

A photograph showing a person's hands holding a smartphone. The screen of the phone displays various financial charts and data. In the background, there are several sheets of paper with more charts and graphs, along with some pens and a calculator. This imagery represents the analytical and data-driven nature of a bank loan case study.

Bank Loan Case Study

project by Yash Sangwan



CONTENT

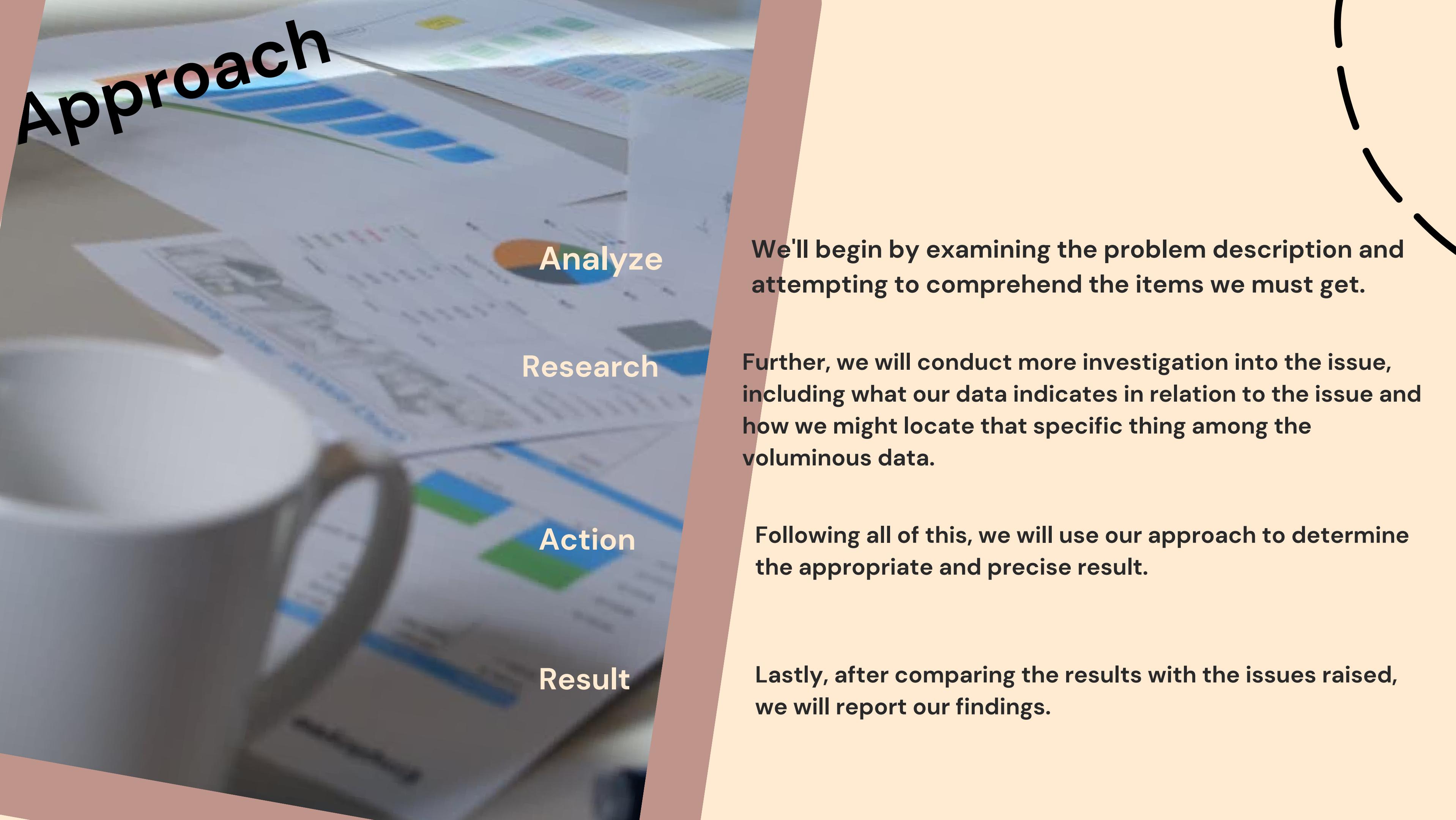
- 01 Overview
- 02 Approach
- 03 Tech Stack
- 04 Outcomes
- 05 Observation
- 06 Conclusion



Overview

In this case study, we'll examine certain customer-related information, including when we ought to approve or deny a loan application.

