CREDIT SCORE FORECASTING

Malak Diab, Mohammed Hourani, Mayas Masalmeh

TABLE OF CONTENT

01

DATA INTRODUCTION

Problem Statement and the dataset

MALAK DIAB, MOHAMMED HOURANI, MAYAS MASALMEH

We’re working as a data scientists. Our dataset includes basic bank details with a lot of credit-related information. We’ll be analyzing the customer behavior based on their credit score in order to know the reason behind the low and good credit score in order to provide advices for the customers to increase their credit score.

PROBLEM STATEMENT

IN 8 MONTHS

100,000

Number of Rows

28

Number of Columns

12,500

Number of Customers

MONTHLY BALANCE

Represents the monthly balance amount of the customer (in USD)

MONTHLY IN HAND SALARY

Represents the monthly base salary of a person (in USD)

INTEREST RATE

Represents the interest rate on credit card (Ratio)

DELAY FROM DUE DATE

Represents the number of days delayed from the payment date

NUMBER OF DELAYED PAYMENTS

Represents the average number of payments delayed by a person

PAYMENT BEHAVIOR

Represents the payment behavior of the customer (in USD)

OUTSTANDING DEBT

Represents the remaining debt to be paid (in USD)

PAYMENT OF MIN AMMOUNT

Represents whether only the minimum amount was paid by the person

TOTAL EMI PER MONTH

Total amount that includes interest on the loan and a part of the principal amount

CREDIT SCORE

Represents the bracket of credit score (Poor, Standard, Good)

02

DATA ANALYSIS

Project Steps, Hypothesis, Data Preprocessing and Charts

MALAK DIAB, MOHAMMED HOURANI, MAYAS MASALMEH

PROJECT STEPS:

DATA EXPLORING

Overview of the Data

01

02

HYPOTHESIS

Hypotheses Generation

DATA PROCESSING

- Defining Categories
- Features Cleaning

03

04

CORRELATION

Discovering correlations between the features and the target

EXTRACTION

Feature Extraction and Selection with some Algorithms

05

06

CONCLUSION

Accept and reject hypothesis

01 DATA EXPLORING

MISSING VALUE & DUBLICATES

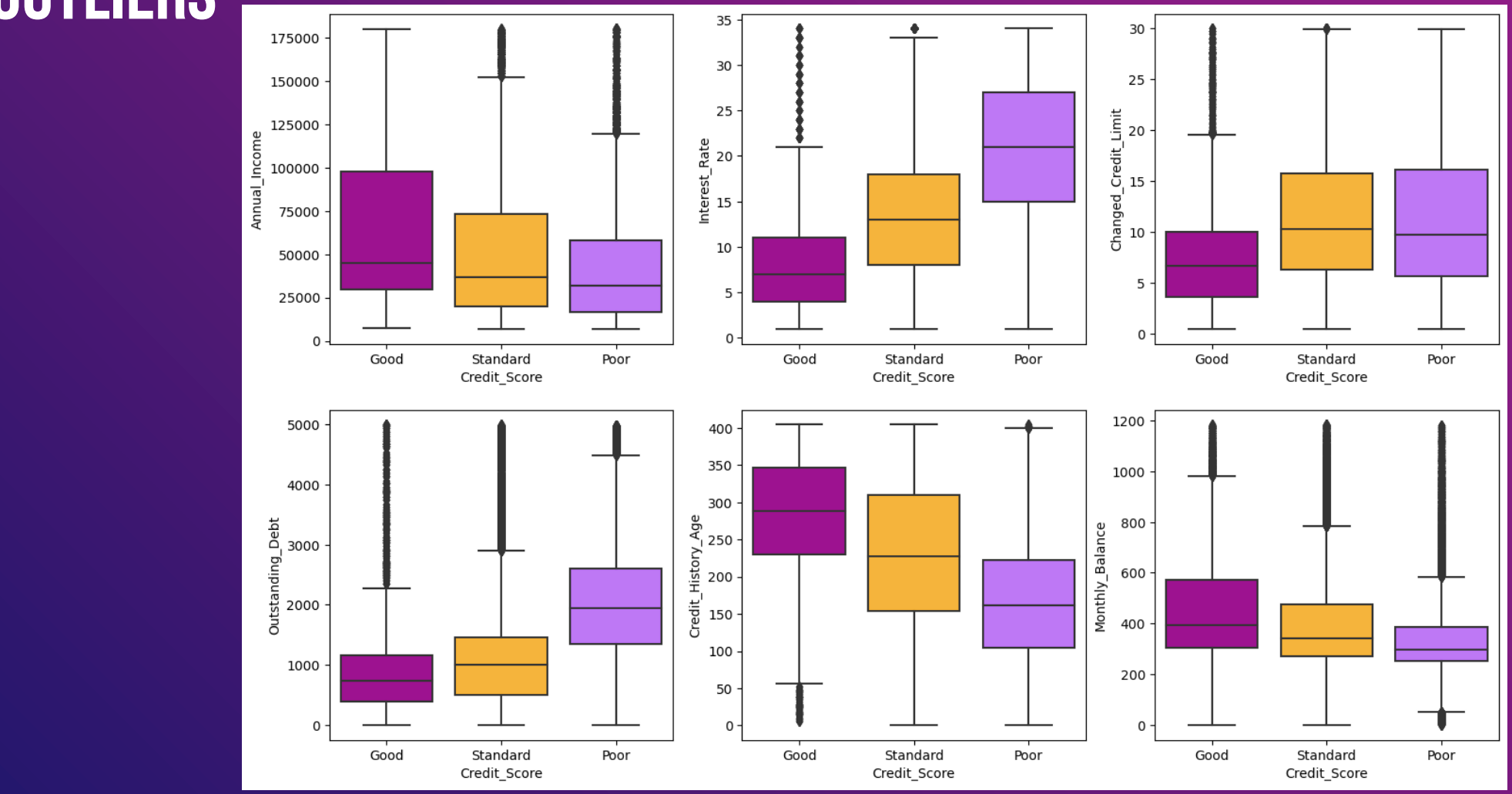
We don’t have any missing values or dublicates 😇

OUTLIRES & SKEWNESS

We devided the outlires feauters into categoris.

We’ll talk about categories in the next following slides

OUTLIERS


Annual_Income
Interest_Rate
Changed_Credit_Limit
Outstanding_Debt
Credit_History_Age
Monthly_Balance
Good
Standard
Poor
Credit_Score

02 HYPOTHESIS

H1 ✓

While the number of Loan increase, the degree of the Credit Score decrease.

H2 ✗

When the History age increase, the degree of the Credit Score increase.

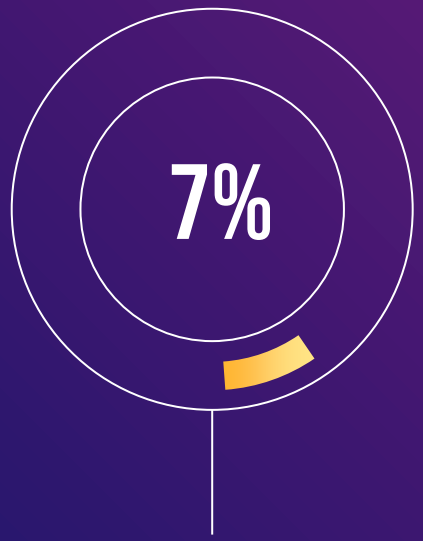
✗ H3

While the payment behavior is Low-Spent, the ratio of the (Good) Credit Score will be the highest.

