

CREDIT SCORE CLASSIFICATION

GROUP 1

22056764

Devayani A/P Balkrishnan

S2195613

Li Xin Qi

S2155659

Lim Yu Xuan

S2192763

Navaneeta A/P P Shanmugam

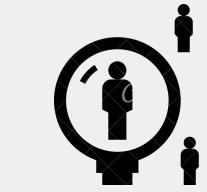
CONTENT

- 
- 01** INTRODUCTION & OBJECTIVES
 - 02** DATASET DESCRIPTION
 - 03** SAMPLE
 - 04** EXPLORE: UNIVARIATE ANALYSIS
 - 05** EXPLORE: BIVARIATE ANALYSIS
 - 06** EXPLORE: MULTIVARIATE ANALYSIS
 - 07** CONCLUSION

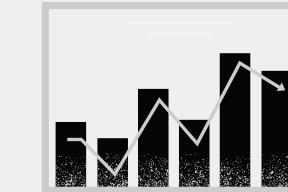
INTRODUCTION

Credit Score Classification Model

- Reflect individual's creditworthiness
- Allow financial institution to:
 - optimize capability in determining individual's risk level
 - minimize credit risk and default cost
 - increase sales and overall expected profits



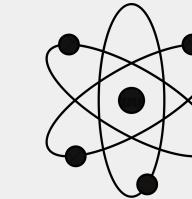
01 Sample



02 Explore



03 Modify



04 Model



05 Assess

OBJECTIVES



Objective 1

Evaluate the borrower's creditworthiness.



Objective 2

Assess borrower's credit-related information and classify them using credit score

DATASET DESCRIPTION

No. of records: 100,000

No. of unique customer: 12,500

No. of attributes: 28

Before

- Ordinal: -
- Interval: 9 attributes
- Nominal: 19 attributes

After

- 3 attributes
- 9 attributes
- 16 attributes

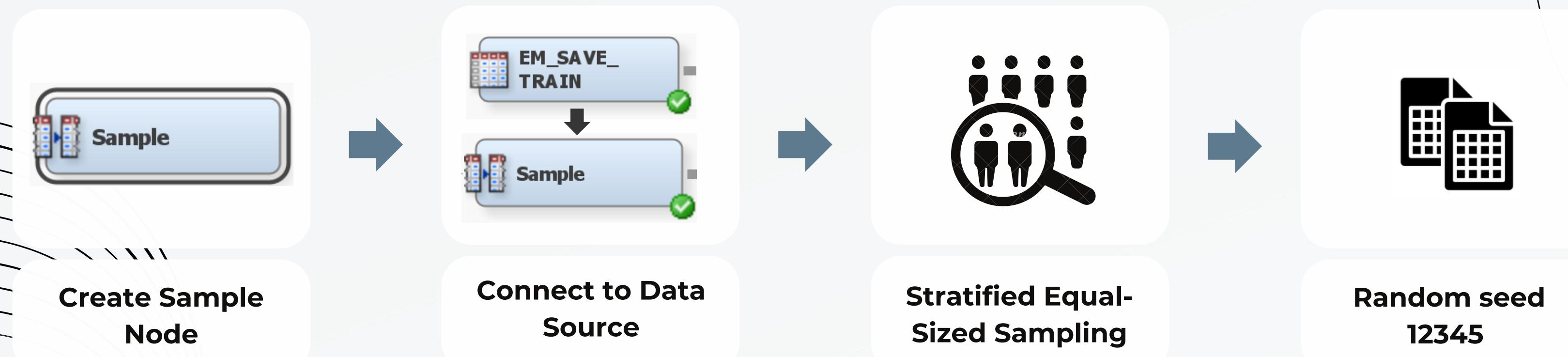
DATASET DESCRIPTION

Reclassification of Variables Roles and Level

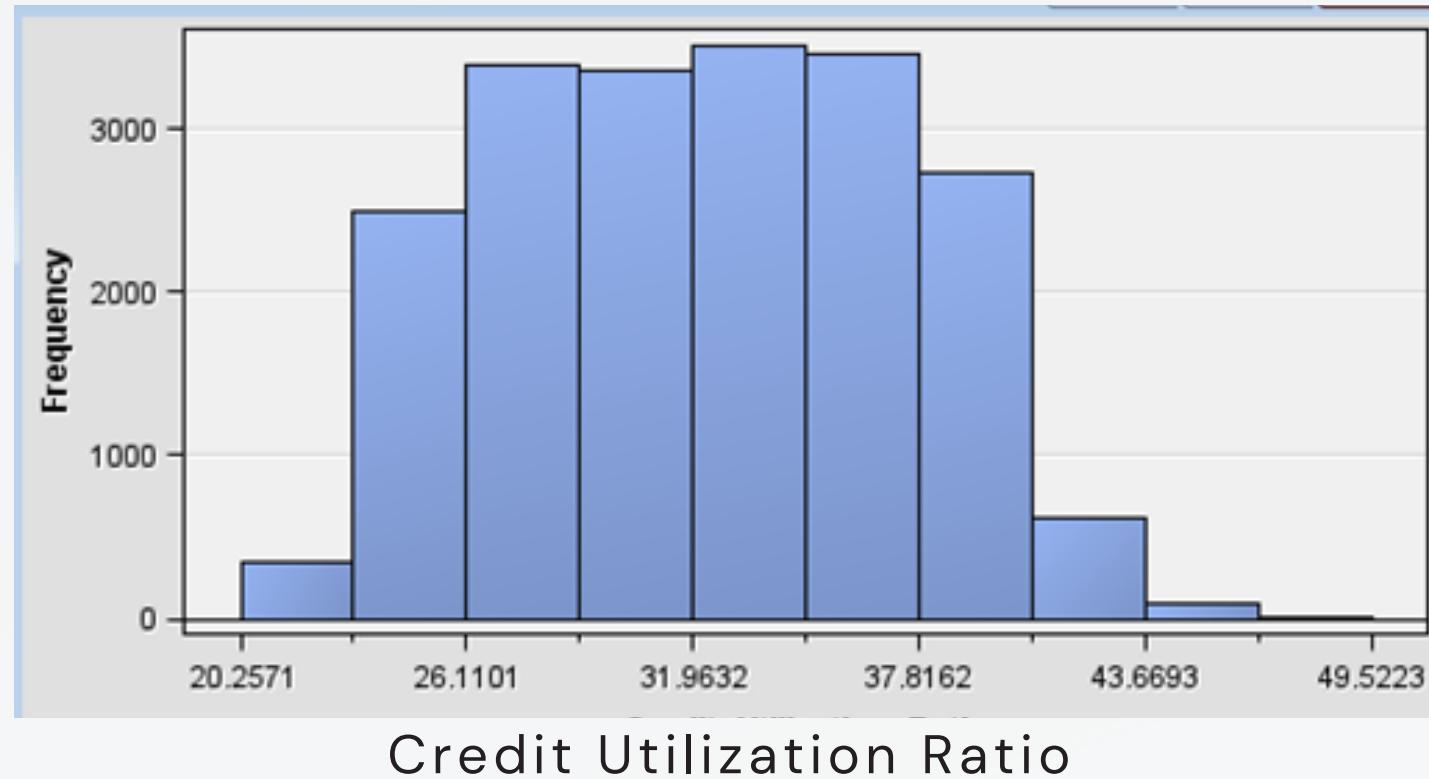
Before			After			Final	
Variables	Role	Level	Variables	Role	Level	Variables	Level
Age	Input	Nominal	Age	Input	Nominal	Age	Nominal
Amount_invested_monthly	Input	Nominal	Amount_invested_monthly	Input	Nominal	Amount_invested_monthly	Interval
Annual_Income	Input	Nominal	Annual_Income	Input	Ordinal	Annual_Income	Interval
Changed_Credit_Limit	Input	Nominal	Changed_Credit_Limit	Input	Nominal	Changed_Credit_Limit	Interval
Credit_History_Age	Input	Nominal	Credit_History_Age	Input	Nominal	Credit_History_Age	Ordinal
Credit_Mix	Input	Nominal	Credit_Mix	Input	Ordinal	Credit_Mix	Ordinal
Credit_Score	Input	Nominal	Credit_Score	Target	Ordinal	Credit_Score	Ordinal
Credit_Utilization_Ratio	Input	Interval	Credit_Utilization_Ratio	Input	Interval	Credit_Utilization_Ratio	Interval
Customer_ID	ID	Nominal	Customer_ID	Input	Nominal	Customer_ID	Nominal
Delay_from_due_date	Input	Interval	Delay_from_due_date	Input	Interval	Delay_from_due_date	Interval
ID	ID	Nominal	ID	ID	Nominal	ID	Nominal
Interest_Rate	Input	Interval	Interest_Rate	Input	Interval	Interest_Rate	Interval
Month	Input	Nominal	Month	Input	Nominal	Month	Nominal
Monthly_Balance	Input	Interval	Monthly_Balance	Input	Interval	Monthly_Balance	Interval
Monthly_Inhand_Salary	Input	Interval	Monthly_Inhand_Salary	Input	Interval	Monthly_Inhand_Salary	Interval
Name	Rejected	Nominal	Name	Rejected	Nominal	Name	Nominal
Num_Bank_Accounts	Input	Interval	Num_Bank_Accounts	Input	Interval	Num_Bank_Accounts	Interval
Num_Credit_Card	Input	Interval	Num_Credit_Card	Input	Interval	Num_Credit_Card	Interval
Num_Credit_Inquiries	Input	Interval	Num_Credit_Inquiries	Input	Interval	Num_Credit_Inquiries	Interval
Num_of_Delayed_Payment	Rejected	Nominal	Num_of_Delayed_Payment	Input	Nominal	Num_of_Delayed_Payment	Interval
Num_of_Loan	Rejected	Nominal	Num_of_Loan	Input	Nominal	Num_of_Loan	Interval
Occupation	Input	Nominal	Occupation	Input	Nominal	Occupation	Nominal
Outstanding_Debt	Rejected	Nominal	Outstanding_Debt	Input	Nominal	Outstanding_Debt	Interval
Payment_Behaviour	Input	Nominal	Payment_Behaviour	Input	Nominal	Payment_Behaviour	Nominal
Payment_of_Min_Amount	Input	Nominal	Payment_of_Min_Amount	Input	Nominal	Payment_of_Min_Amount	Nominal
SSN	Rejected	Nominal	SSN	Rejected	Nominal	SSN	Nominal
Total_EMI_per_month	Input	Interval	Total_EMI_per_month	Input	Interval	Total_EMI_per_month	Interval
Type_of_Loan	Text	Nominal	Type_of_Loan	Input	Nominal	Type_of_Loan	Nominal