However, we first performed an EDA on our data, the results of which are available [here](#), before moving on to the deeper analytics portion.



Approach

Analyze

Research

Action

Result

We'll begin by examining the problem description and attempting to comprehend the items we must get.

Further, we will conduct more investigation into the issue, including what our data indicates in relation to the issue and how we might locate that specific thing among the voluminous data.

Following all of this, we will use our approach to determine the appropriate and precise result.

Lastly, after comparing the results with the issues raised, we will report our findings.

Tech Stack



General strategy for this analysis



Problem Statement

The bank problem statement inquires as to the reasons behind bank loan defaulters and seeks to differentiate between those who can and cannot repay the loan.

Insights

Two significant datasets, "application.csv" and "previous application.csv," with numerous unwanted columns and blank spaces, were given to us for this case study.

Analysis Approach

Due to the unorganised nature of the data, we first performed an EDA on both datasets, deleting all non-useful columns and filling in or removing blanks as appropriate for each row. Finally, we presented the findings from both datasets in the following slides.

Identifying Missing Data



Problem Statement

Identify the missing data and use appropriate method to deal with it

Insights

Our datasets comprise a total of 161 columns, many of which were unwanted.

Analysis Approach

We eliminated columns that were not significant and rows with more than 13% blank data.

The remaining rows' blank data was attempted to be filled in in accordance with the column's requirements or by observing other values above or below it, such as None, 0, NA, etc.