✓ H4

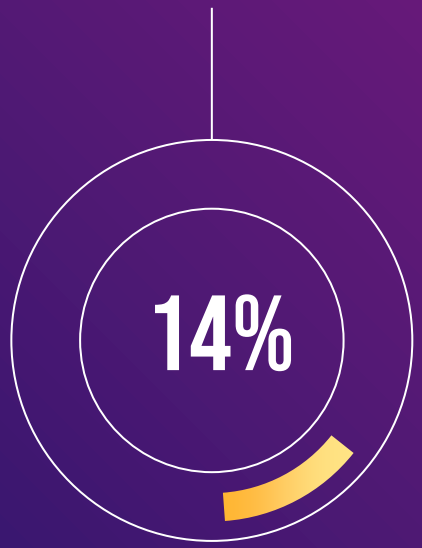
The increasement of the Annual Income results in the ratio of the (Good) Credit Score increment

AGE CATEGORY:



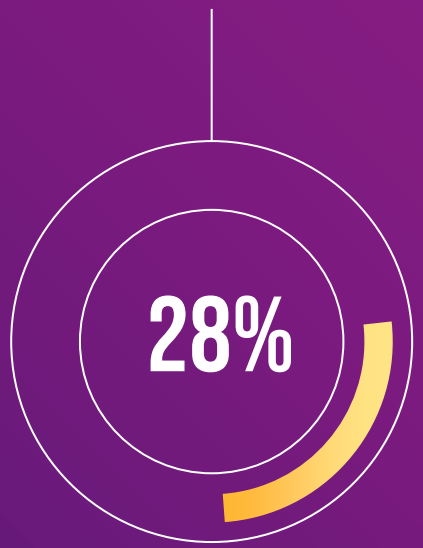
Younger than 18 Years

Between 18-24 Years



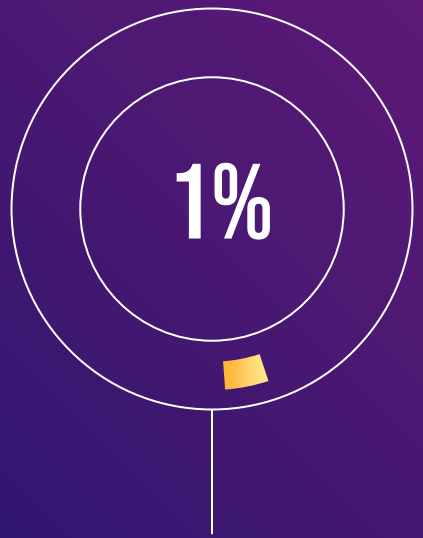
Between 24-30 Years

Between 30-40 Years



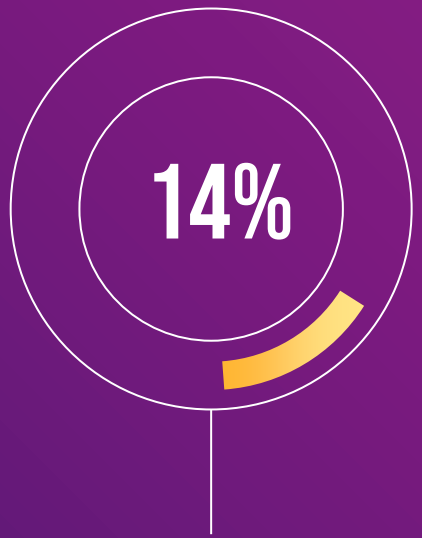
Older than 40 Years

HISTORY AGE CATEGORY:



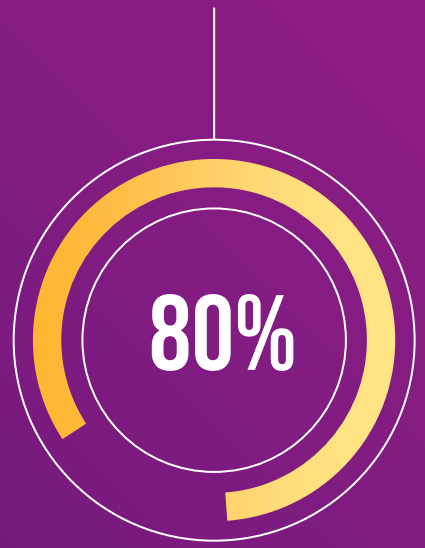
Less than a Year

Between 1-5 Years

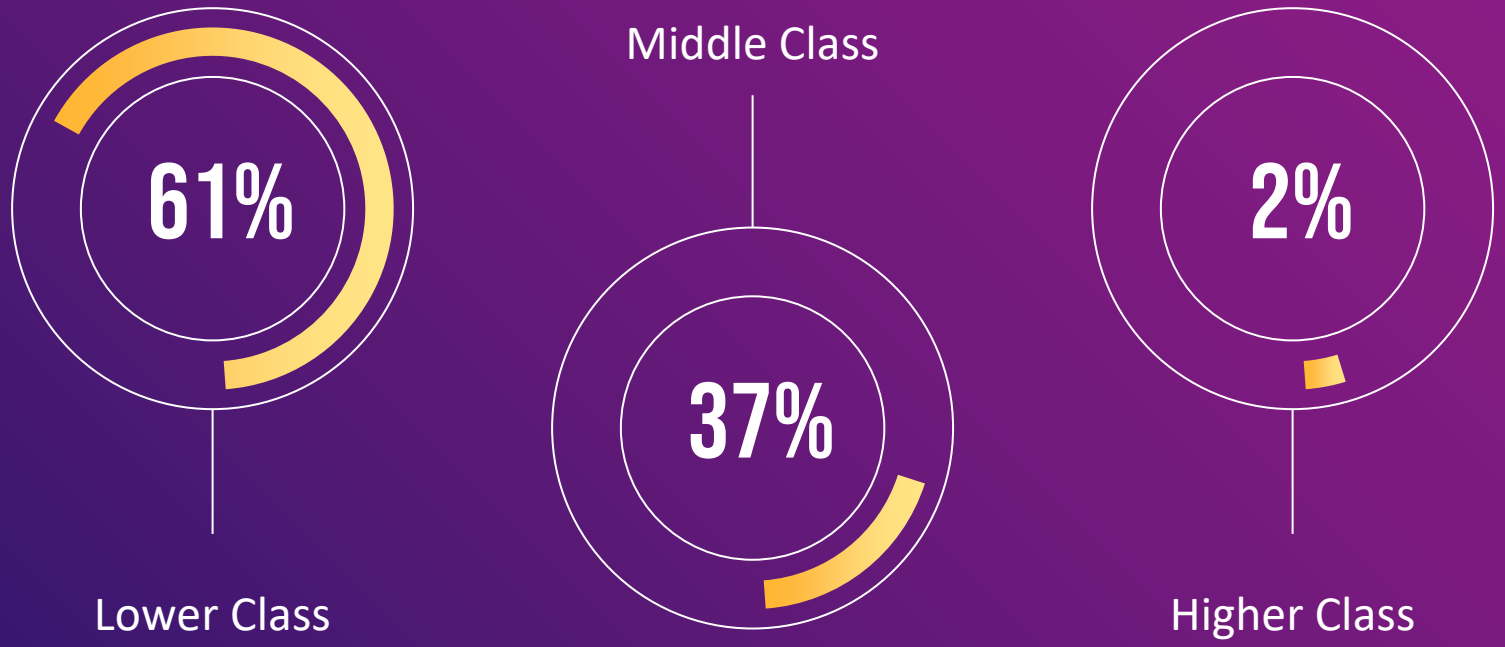


Between 5-10 Years

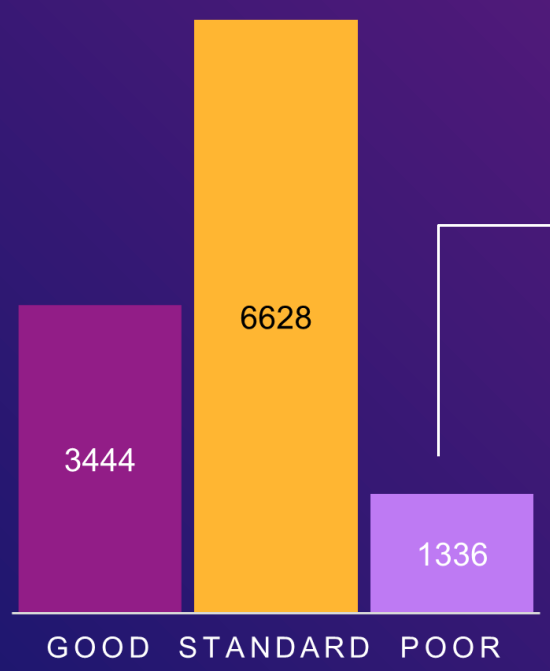
Older Than Ten Years



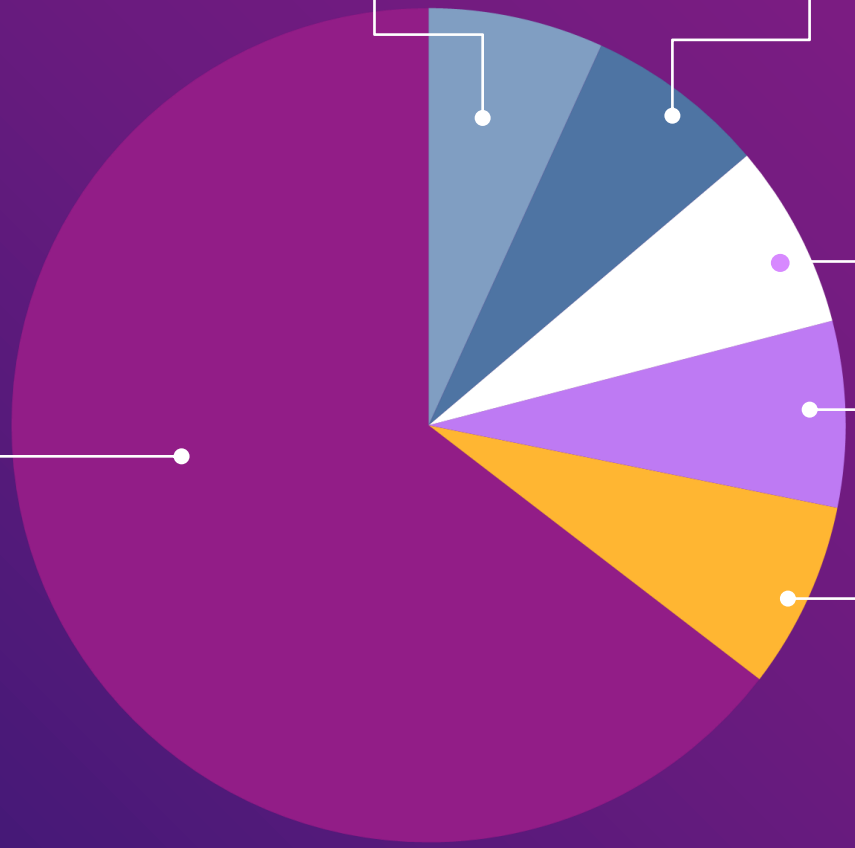
COMMUNITY CATEGORIES:



TYPE OF LOANS



65%
No Loans



7%
Payday Loan

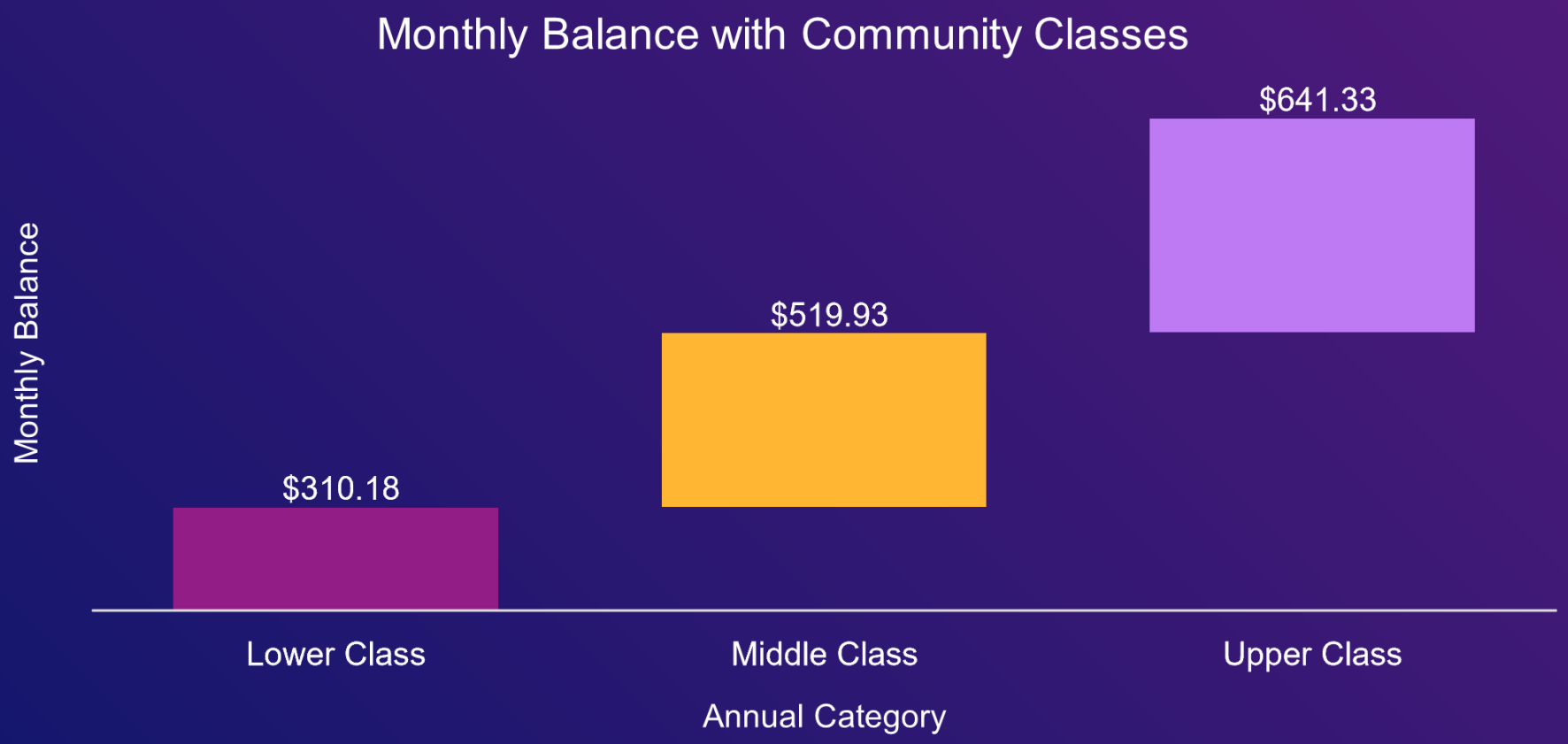
7%
Debt Consolidation Loan

7%
Student Loan

7%
Personal Loan

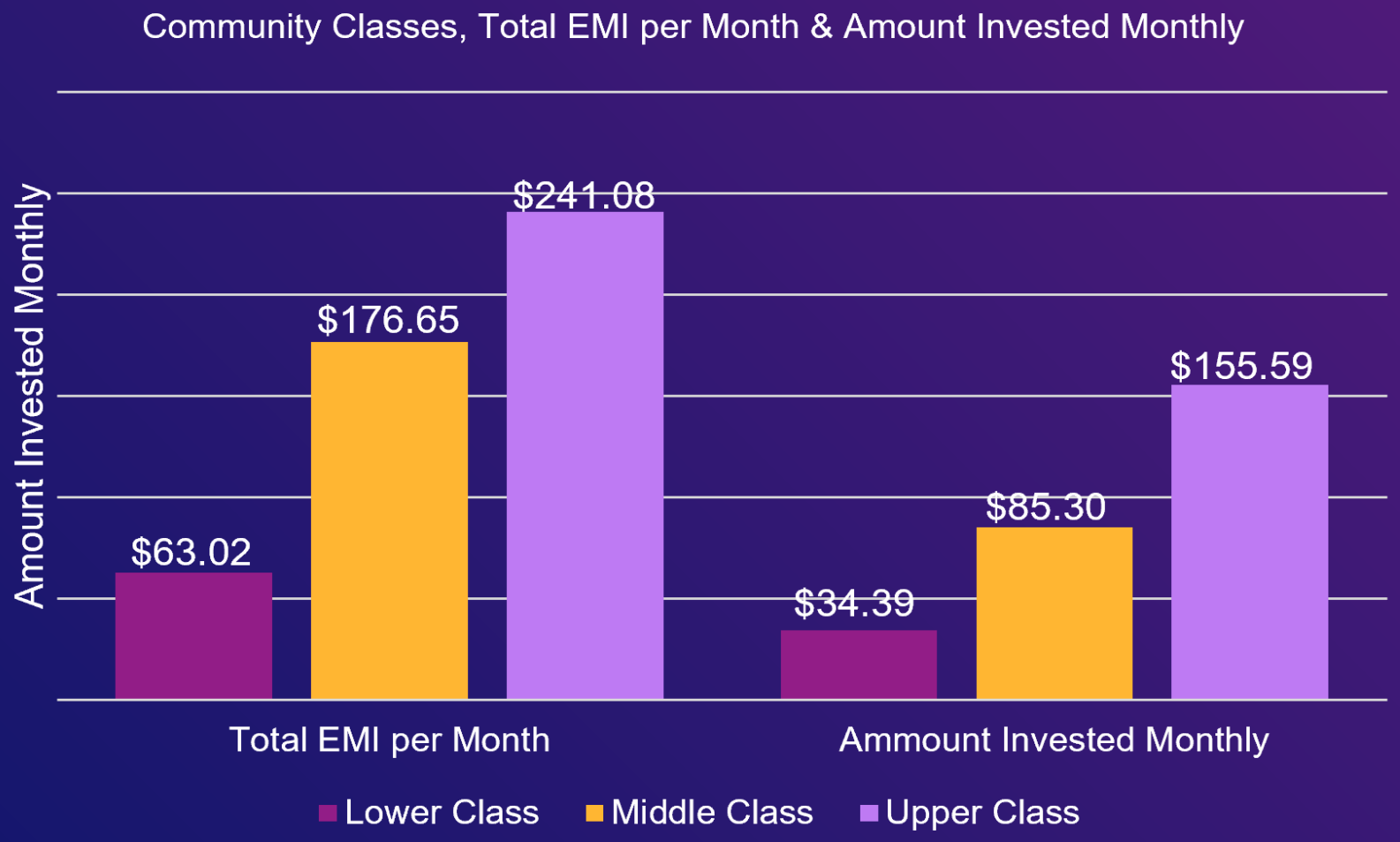
7%
Credit Builder Loan

04 RELATIONSHIPS:



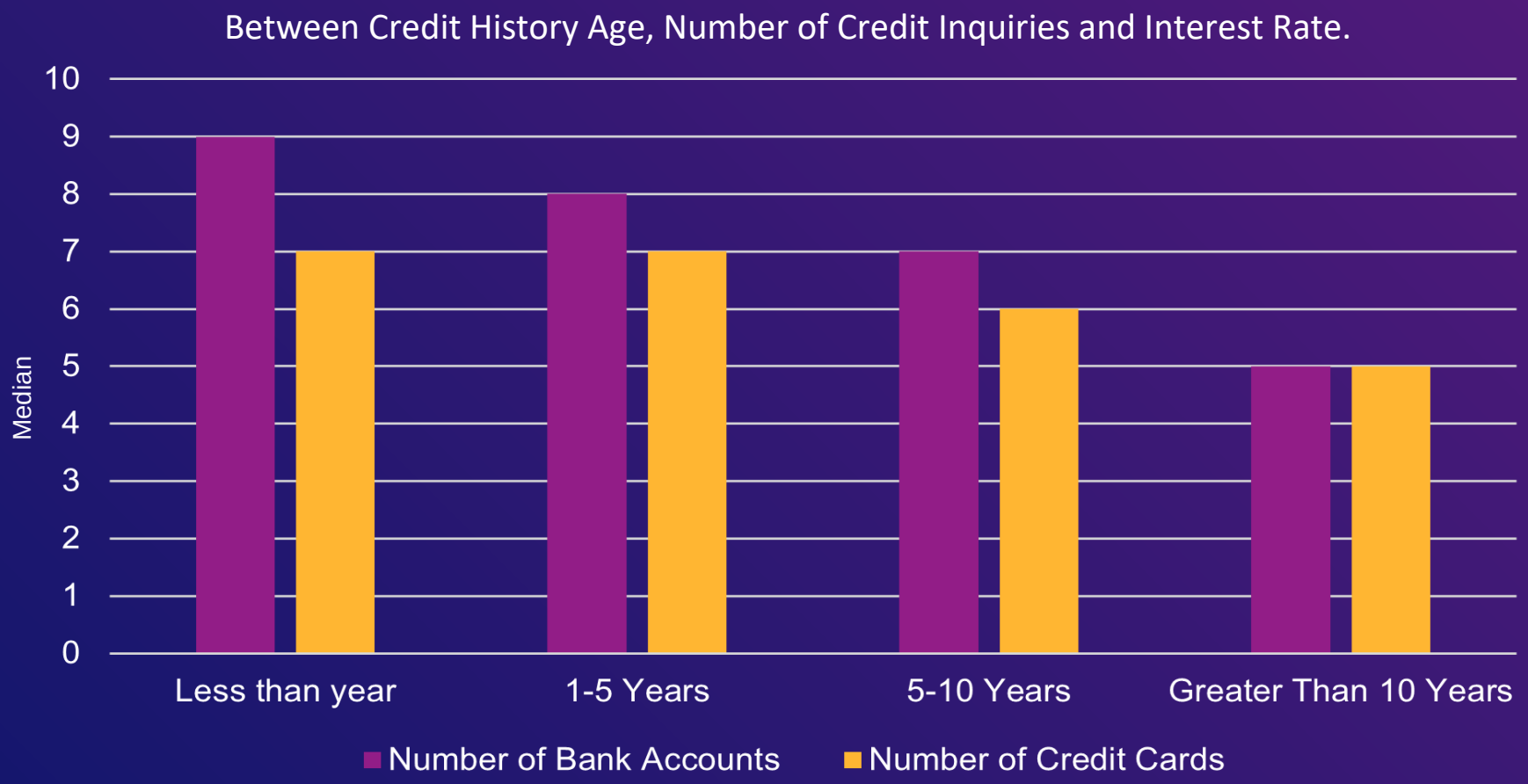
↑ Community Classes
↑ Monthly Balance

RELATIONSHIPS:



↑ Community Classes
↑ Total EMI per month
↑ Amount invested monthly

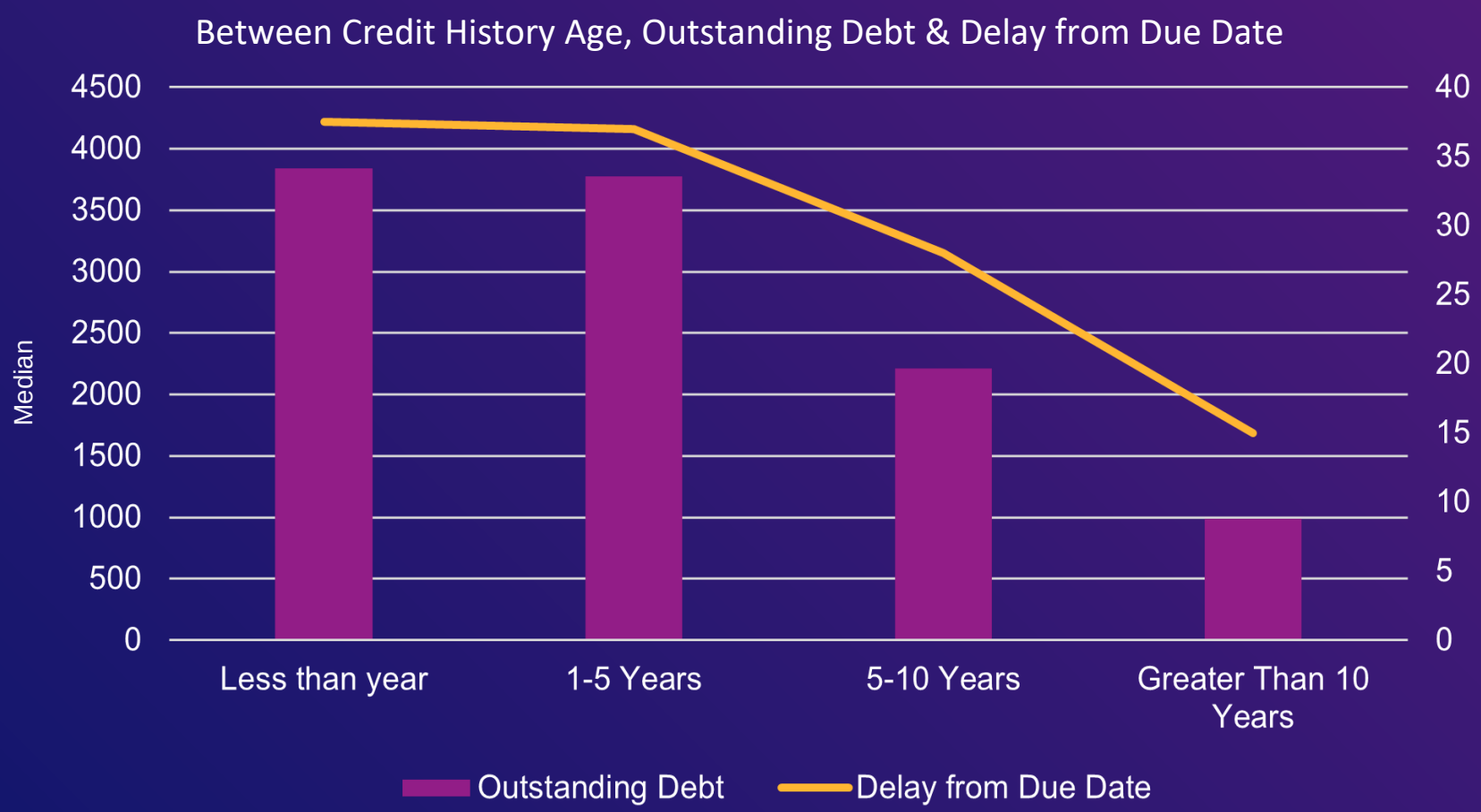
RELATIONSHIPS :



↑ Credit history age
↓ Number of bank accounts & Credit cards.