SAMPLE

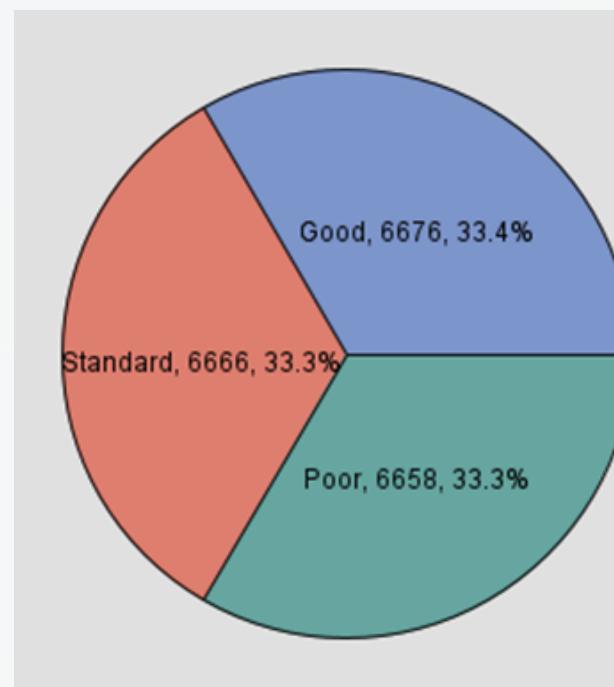
Subset of dataset selected to represent the entire dataset, which is used for variable identification purpose.



