

# Prediction Model on a Revolving Credit Line

Group Name: **GROUP-6**

Mentor: **MUNMUN**

Date: **09-AUG-2020**

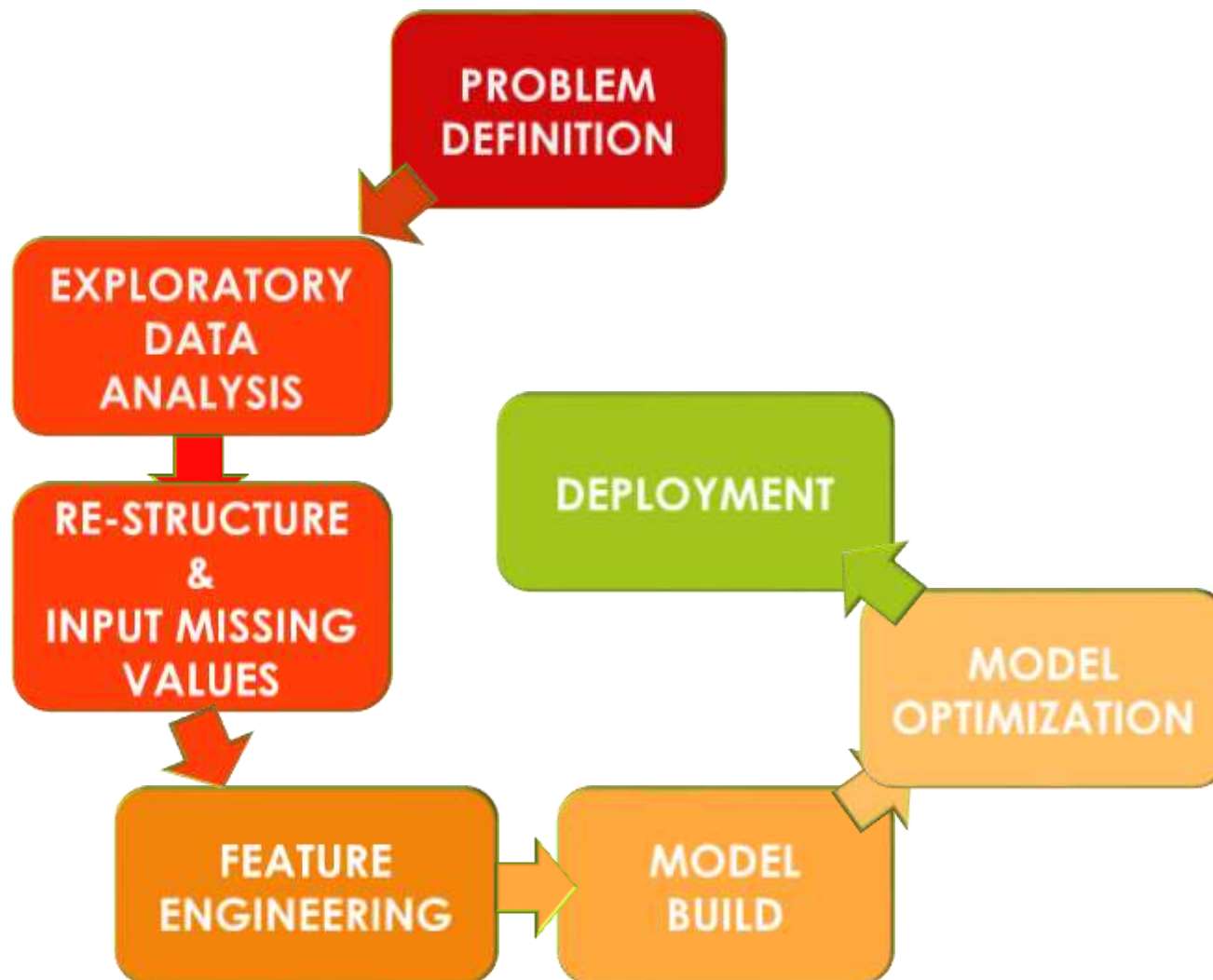
## Business Problem:

- Revolving Credit means you're borrowing against a Line of Credit.
- Users can borrow the amount of credit allowed to use each month known as *“Credit line or Credit Limit”*.
- Similar to a **Credit card** with the only difference being *“Lesser Interest Rates and Secured Business Assets”*.
- At the end of each statement period, a Bill gets generated.  
*If not paid fully, the balance is carried over, or revolved over to the next month along with Interest incurred on the remaining balance.*
- *As you pay down the balance, more of your credit line becomes available.*