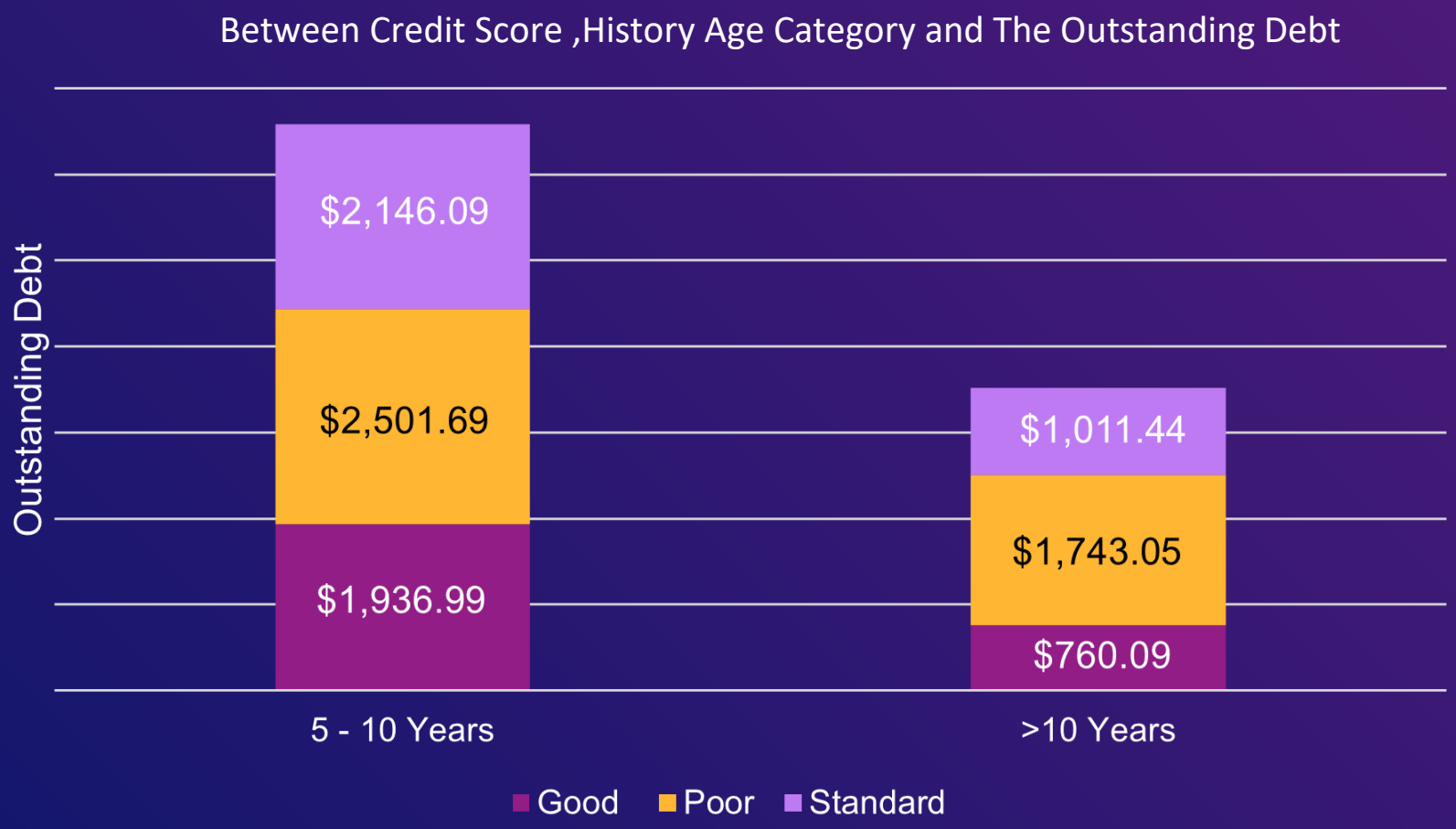
↑↓ Number of bank
↑↓ Number of credit card.

RELATIONSHIPS:



↑ Credit history age
↓ Delay from due date &Outstanding debt.

RELATIONSHIPS:

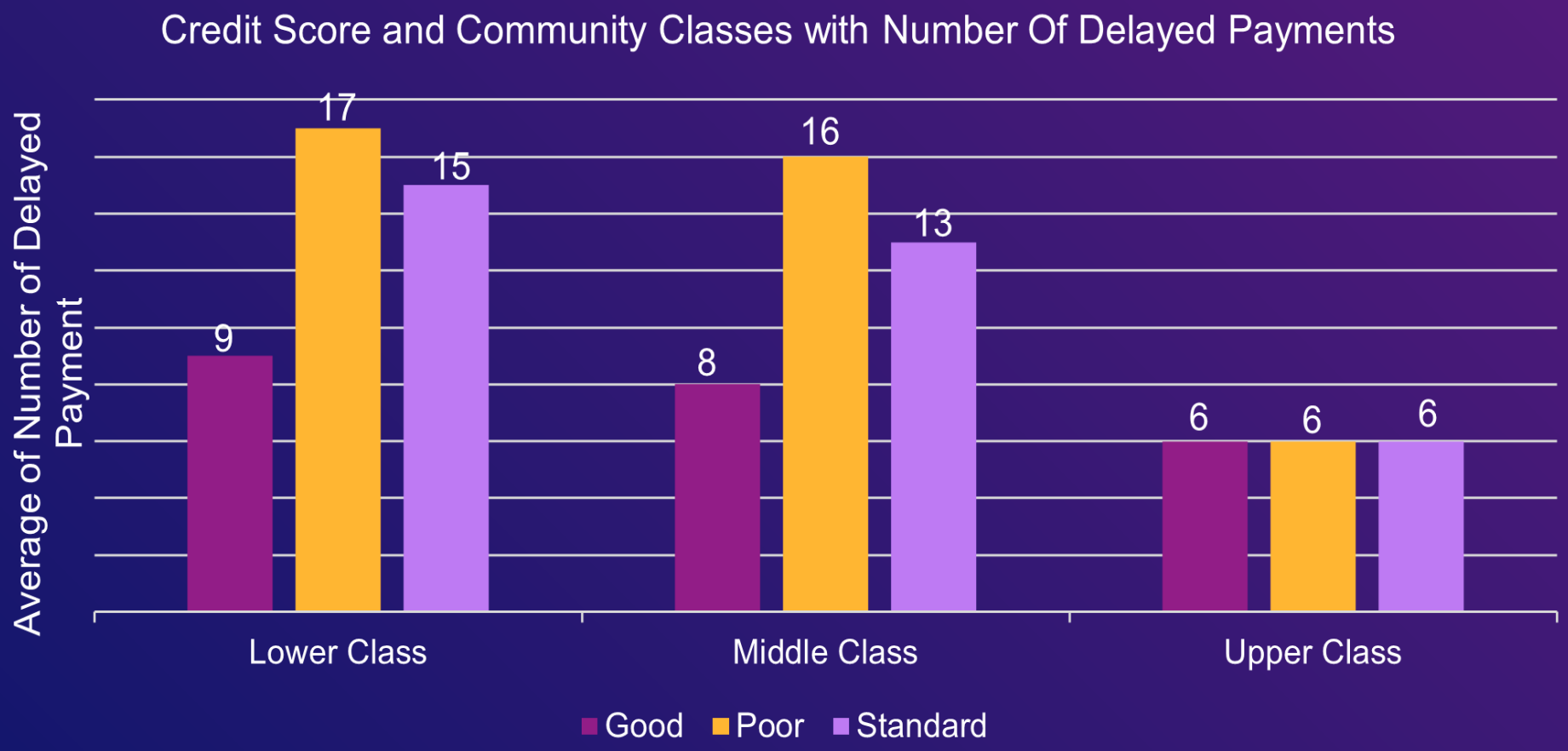


↑ History Age ↓ Outstanding Debt

Hypotheses Number 2 is rejected depending on this chart, because the history age category isn't affecting the credit score.