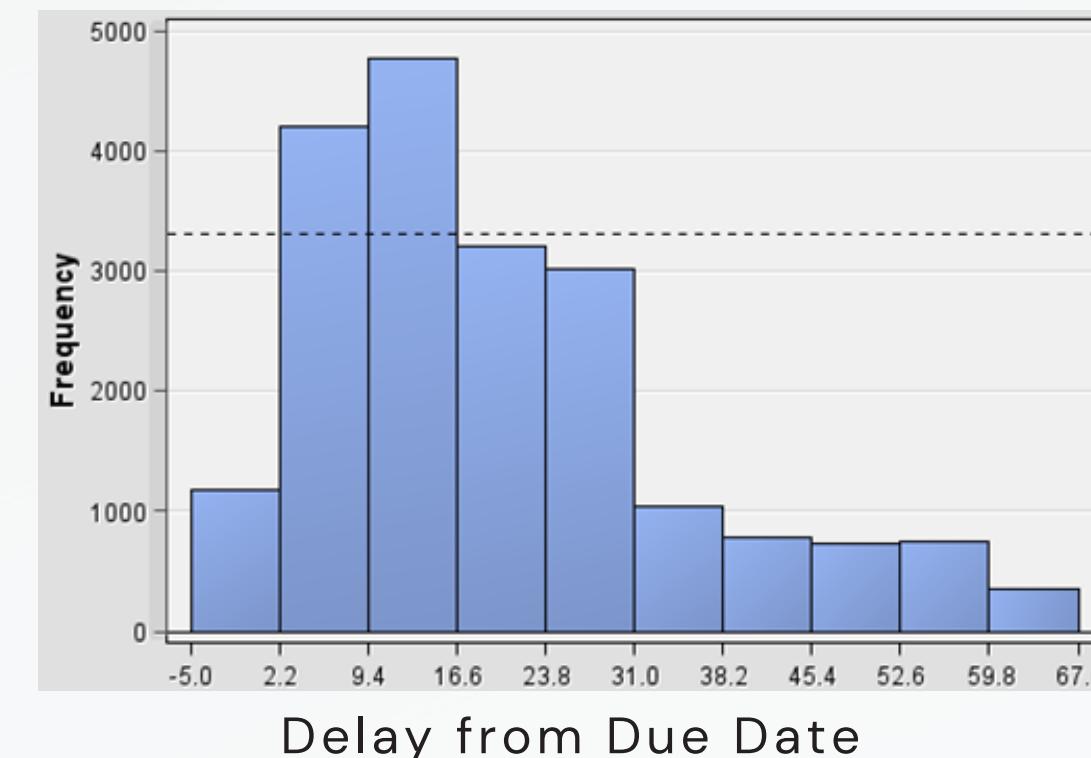
EXPLORE: UNIVARIATE



Credit Utilization Ratio



Credit Score



Delay from Due Date

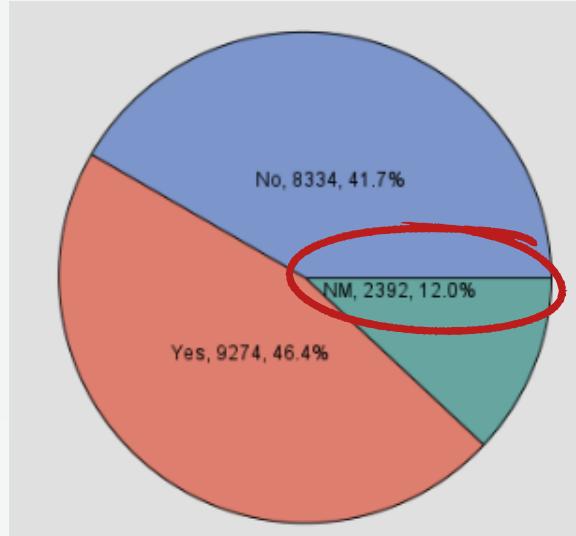


Clean Data

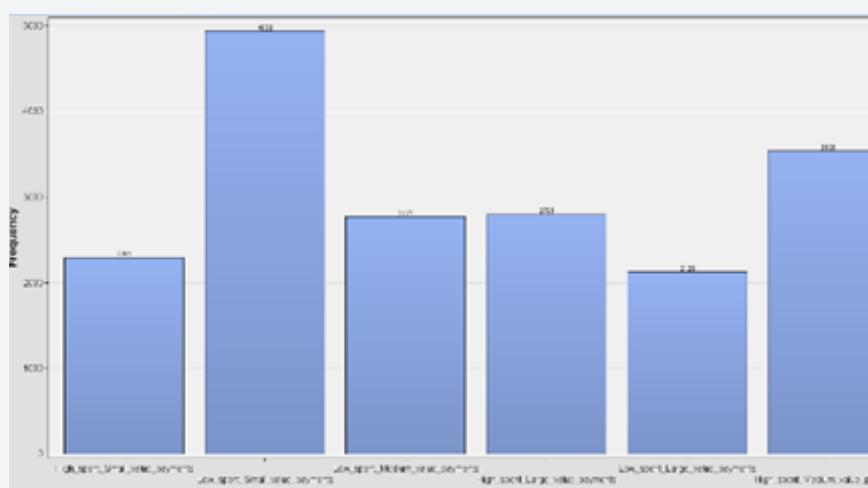
Input and target variables with clean data:

- Credit Utilization Ratio
- Delay from Due Date
- Credit Score

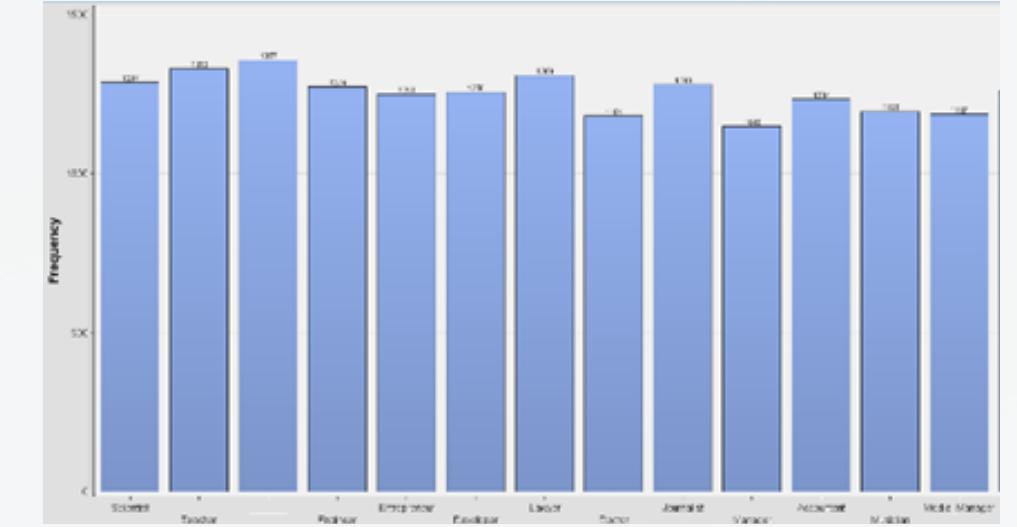
EXPLORE: UNIVARIATE



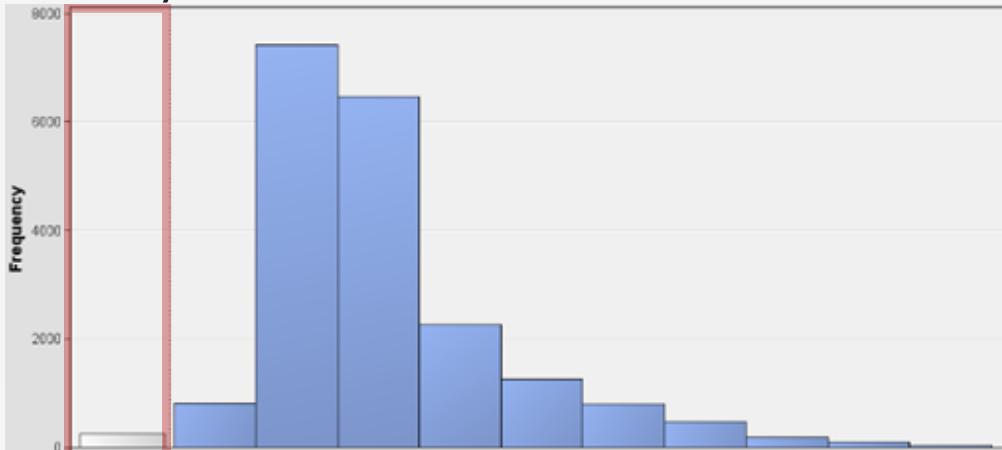
Payment of Minimum Amount



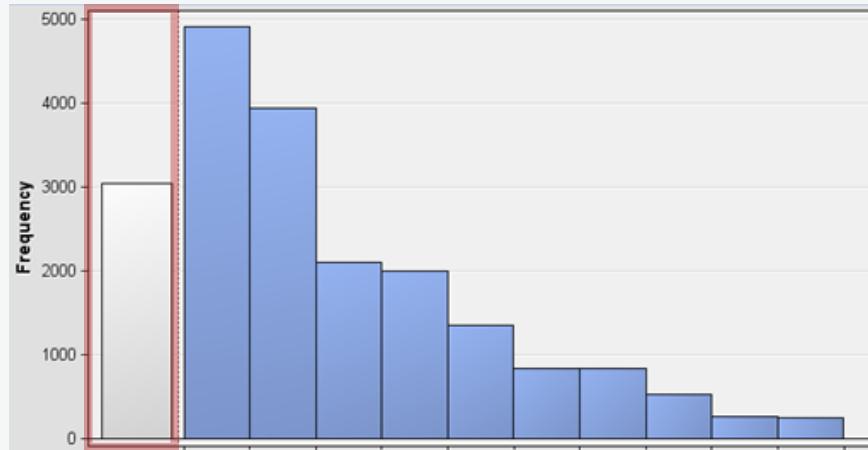
Payment Behavior



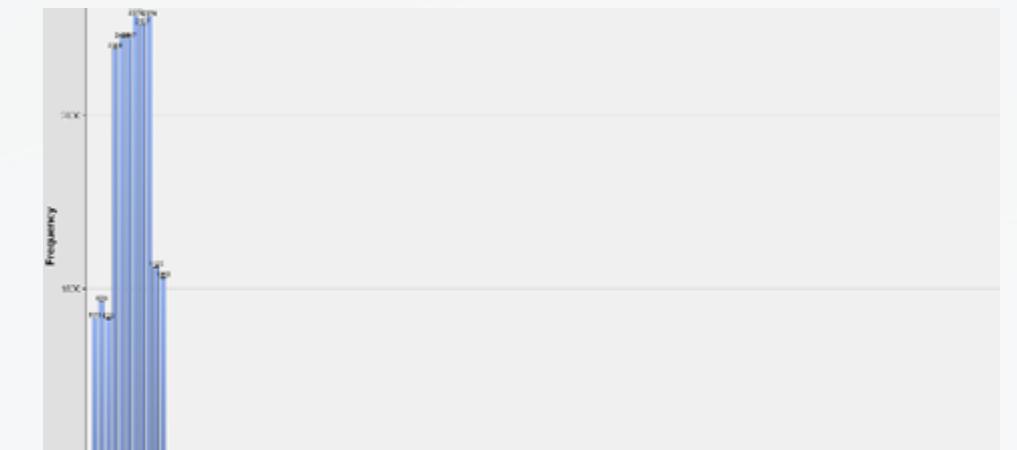
Occupation



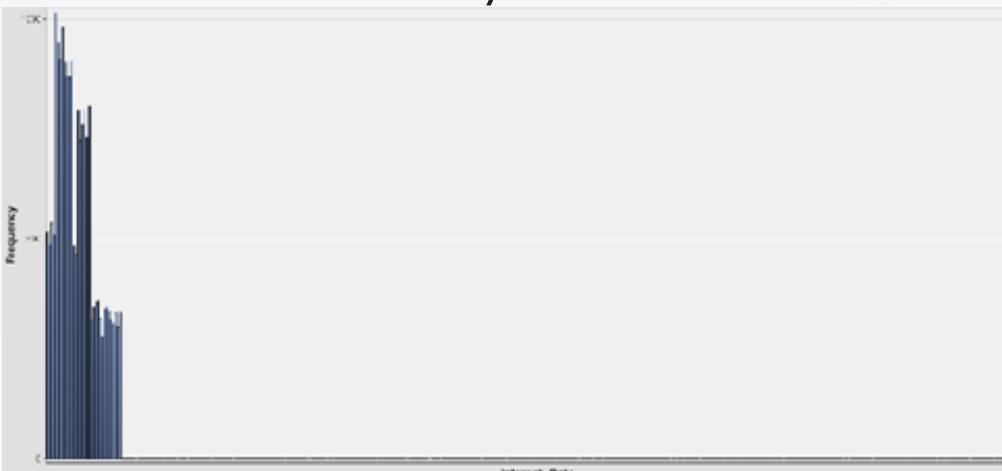
Monthly Balance



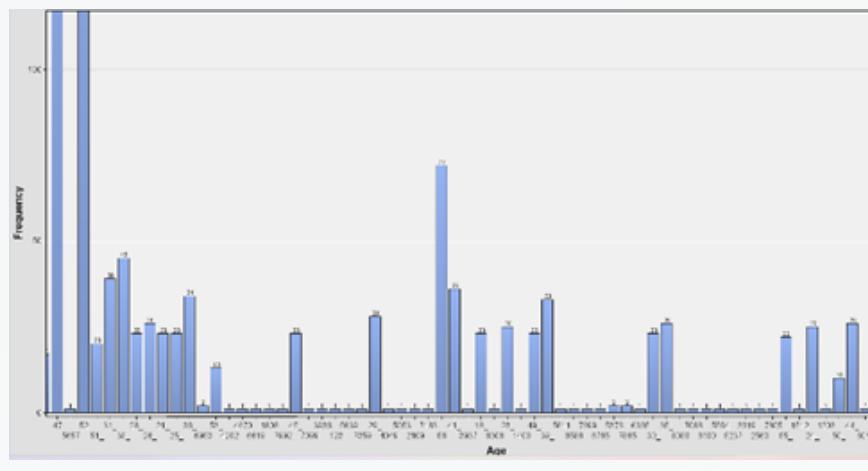
Monthly In-Hand Salary



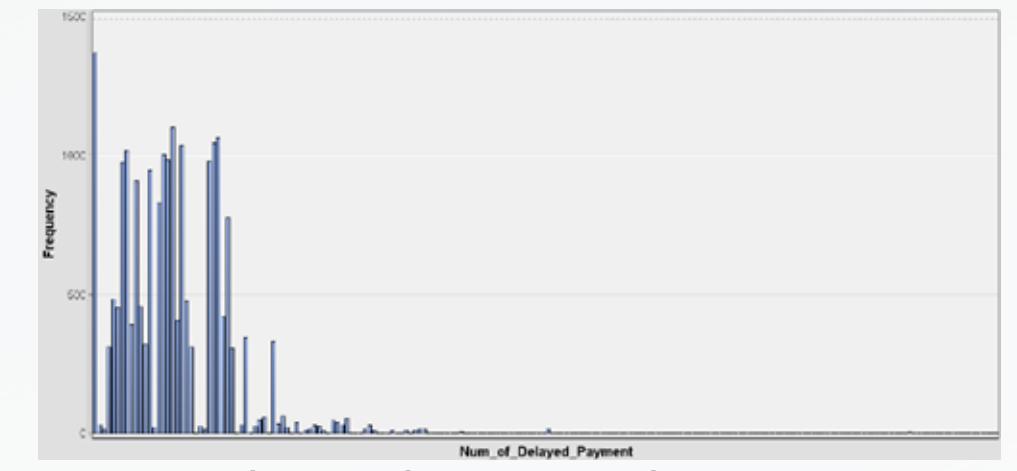
Number of Bank Accounts



Interest Rate



Age



Number of Delayed Payment



Dirty & Noisy Data

Input variables that consist of dirty and noisy data due to several factors:

- Missing Values
- Invalid Values
- Spelling Errors
- Outliers

DATA ERROR TYPE

INCOMPLETE

- Name
- Occupation
- Monthly_Inhand_Salary
- Type_of_Loan
- Num_of_Delayed_Payment
- Changed_Credit_Limit
- Num_Credit_Inquiries
- Credit_Mix
- Credit_History_Age
- Payment_of_Min_Amount
- Amount_invested_monthly
- Monthly_Balance