## Objective:

*“To predict the revolving balance maintained by each customer to derive market strategies for a bank or investment firm”.*

# Project Architecture / Project Flow



# Exploratory Data Analysis (EDA)

*“Exploratory data analysis can never be the whole story, but nothing else can serve as the foundation stone” — John Tukey*

## Steps for EDA:

- Describe Dataset: Rows/Columns, Missing Data, Data Types, Preview
- Clean corrupted data; Handle missing data, Invalid data types, Incorrect values.
- Visualize data distributions; Bar charts, Histograms, Box plots.
- Calculate and Visualize Correlations (Relationships) between variables; Heat map.

Dataset info

Number of variables	34
Number of observations	887379
Total Missing (%)	9.9%
Total size in memory	230.2 MiB
Average record size in memory	272.0 B

Variables types

Numeric	21
Categorical	13
Boolean	0
Date	0
Text (Unique)	0
Rejected	0
Unsupported	0

	member_id	loan_amnt	terms	batch_ID	Rate_of_intrst	grade	sub_grade	Emp_designation	Experience	home_ownership	...	recoveries	collection
0	58189336	14350	36 months		19.19	E	E3	clerk	9 years	OWN	...	0.0	
1	70011223	4800	36 months	BAT1586599	10.99	B	B4	Human Resources Specialist	< 1 year	MORTGAGE	...	0.0	
2	70255675	10000	36 months	BAT1586599	7.26	A	A4	Driver	2 years	OWN	...	0.0	
3	1893936	15000	36 months	BAT4808022	19.72	D	D5	Us office of Personnel Management	10+ years	RENT	...	0.0	
4	7652106	16000	36 months	BAT2833642	10.64	B	B2	LAUSD-HOLLYWOOD HIGH SCHOOL	10+ years	RENT	...	0.0	
5	10247268	15000	36 months	BAT2575549	8.90	A	A5	Design Consultant	2 years	MORTGAGE	...	0.0	

# Exploratory Data Analysis (EDA)

## Numerical Type

mths_since_last_record	0.845553
mths_since_last_major_derog	0.750160
mths_since_last_delinq	0.511971
tot_curr_bal	0.079195
tot_colle_amt	0.079195
collections_12_mths_ex_med	0.000163
inq_last_6mths	0.000033
acc_now_delinq	0.000033
delinq_2yrs	0.000033
total_credits	0.000033
pub_rec	0.000033
numb_credit	0.000033
annual_inc	0.000005

MISSING  
VALUES???

## Categorical Type

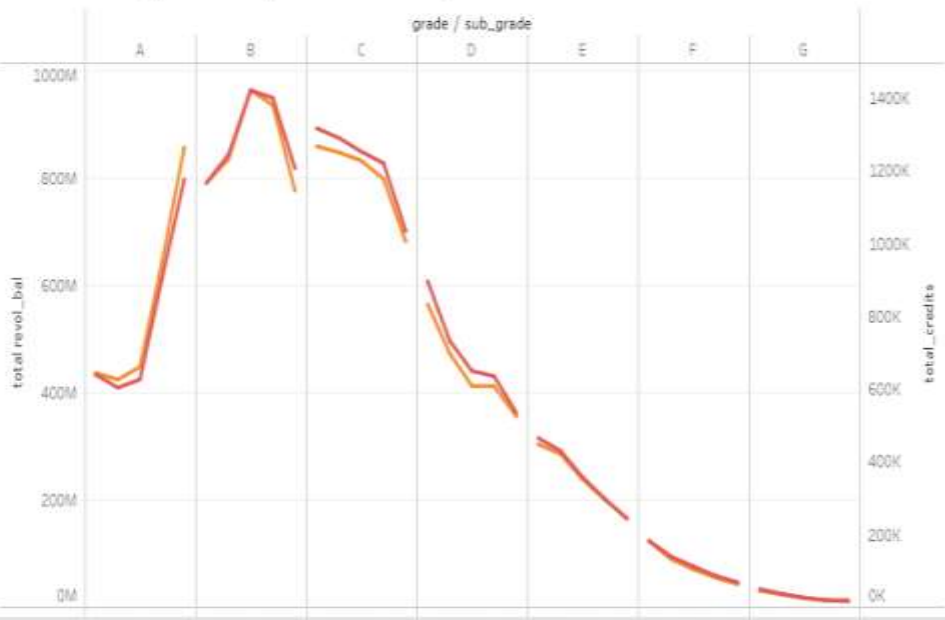
verification_status_joint	0.999424
Emp_designation	0.057993
Experience	0.050514