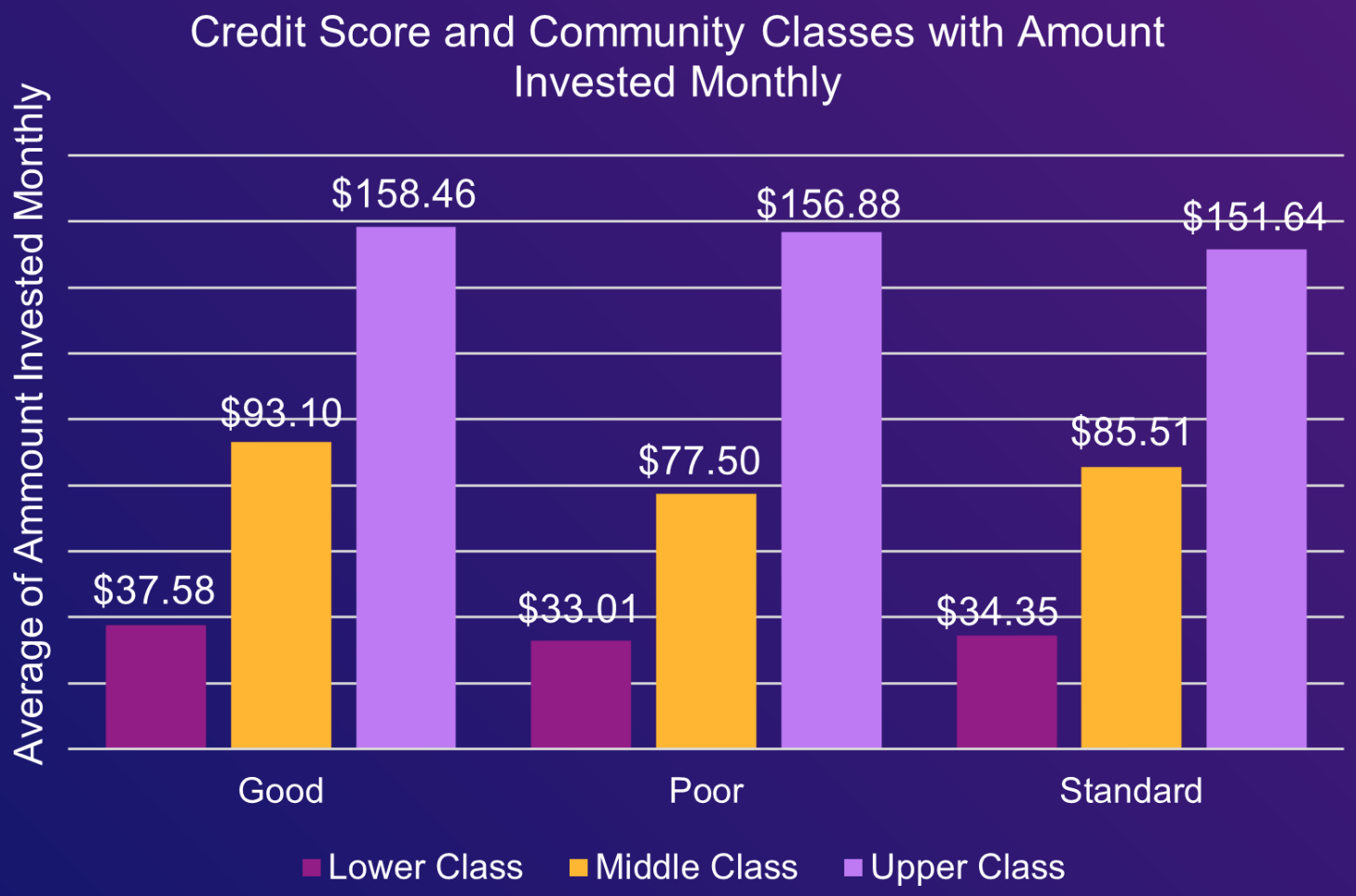
06

RELATIONSHIPS:



People in Lower& Middle class have more delays than people in Upper class.

RELATIONSHIPS:



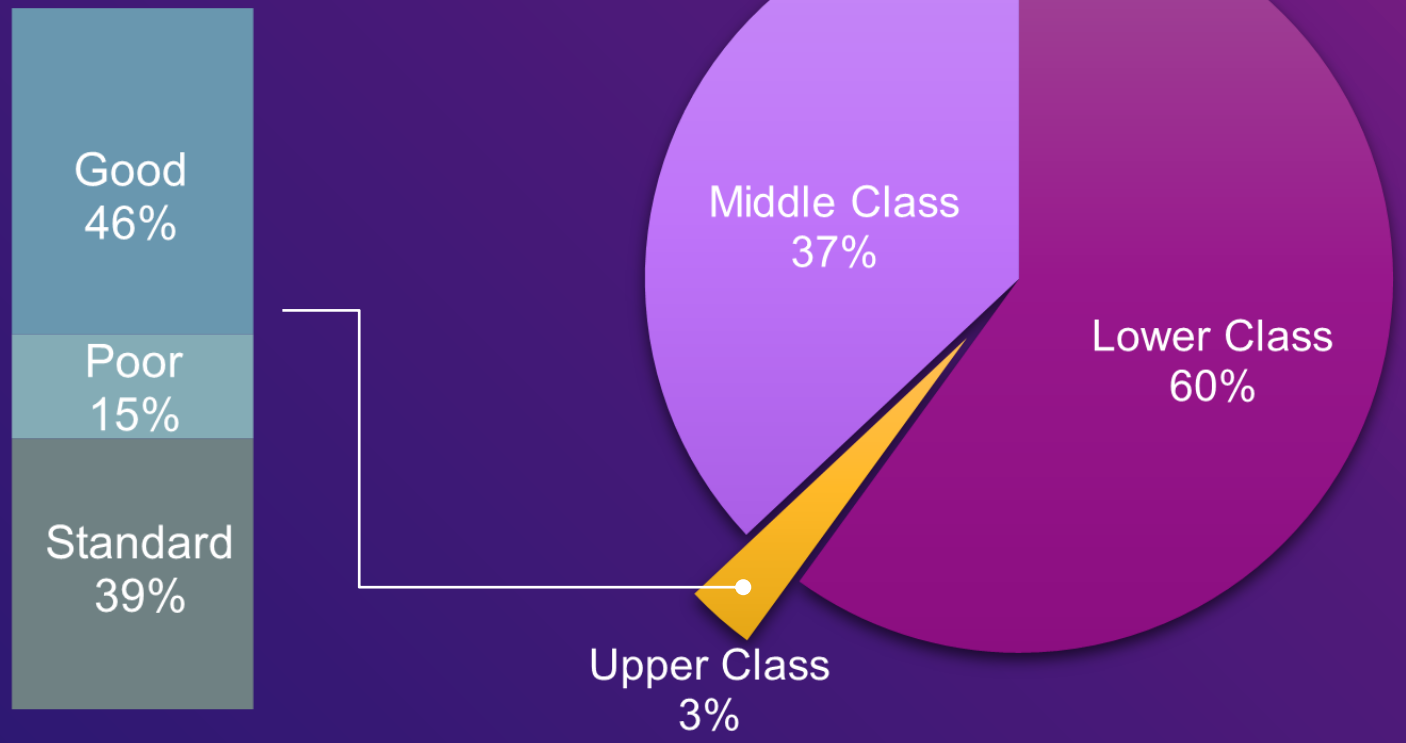
↑ Community Class
↑ Number Invested Monthly

RELATIONSHIPS:

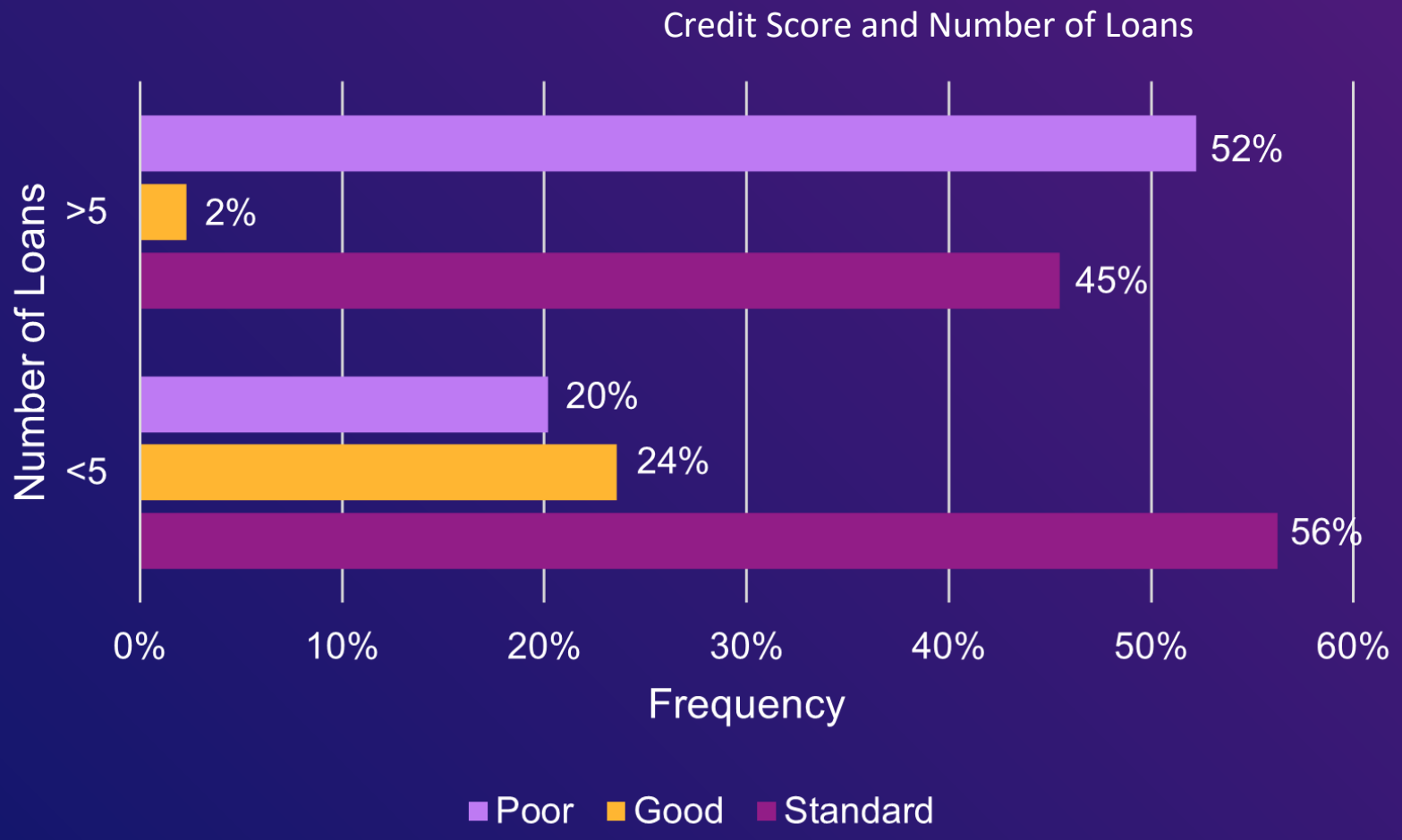
↑ Community Class ↑ Credit Score

Hypotheses number 4 is accepted, as this chart shows that the increment of the Annual Income results in the increasement of Good Credit Score Ratio

06



RELATIONSHIPS:



People with Number of Loan lower than 5 have higher Standard and Good Credit score frequency .

People with Number of Loan higher than 5 have higher Poor Credit score frequency.

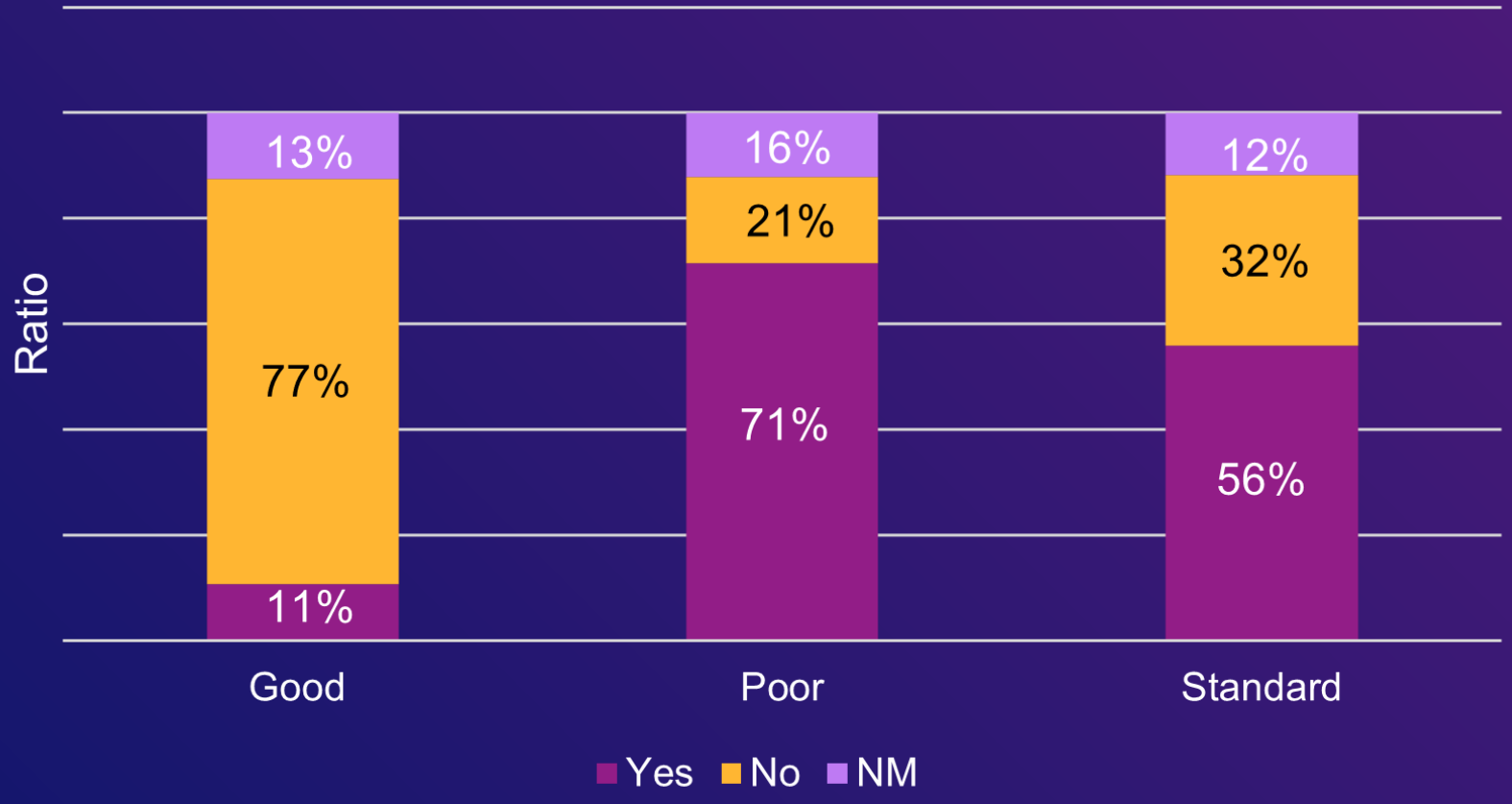
Hypothesis Number 1 is accepted depending on this chart

RELATIONSHIPS:

Matching Ratio between Credit Mix and Credit Score

RELATIONSHIPS:

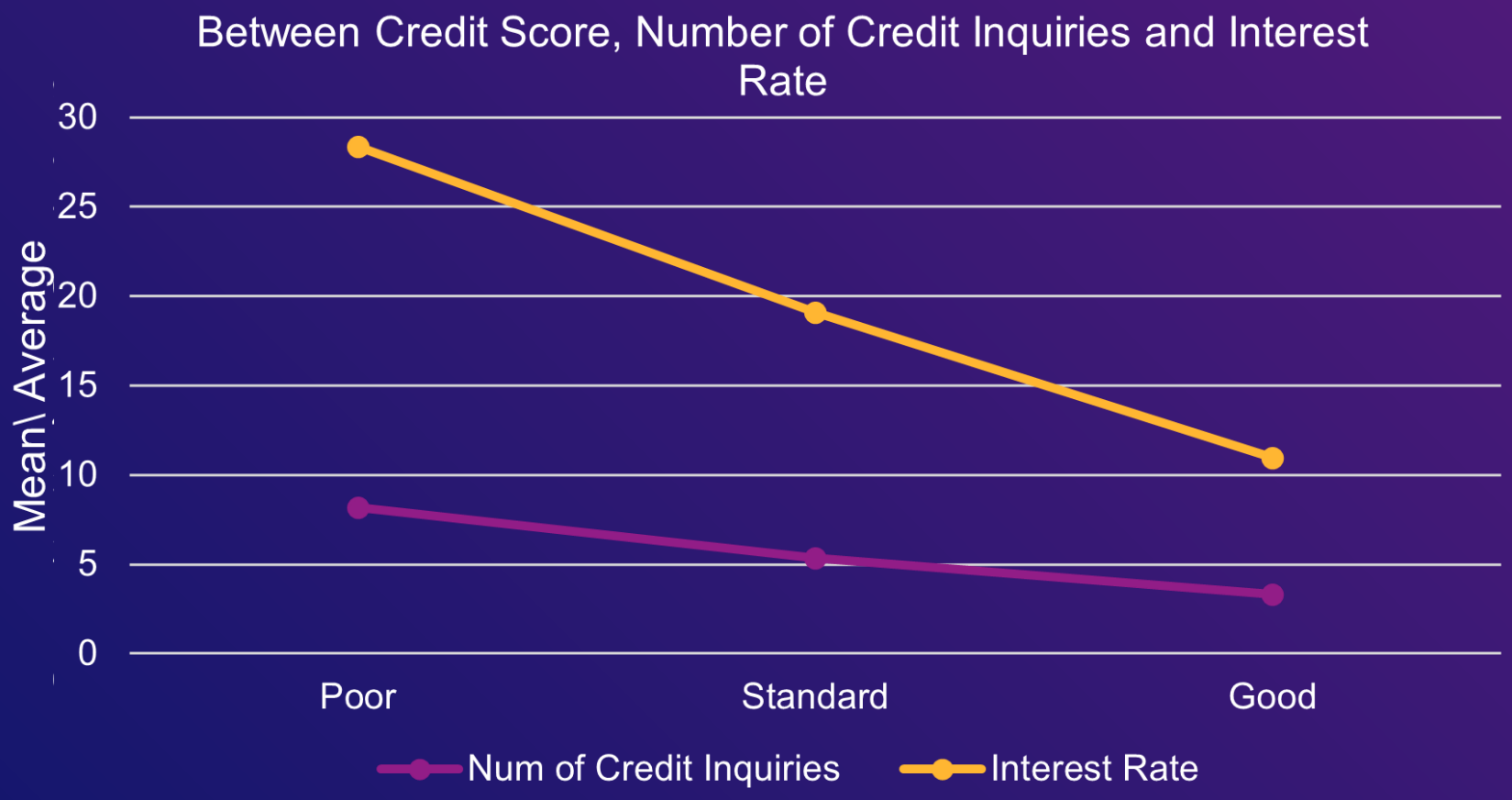
Between Credit Score, Payment of Minimum Amount