NOISY

- Age
- Occupation
- Annual_Income
- Num_Bank_Accounts
- Num_Credit_Card
- Num_of_Loan
- Num_of_Delayed_Payment
- Changed_Credit_Limit
- Outstanding_Debt
- Amount_invested_monthly
- Payment_Behaviour
- Interest_Rate
- Total_EMI_per_Month

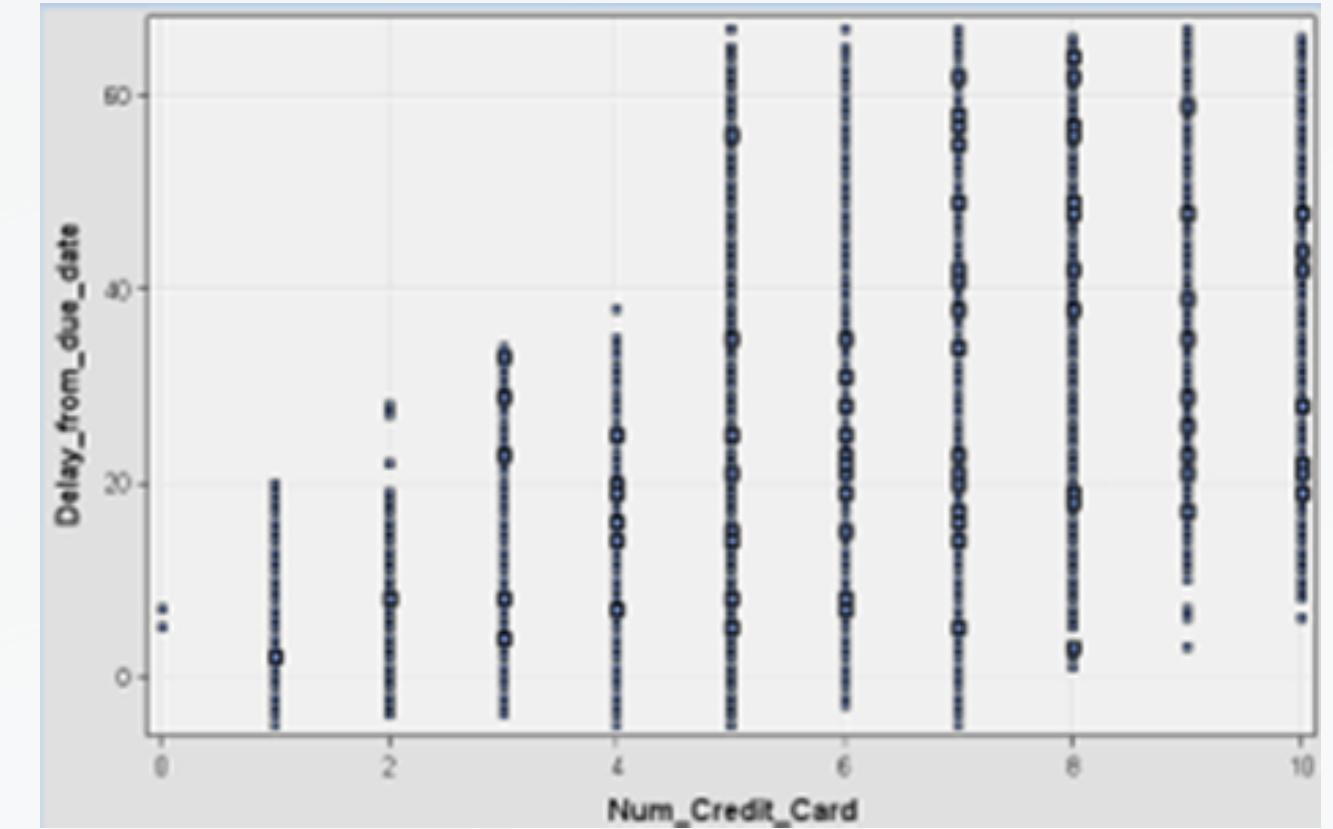
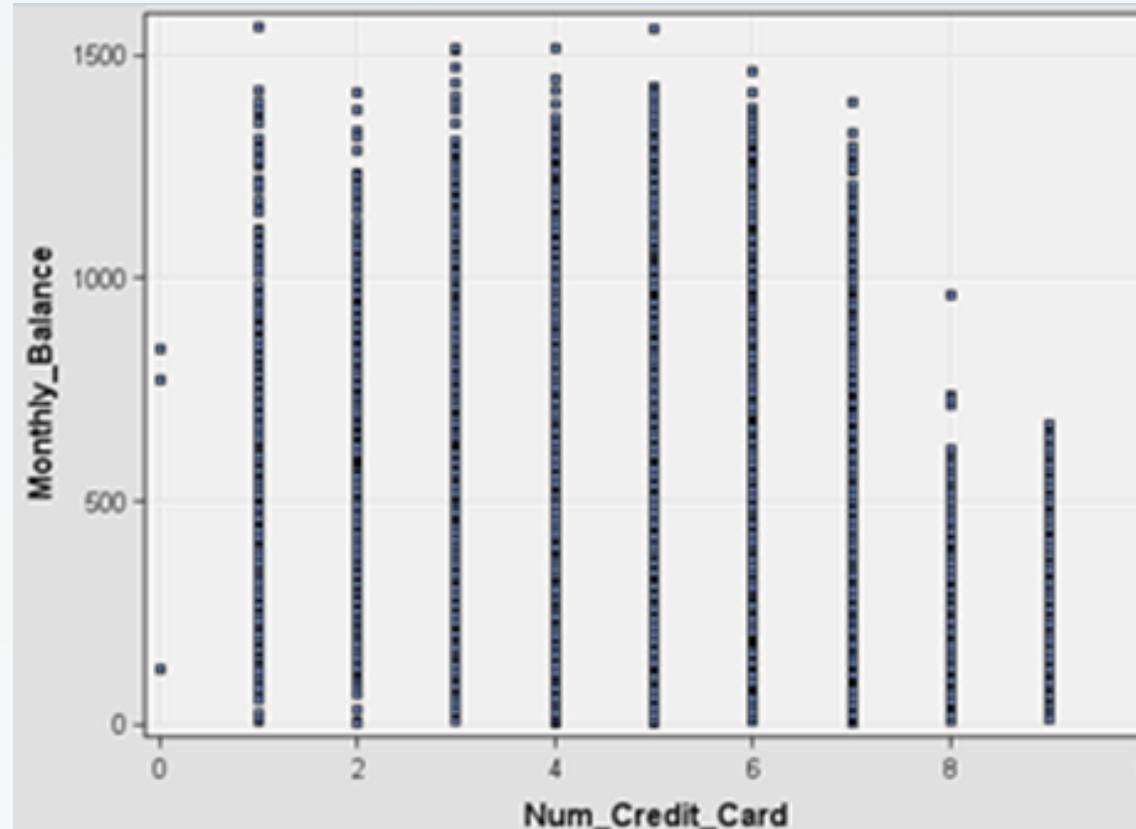
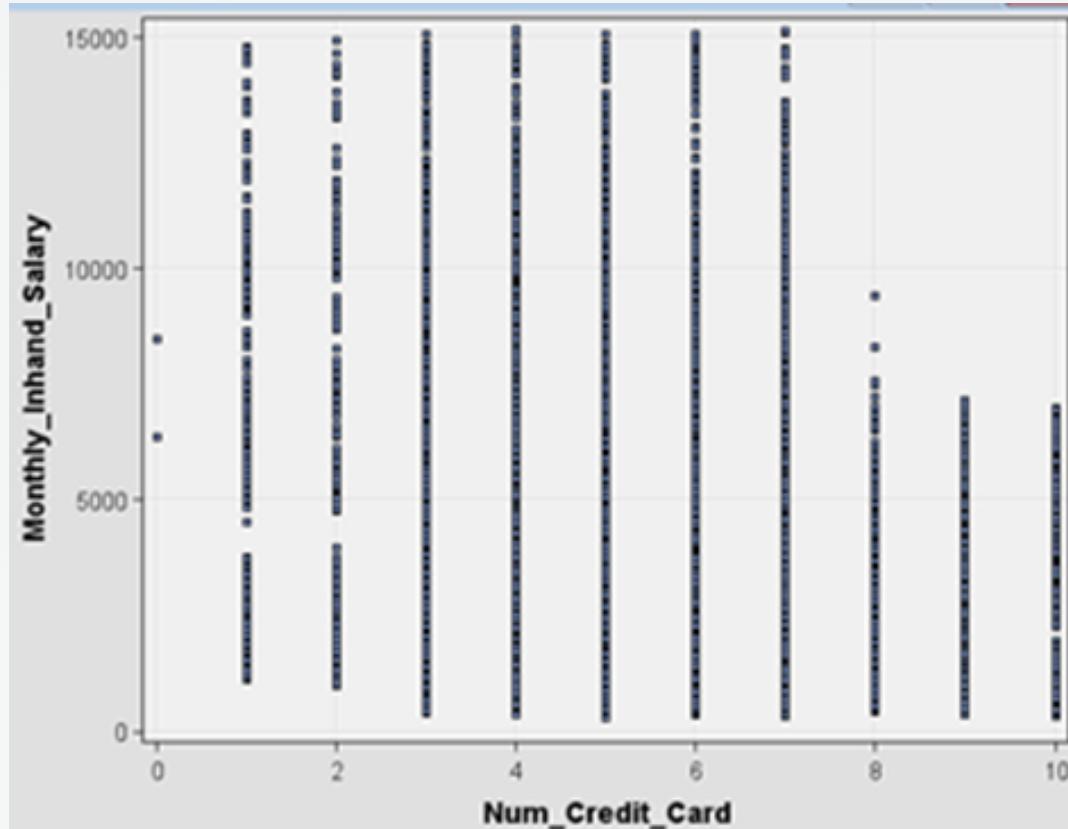
INCONSISTENT

