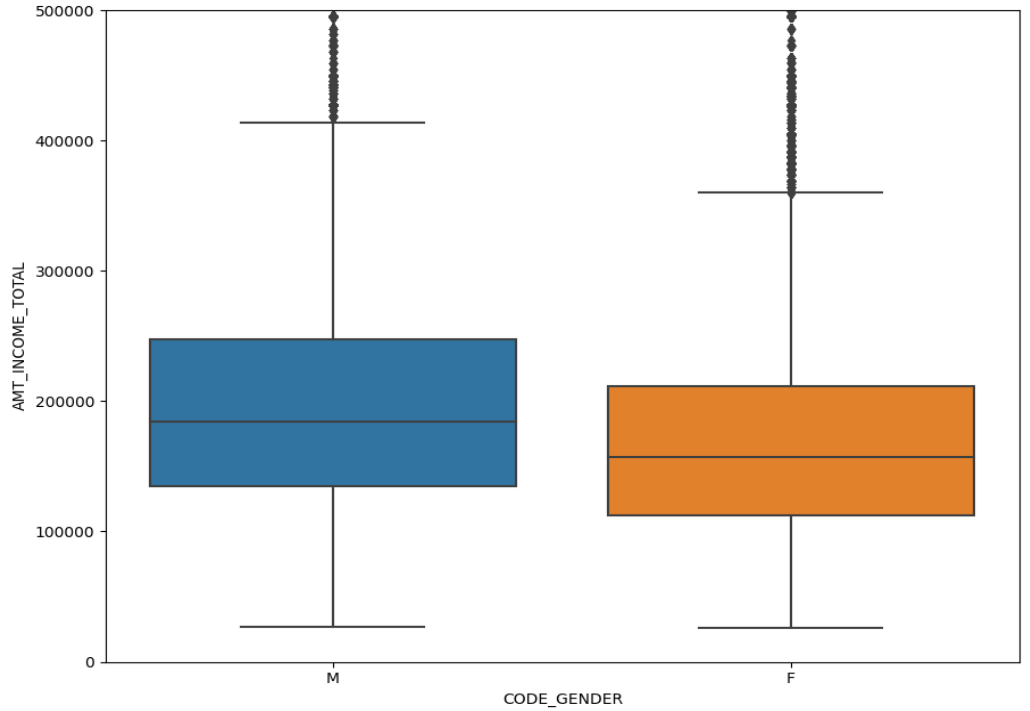




**Income Distribution by Education**

# BIVARIATE ANALYSIS

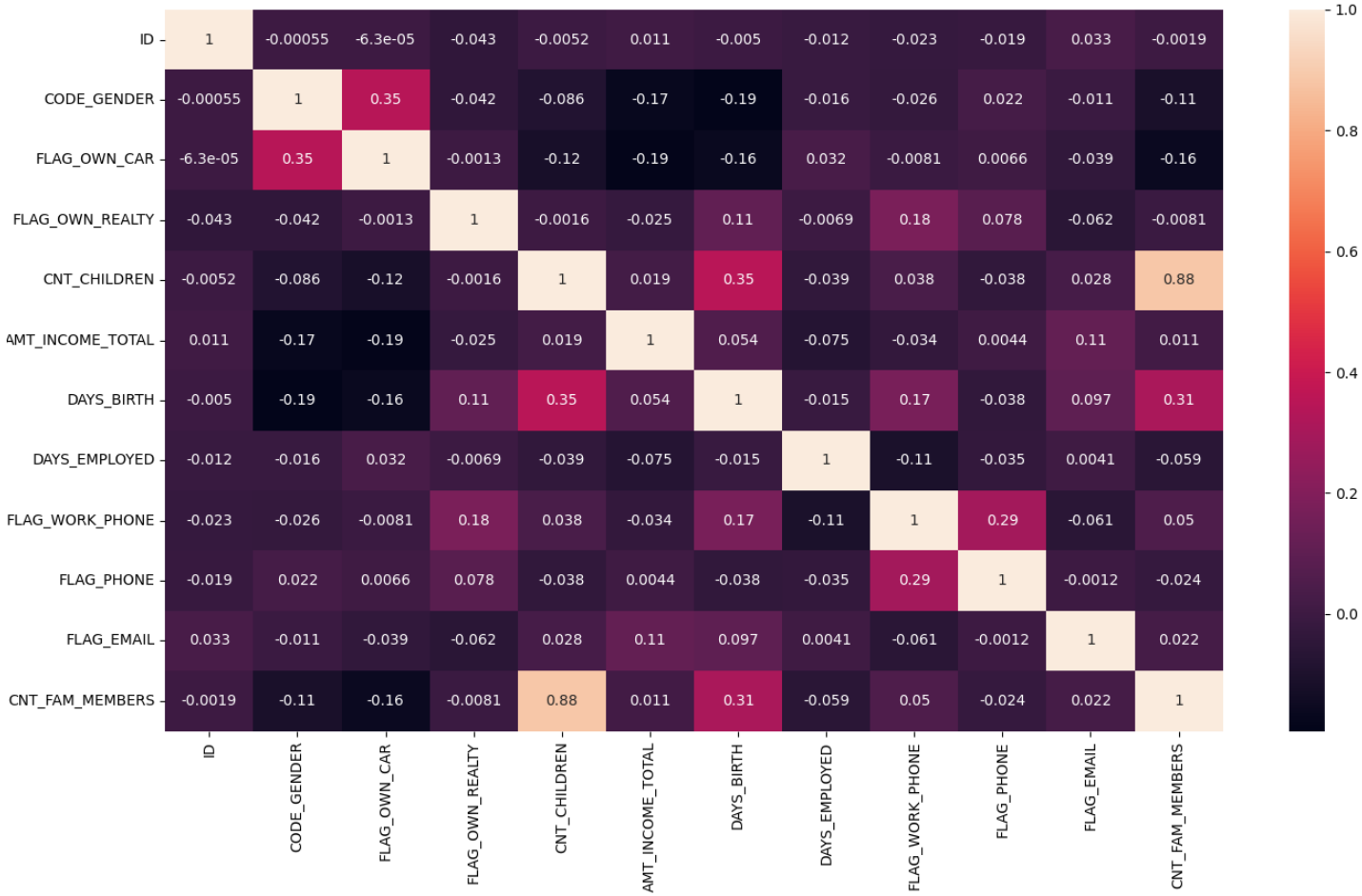



Income Distribution by Age




Income Distribution by Gender

# MULTIVARIATE ANALYSIS

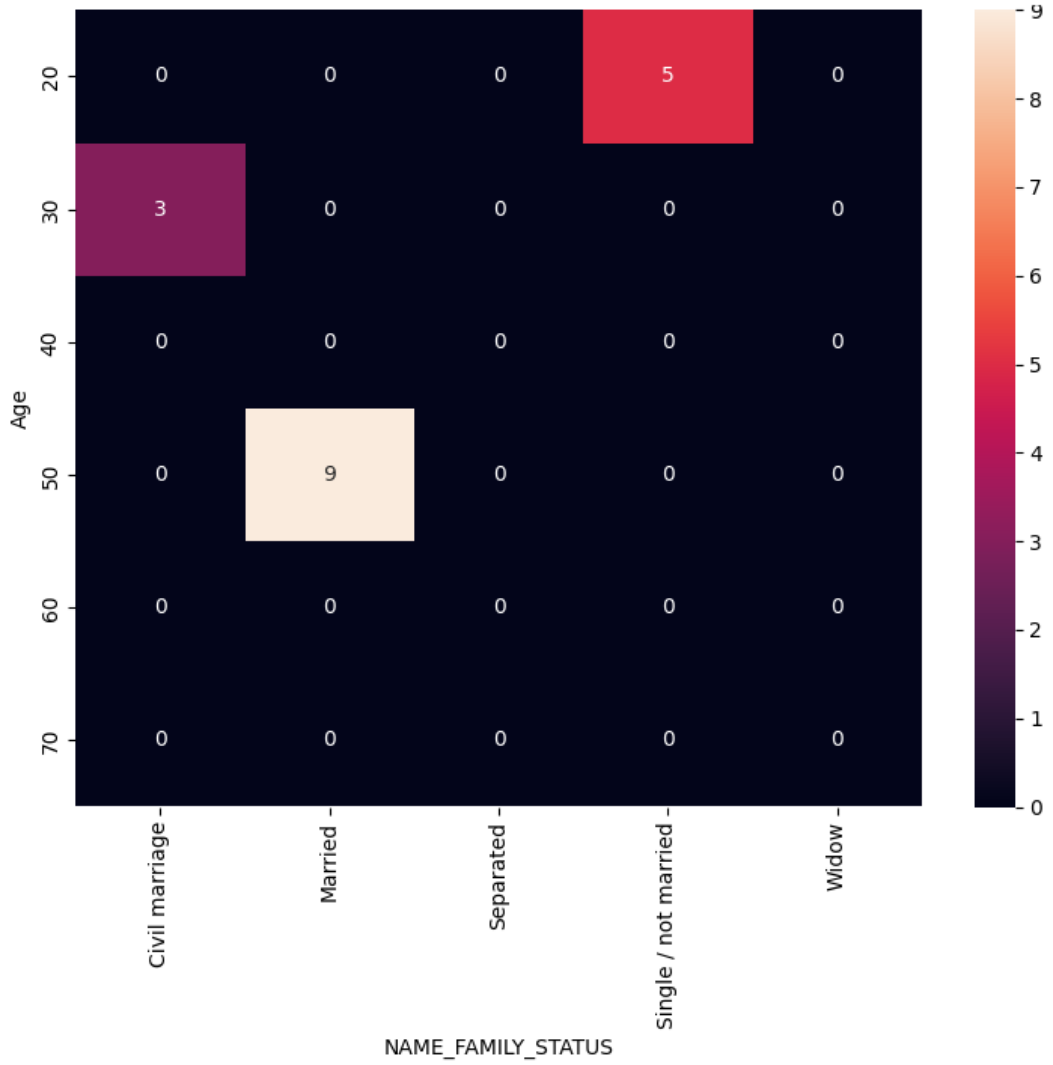



Number Children & Number of Family Members highly correlated.
Number Children & DAYS_BIRTH (Age * -1) mildly inversely correlated.
Gender & Income mildly inversely correlated.
Gender & Age mildly inversely correlated.

More kids = more family members.
People only have kids in their 30s-40s.
Men make more money on average.
Women tend to live longer than men.

# MULTIVARIATE ANALYSIS

## WHY DO STUDENTS EARN SO MUCH




Most students are over 30 years old and married. Household income from applicant's martial partner skewing the mean value. Alternatively, older students are receiving a corporate sponsorship while employed and earning salary.

Dataset is not representative of real world distribution where majority of students are 20 years old and single.

# THANK YOU