Identifying Outliers



Problem Statement

Identify if there are outliers in the dataset

Insights

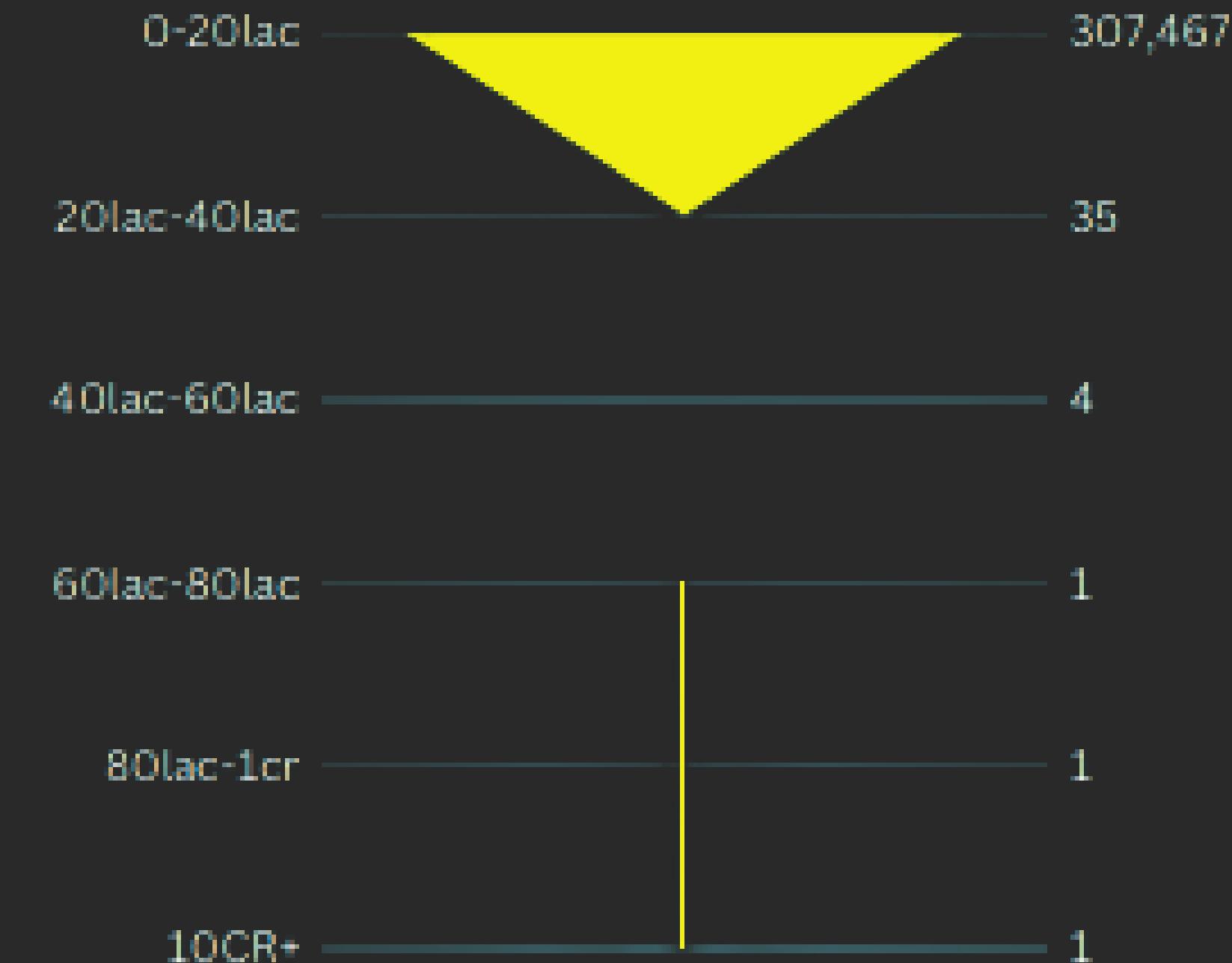
Because outliers can only be detected in numerical data, we have limited our study to those numerical data that are critically important for obtaining the correct insight.

Analysis Approach

Due to the unorganised nature of the data, we first performed an EDA on both datasets, deleting all non-useful columns and filling in or removing blanks as appropriate for each row. Finally, we presented the findings from both datasets in the following slides.