- Num_of_Loan

EXPLORE: BIVARIATE

ANALYSIS ON INPUT VARIABLES

Some of the graphs needed to be analyzed close-up to be able to account for the outliers and find hidden patterns.



A huge drop in the monthly in-hand salary for customers owned more than 7 credit cards.

Similar trend as previous graph

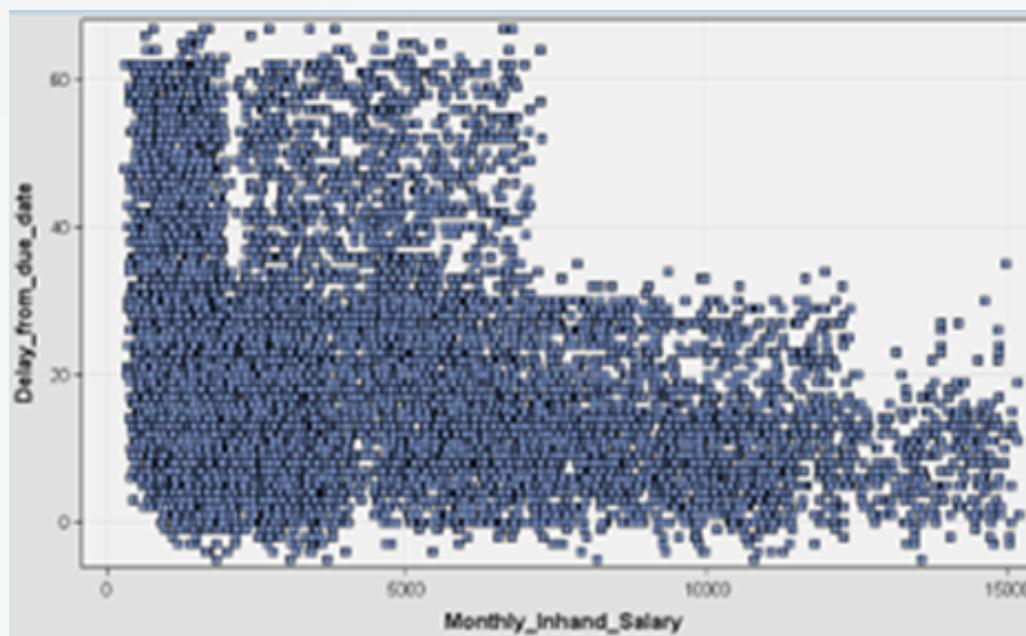
Number of days delayed from the payment due date increases sharply if the customer holds more than 4 credit cards