- **36 months** in 'terms' column has **70%** data compared to **60 months**.
- **33%** of the customers have over **10 years of Experience**.
- **50%** of 'home\_ownership' have **MORTGAGED** while **40%** of them are on **RENT**.
- **Almost 60%** of loans taken cater the purpose of 'debt\_consolidation'.
- **15% (MAX)** of customers are based out of **California State** while only **12 users(MIN)** are from **Idaho state**.
- **80%** of the customers **DON'T HAVE delinquent accounts for a span of 2 years** while **56%** customers **DON'T HAVE delinquent accounts since 6 months**.
- **Almost 80%** of 'tot\_colle\_amt' (total collection amount ever owed) is **ZERO** dollars.
- 'mths\_since\_last\_record', 'mths\_since\_last\_major\_derog' and 'verification\_status\_joint' have **LARGEST** amount of missing values – **84%, 75% and 99.9%** respectively.
- 'annual\_inc', 'debt\_income\_ratio', 'collection\_recovery\_fee', 'acc\_now\_delinq' and 'tot\_colle\_amt' are **HIGHLY SKEWED**.
- 'State' and 'last\_week\_pay' have **HIGH CARDINALITY**.

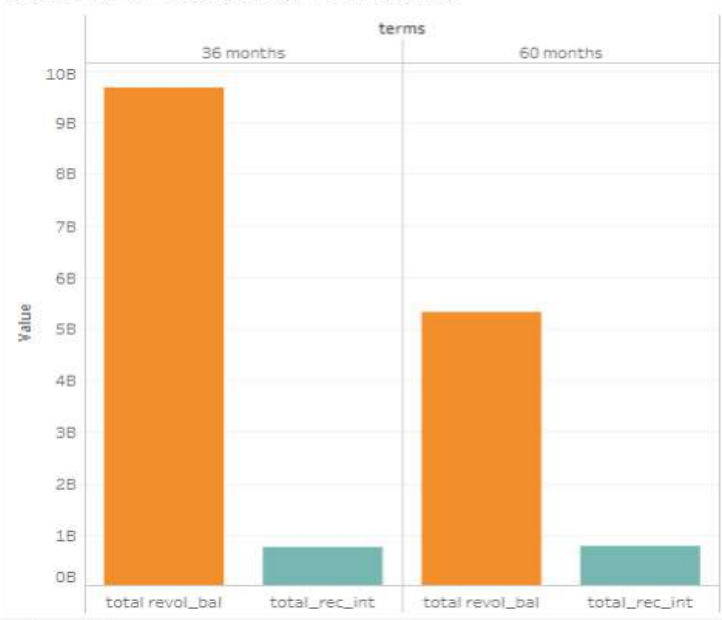
# EDA( VISUALIZATION)

Measure Names  
total\_revol\_bal  
total\_credits

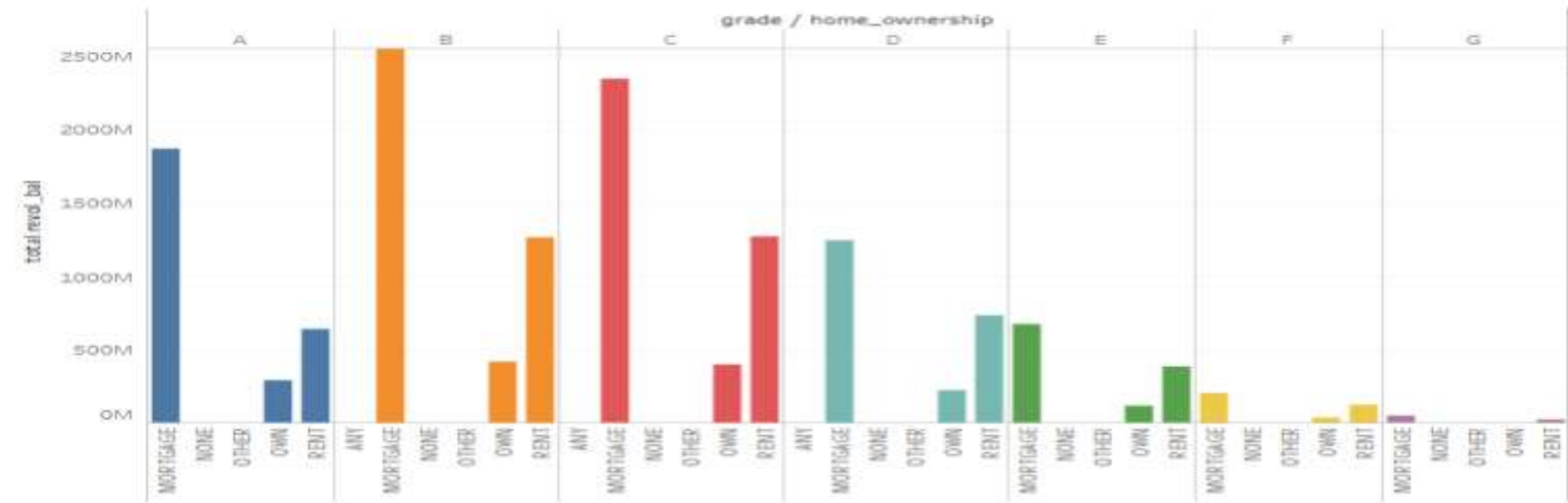
conclusion:-high revolving balance with high amount of credit lines and vice versa