Insights

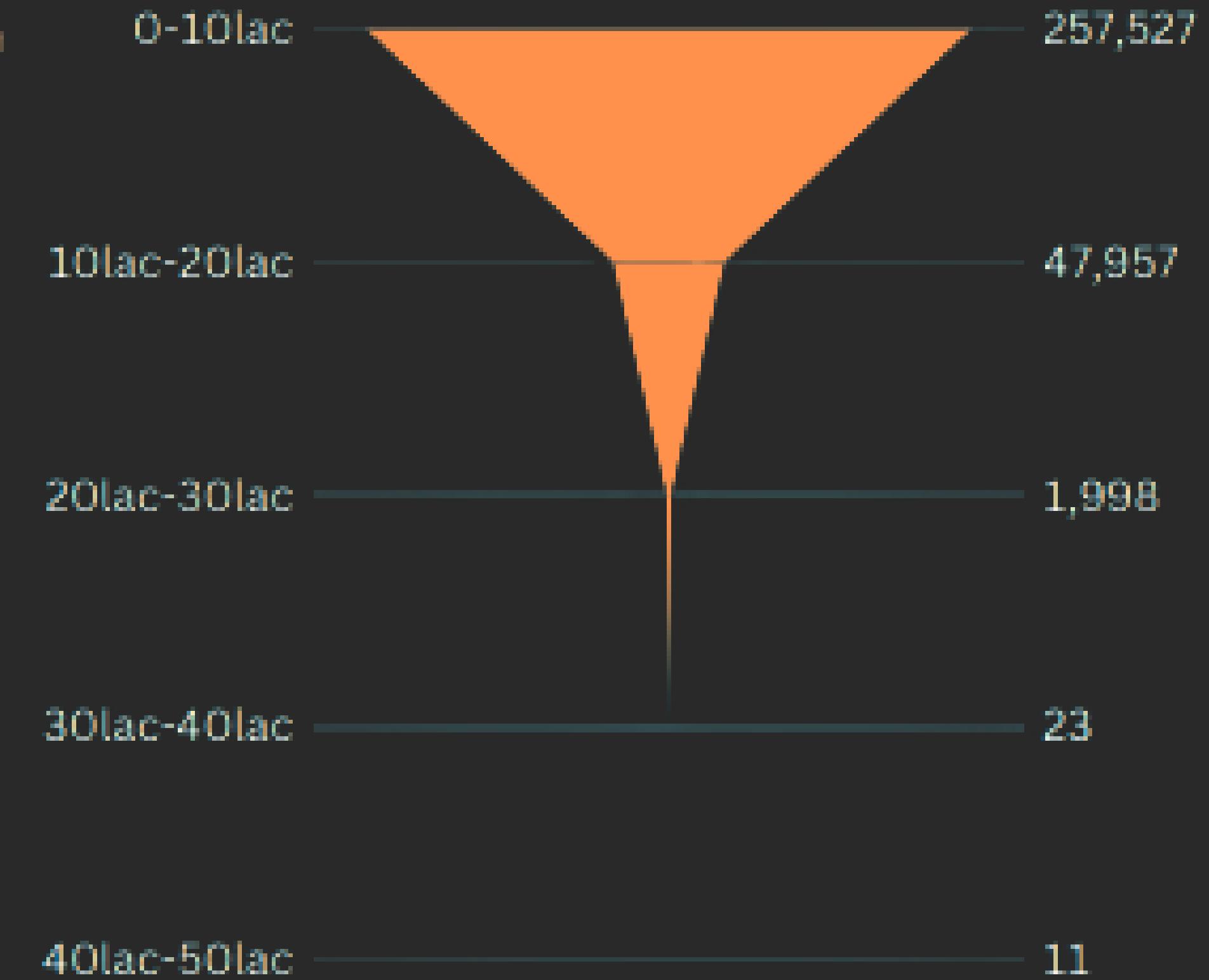
AMT_INCOME_TOTAL



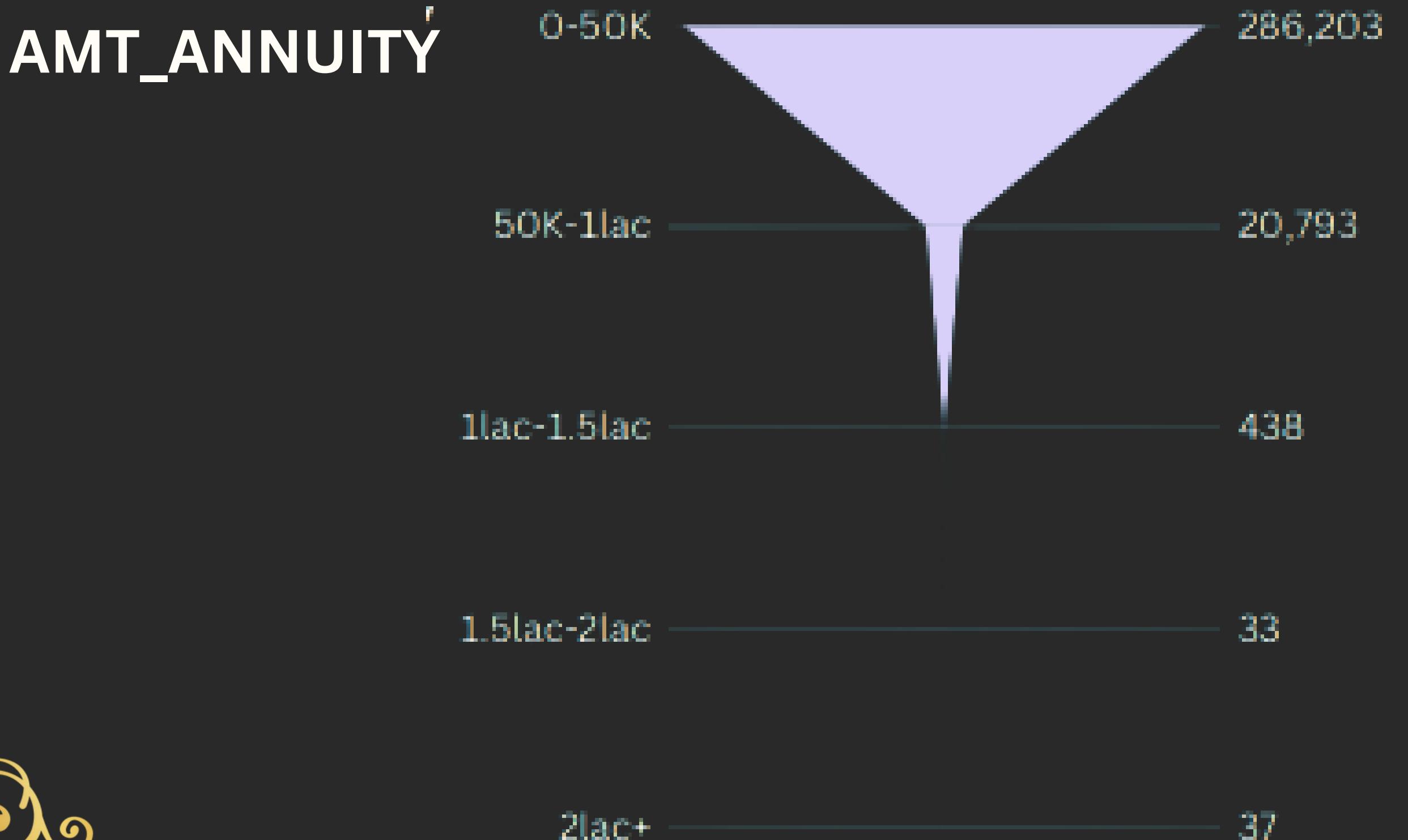
It is evidently an outlier because of the tiny yellow line beneath the large yellow triangle.

Insights

AMT_CREDIT



Insights



Analyzing Data Imbalance



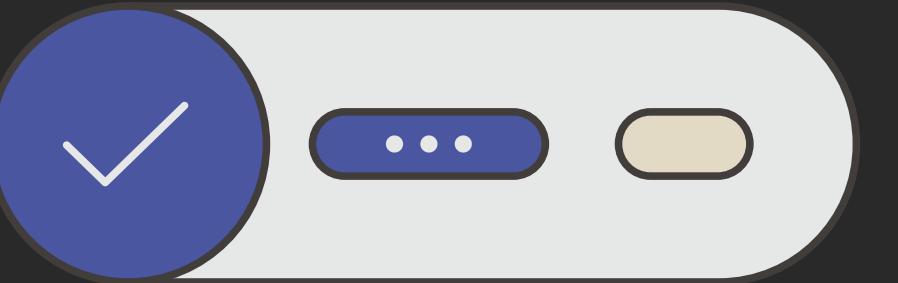
Insights

The total income column of the applicants, for example, displays data ranging from 25k to ending with 11Cr+. Also, the pattern of outliers in the previous presentations is another important indicator of the dataset's enormous imbalance.