EXPLORE: BIVARIATE

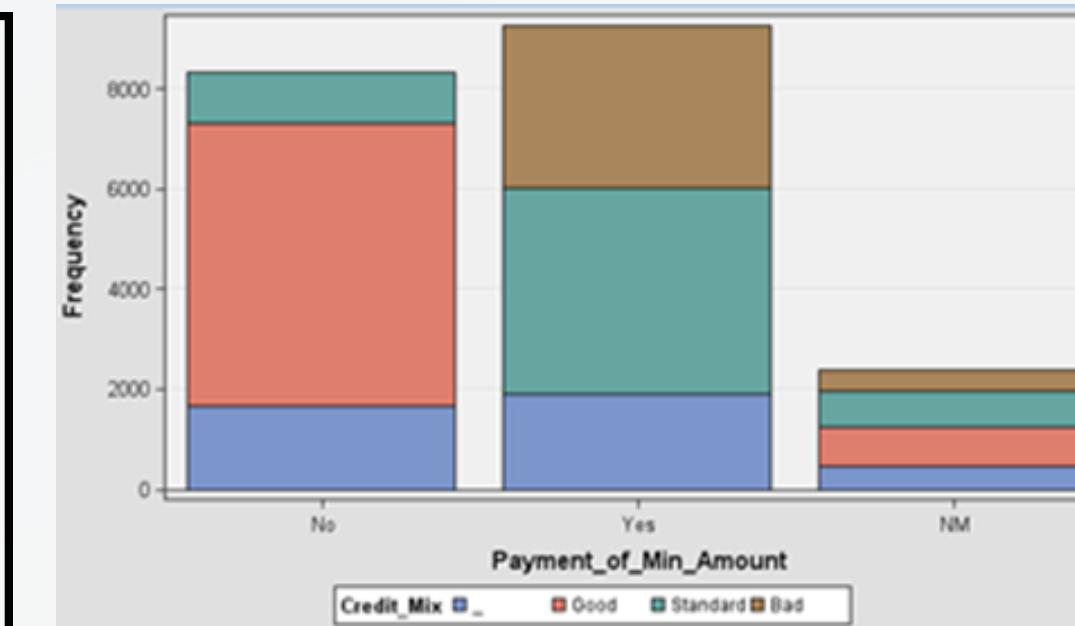
ANALYSIS ON INPUT VARIABLES



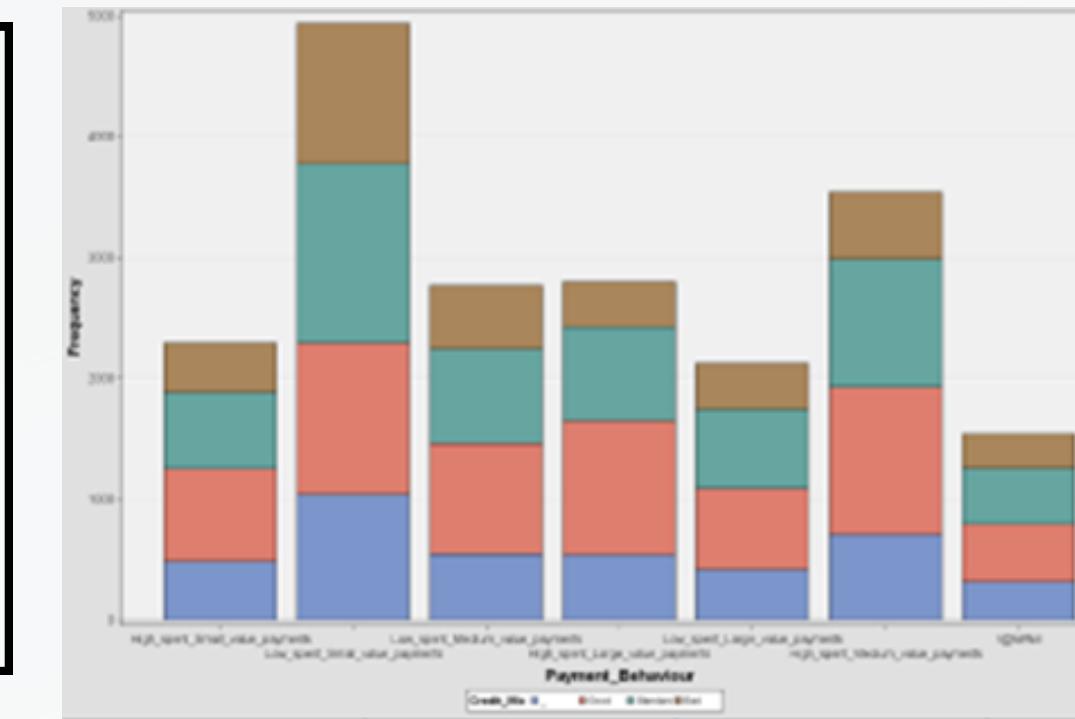
Higher the monthly in-hand salary, the higher the monthly balance of the customer



Customers with higher monthly in-hand salary pay their debt earlier, as compared to those with lower monthly in-hand salary, with a sharp drop. Similar trend was seen a total of 4 times.

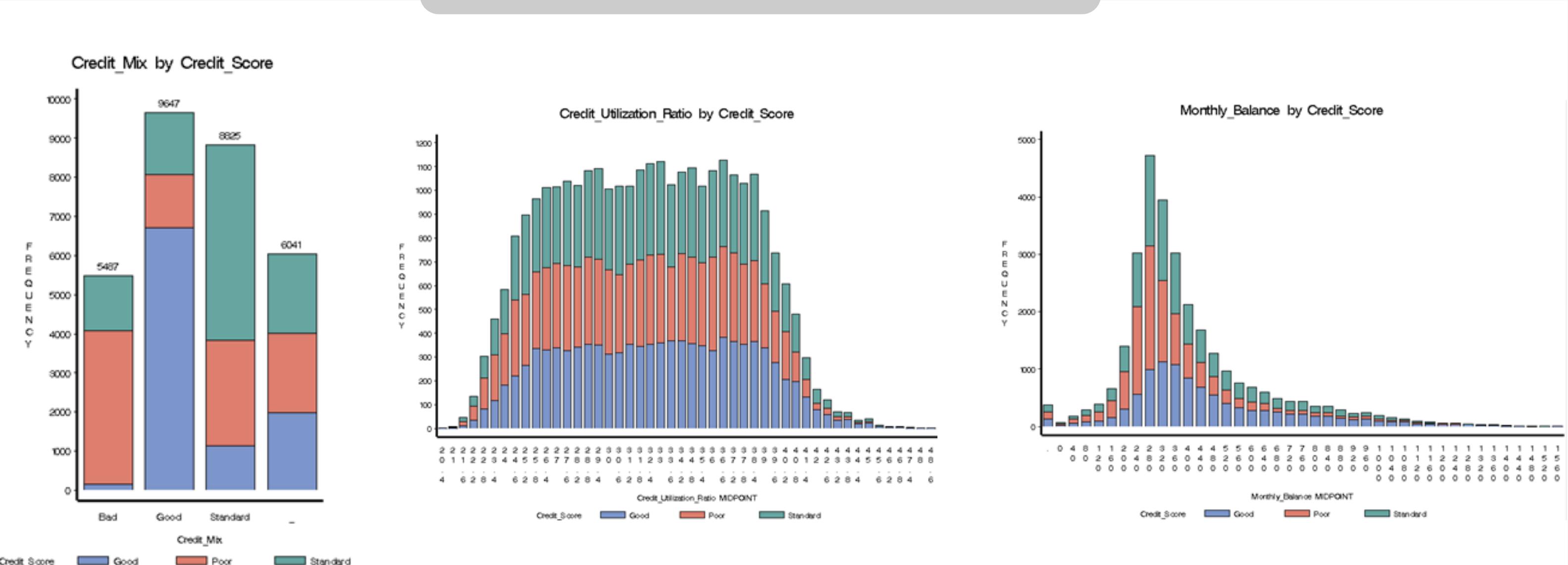


Customers who paid in full amount and have better payment behavior have better credit mix.