comparision of revolving balance andTotal interest received till date under loan terms

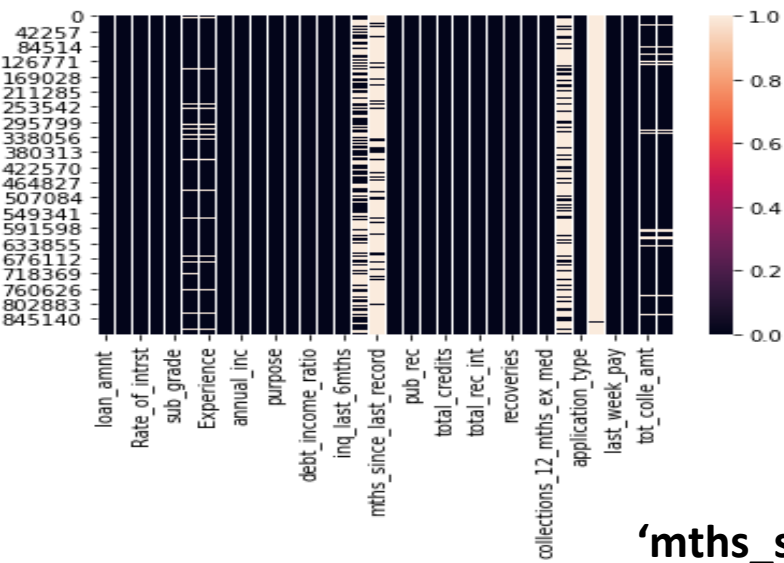


revolving credit balance is maximum for all grades under mortgage type of home ownership