Results

After doing all the computations, it is discovered that the imbalance ratio is: 10.6

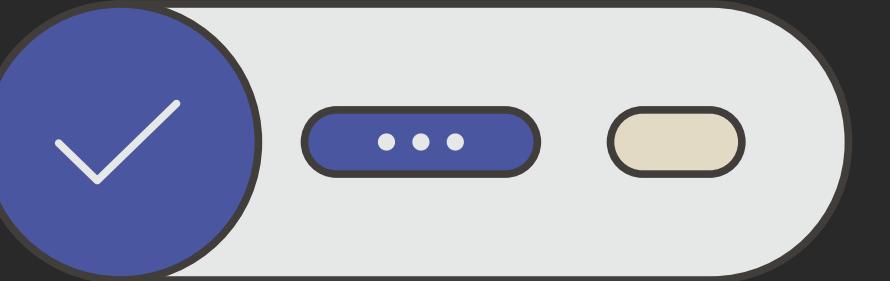
Univariate Analysis



Results

- Higher-income applicants are less likely to seek for loans.
- The credit for loan normally falls between \$45,000 and \$104,50000.
- Those between the ages of 35 and 50 have applied for the majority of the loans.
- The majority of people who apply for loans have between 0 and 8 years of job experience.
- Homeowners are more likely than non-owners to seek for loans.
- Married people have borrowed more money.
- Working folks have demanded more loans.
- Further loans have been requested by unaccompanied kids.

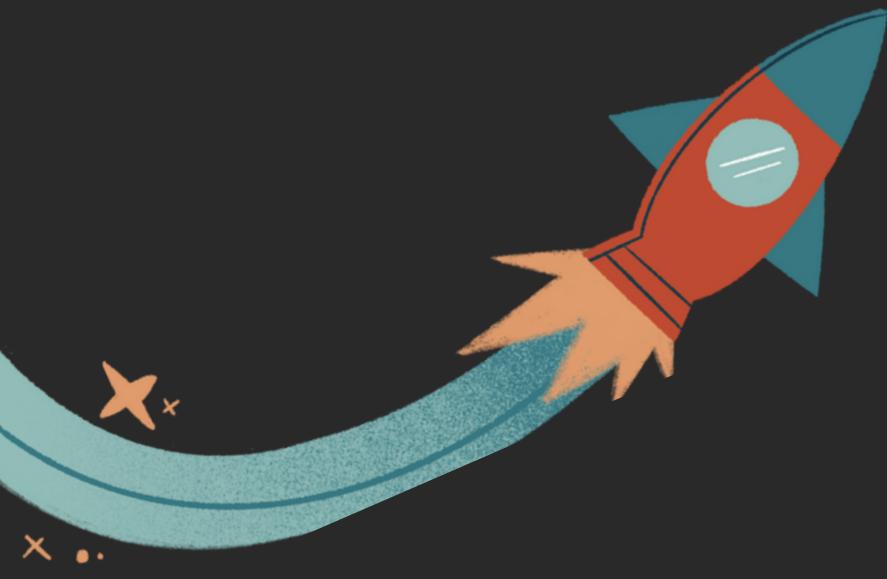
Bivariate Analysis



Results

- Civil marriage, marriage, and separation as a family status for academic degree education all have more credits than the other family statuses.
- The salary amount and family status are generally equal for education type "Higher education."
- For loan objectives such as "buying a home," "buying a land," "buying a new car," and "building a house," the credit amount is larger.
- Less credits are requested when money is needed for a hobby or third party.
- As co-op apartments are having trouble making their payments, banks should refrain from lending to this type of property. For successful payments, banks might concentrate mostly on dwelling

TOP 10 Correlations





Conclusion

FOR EFFECTIVE PAYMENTS, BANKS SHOULD CONCENTRATE MORE ON CONTRACT TYPES SUCH AS "STUDENT," "PENSIONER," AND "BUSINESSMAN" WITH DWELLING TYPES OTHER THAN "CO-OP APARTMENTS."



BANKS SHOULD PAY LESS ATTENTION TO THE REVENUE CATEGORY "WORKING," AS THESE TRANSACTIONS FAIL THE MOST FREQUENTLY.



BANKS SHOULD PAY LESS ATTENTION TO THE REVENUE CATEGORY "WORKING," AS THESE TRANSACTIONS FAIL THE MOST FREQUENTLY.

Thank You