People with Poor Credit Score paid only the minimum amount of their debt.

People with Good Credit Score paid more than the minimum amount of their debt.

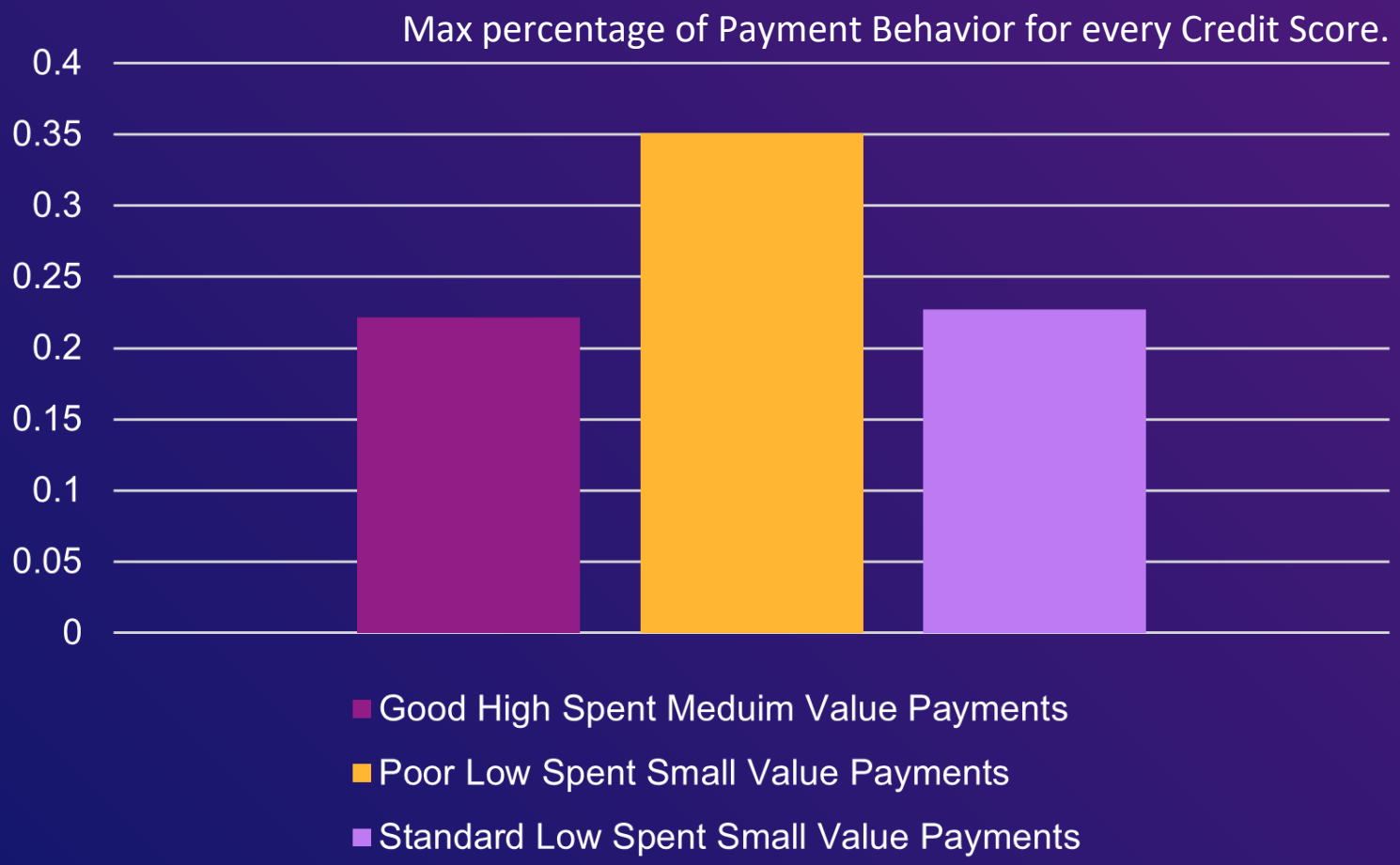
RELATIONSHIPS:



↑ Number of Credit Inquiries & Interest Rate

↓ Credit Score

RELATIONSHIPS:



Hypothesis Number 3 is Rejected depending on this chart, as it shows that Good Credit Score has High Spent behavior depending on the ratio.

06

05 FEATURE EXTRACTION:

DELAY BY CUSTOMER MEAN

The average number of days that the customer had delayed

TYPE OF LOAN

We divided each type in a different column

MONTHLY BALANCE BY CUSTOMER MEAN

Monthly balance average for each customer

```
data_loan['delay_by_customer_mean']= data_loan.groupby(['Name'])['Delay_from_due_date'].transform('mean')
data_loan['monthlyBalance_by_customer_mean']= data_loan.groupby(['Name'])['Monthly_Balance'].transform('mean')
```

FEATURE SELECTION:

After feature extraction and one hot encoding for categorical features, we ended up with 53 columns.

First, we trained the model (RFC) on all the columns, resulting in the test data score of (0.835).

Second, we applied feature selection and used (MIC) algorithm and (MRMR) algorithm ending up selecting the 43 best features. Then, we trained the model resulting in a test data score of (0.836). ☺

03

DATA OBSERVATION

Observation that we have about the data

MALAK DIAB, MOHAMMED HOURANI, MAYAS MASALMEH

DATA OBSERVATION:

(0.0) WITH CREDIT CARD

There’s 4417 record for people with no bank accounts, but they have a credit cards.

TYPE OF LOAN

(No Data) = Number of Loan (0)
Not Specified

AGE LESS THAN 18

There’s 7098 record with ages <18.

NM

In Payment with minimum amount, there’s 12007 record with NM.
Some of those records don’t have a loan, but others status is unknown.

04

ADVICES FOR CUSROMERS

Advices for the customers to increase their Credit Score

Download the Credit Score Advices HTML file attached in the file to view the map

MALAK DIAB, MOHAMMED HOURANI, MAYAS MASALMEH

	<52,000	52,000 - 156,000	>156,000
Number of Bank Accounts	In range of 4 and doesn't exceeds 7 accounts	In range of 3 and doesn't exceeds 6 accounts	Doesn't exceeds 3 accounts
Number of Credit Cards	In range 4 and doesn't exceeds 7 cards	In range 4 and doesn't exceeds 6 cards	In range 4 and doesn't exceeds 5 cards
Number of Loans	In range 2 and doesn't exceeds 5 loans	In range 2 and doesn't exceeds 4 loans	Doesn't exceeds 2 loans
Delay in Days	In range 10 and doesn't exceeds 28 days	In range 10 and doesn't exceeds 25 days	In range 8 and doesn't exceeds 12 days
Delay in Payment	In range 9 and doesn't exceeds 17 times	In range 8 and doesn't exceeds 16 times	Doesn't exceeds 6 times
Amount Invested Monthly	In range $37 and more is better	In range $93 and more is better	In range $158 and more is better
Payment Behavior	Spend as much as paid	Spend as much as paid	Spend as much as paid

THANKS!

DO YOU HAVE ANY QUESTIONS?

Made with Lots of Love by the Most
Creative Clever People Ever.

MALAK DIAB
MOHAMMED HOURANI
MAYAS MASALMEH