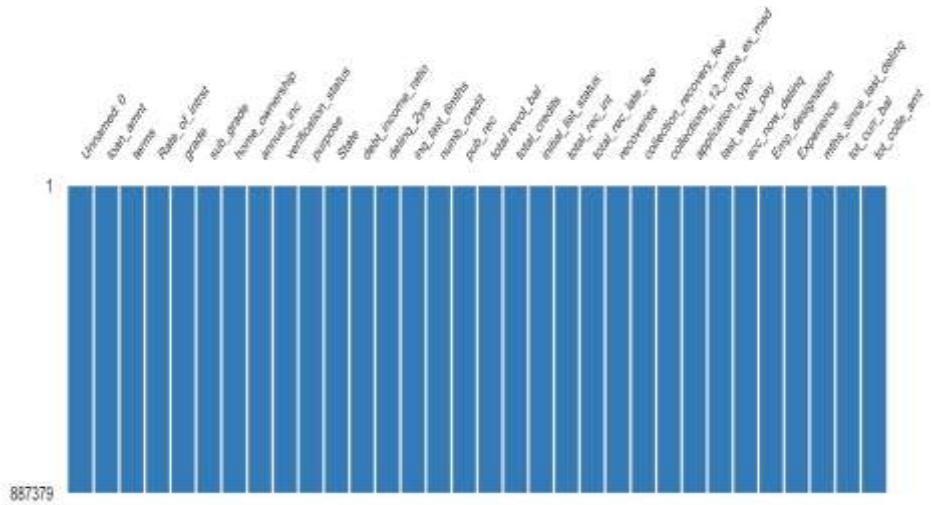


# EDA( VISUALIZATION) Contd..

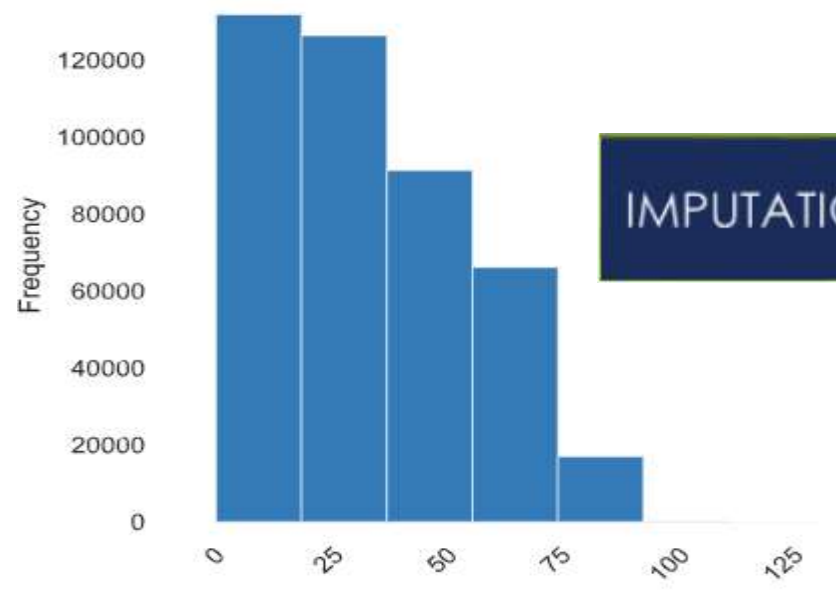
NULL VALUES



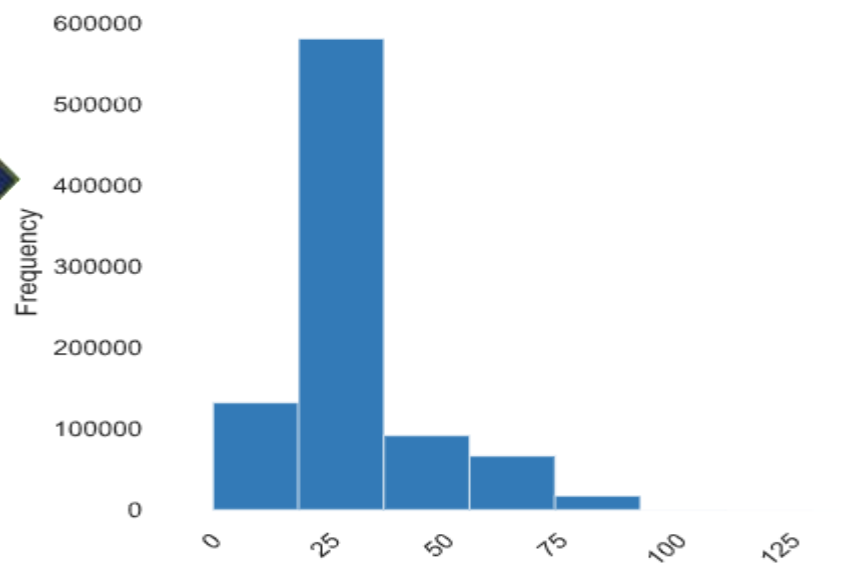
NO NULL VALUES



‘mths\_since\_last\_delinq’



Histogram with fixed size bins (bins=10)