EXPLORE: BIVARIATE

ANALYSIS ON INPUT AND TARGET VARIABLES



Better credit mix gives better credit score

Normal distribution

Right-skewed normal distribution

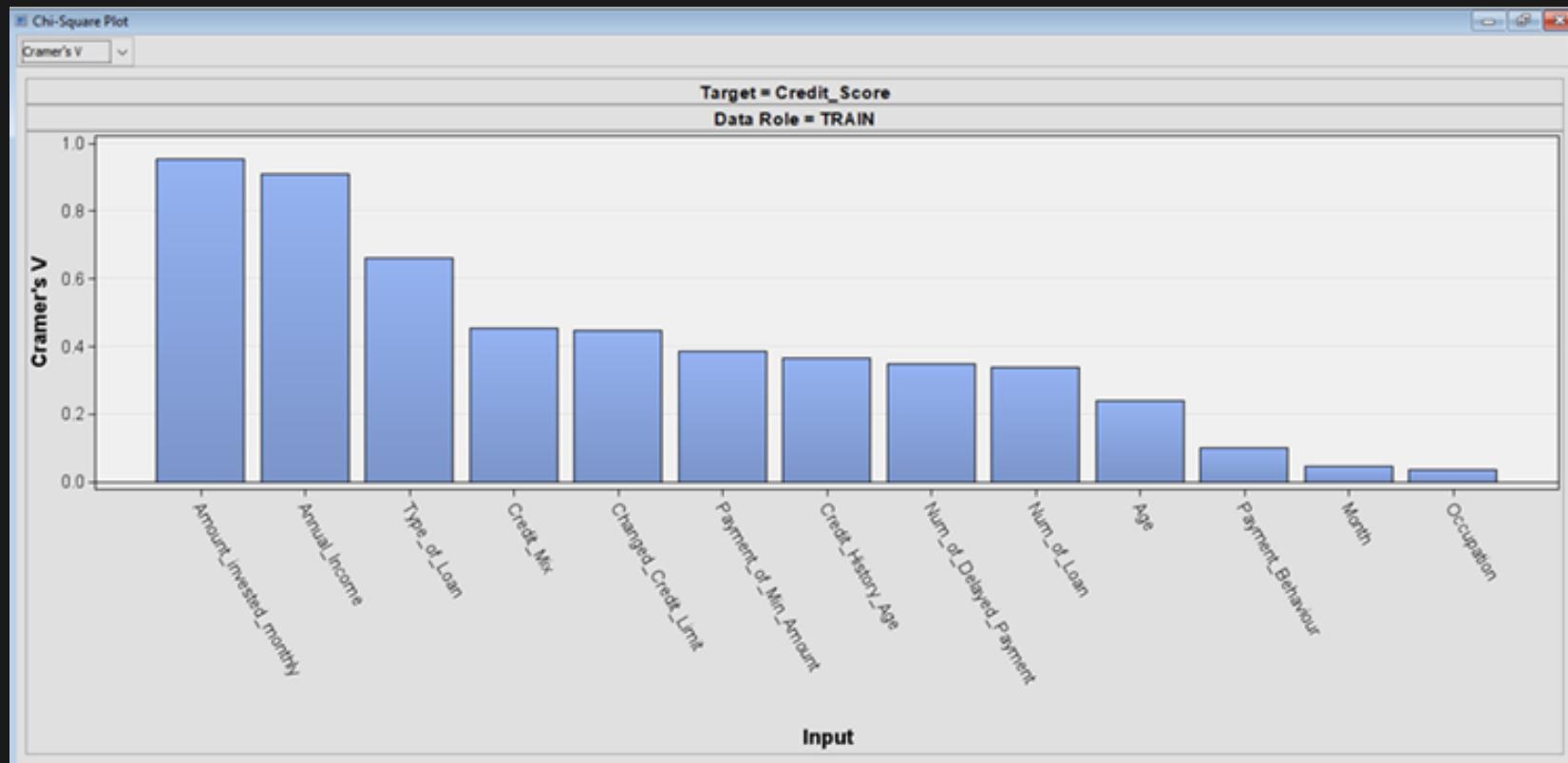
EXPLORE: MULTIVARIATE



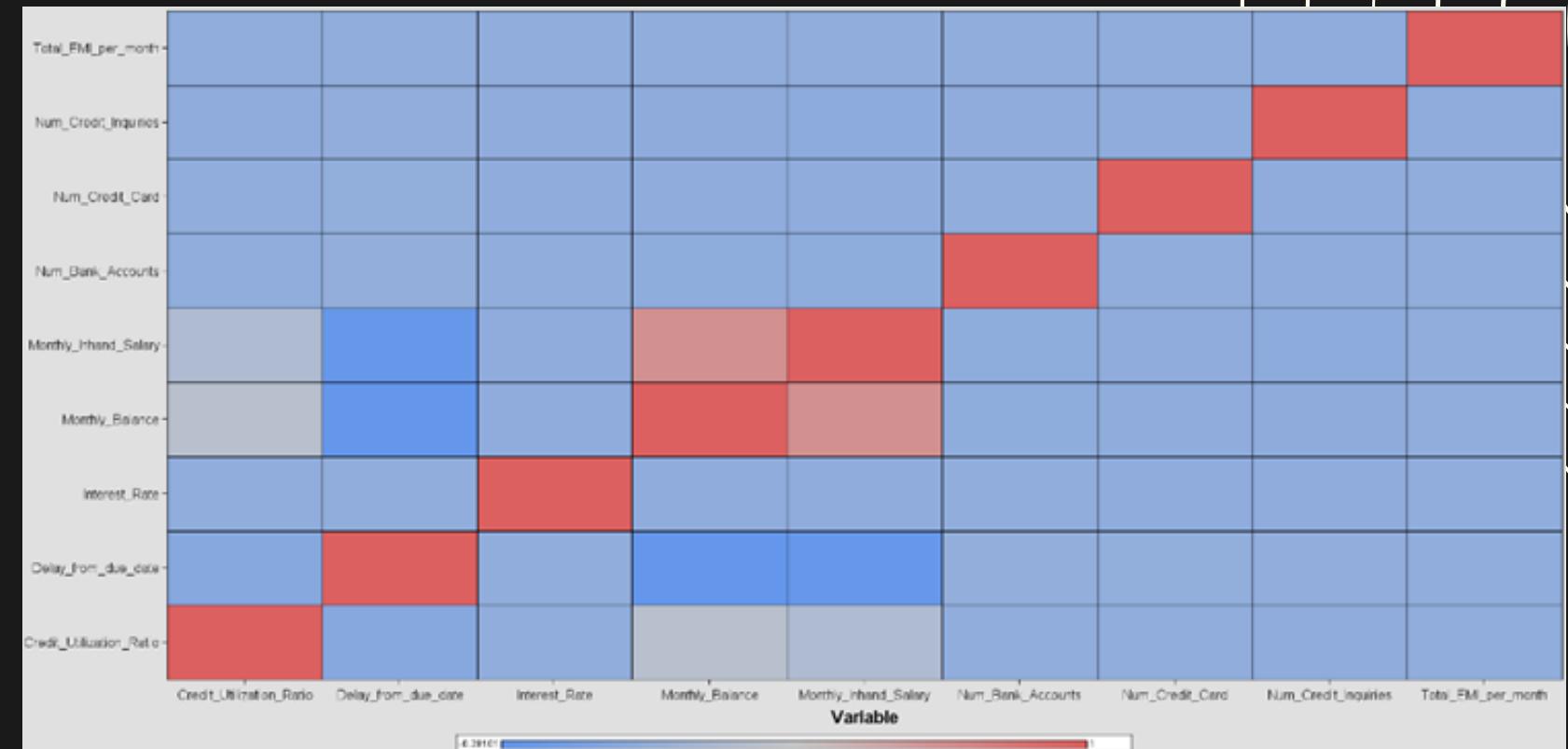
→ Regardless of payment behaviour, a better credit mix will give a better credit score.
High spent large value payment is more likely to give a good credit score.

EXPLORE: MULTIVARIATE

CORRELATION MATRIX



Of the nominal variables, month and occupation show nearly no association with Credit Score.



Of the interval variables, the closest positive correlation value obtained was 0.7 from monthly in-hand salary and monthly balance. No attributes were dropped.

CONCLUSION



Common data type errors found in this dataset are incomplete due to missing values, inconsistent due to negative values that are supposed to be positive value, and lastly noisy data due to outliers.



Invalid data types will be corrected in the Modify stage



Overall, credit mix, monthly balance, delay from payment due date (in days) and payment of only minimum amount show correlation with the credit score. There may be more variables showing correlation with the credit score after applying data transformation in the Modify phase.

**THANKS FOR
LISTENING**