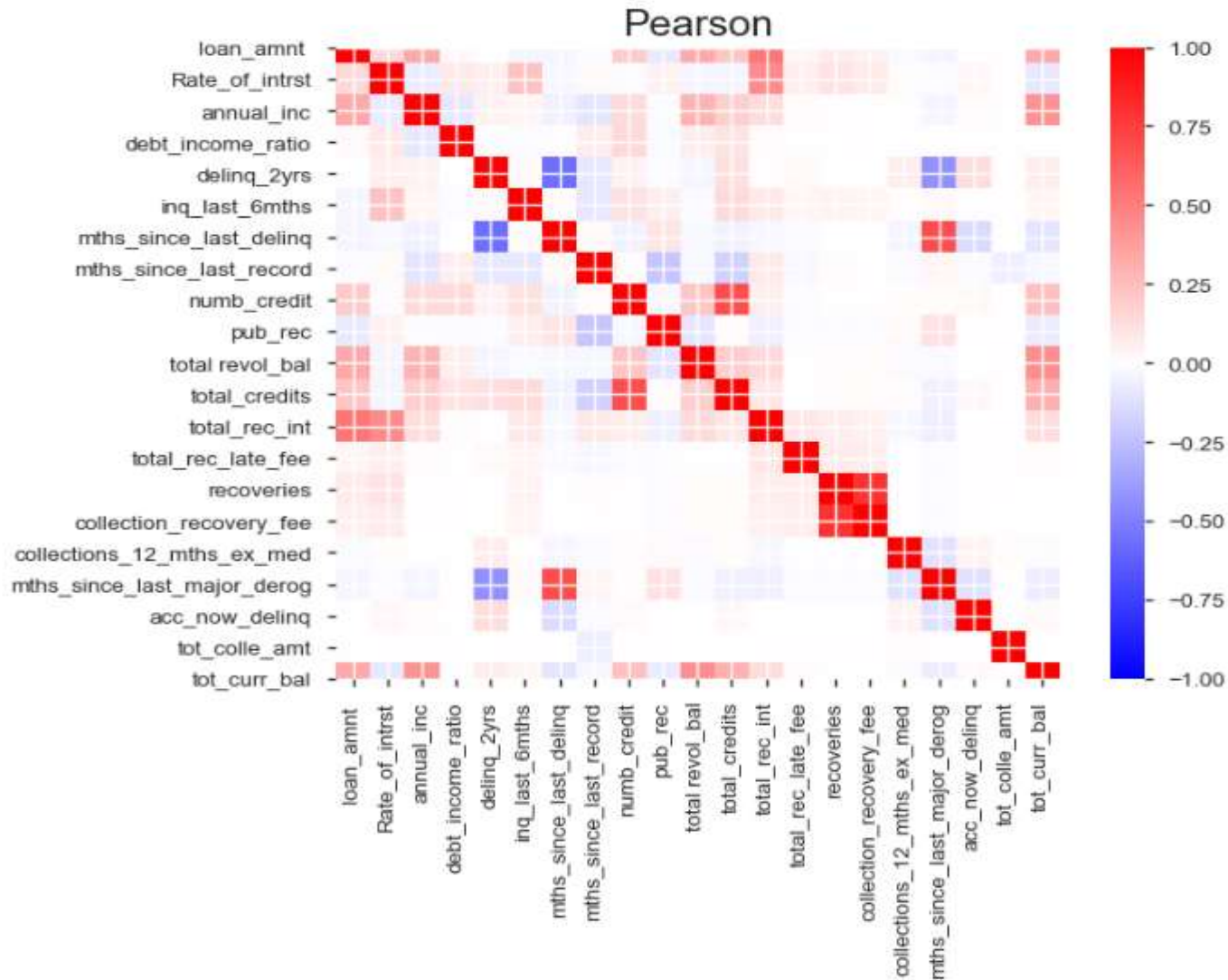
Histogram with fixed size bins (bins=10)

IMPUTATION



ERCA (ECONOMIC RECOVERY AND COLLABORATION ACT)

## CORRELATION



# **Model Building**

## Linear Regression

Dataset Type	Test Size	Random State	R-squared	RMSE
NULL	0.3	0	0.261	18947.63
NOT NULL	0.3	0	0.261	18851.73

# Model Building

## Random Forest

Data Type	Model Type	Test Size	Random State	R-squared (Train)	R-squared (Test)	RMSE (Train)	RMSE (Test)
NOT NULL	Only Numerical	0.3	0	0.66	0.33	13296.07	17496.04
NOT NULL	Only Numerical ( Feature Importance)	0.3	0	0.804	0.33	10092.98	17565.39
NOT NULL	Both Numerical& Categorical	0.3	0	0.805	0.33	10061.32	17574.13
NOT NULL	Both Numerical& Categorical (Feature Importance)	0.3	0	0.814	0.326	9842.85	17629.99

# Model Building

## XGBoost

Dataset Type	Encoding	Min. Child Weight	Max Depth	Learning Rate	Gamma	R-squared (Train)	R-squared (Test)	RMSE (Train)	RMSE (Test)
NOT NULL	Label Encoding	3	6	0.15	0.4	0.55	0.35	15162.58	17242.74
NOT NULL	Frequency Encoding	5	3	0.08	0.1	0.43	0.35	17121.70	17260.07

## **Model Deployment using R shiny / Flask or any other method**

**Challenges faced?**

**How did you overcome?**

**Thank you**