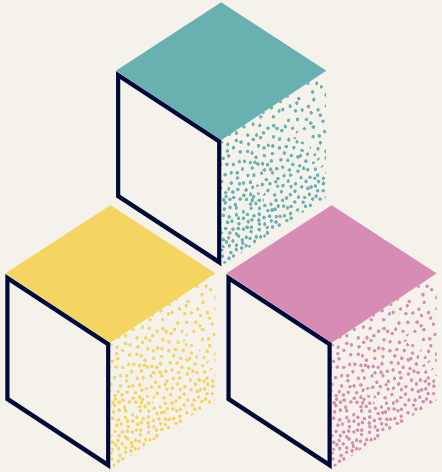


Loan Approval Forecasting

01

Business Overview

Business Goal and Dataset

- KESUS seeks to automate the loan qualifying procedure based on the information provided by the customers.
- KESUS is expecting that the development of the Machine Learning Model will help the company predict loan approval.
- The Machine Learning Model will analyze customer data provided in the dataset.

Data Summary

- The dataset consists of 12 variables
- Eight categorical variables
- Four continuous variables

Variable Name	Description
Gender	Applicant's Gender (Male or Female)
Married	Applicant's Marital Status (Married or Not Married)
Dependents	Number of people dependent on Applicant
Education	Applicant's education (Graduate or Not Graduate)
Self_Employed	Employment Status (Self Employed-Y or Employed by Others - N)
ApplicantIncome	Applicant's monthly salary/income
CoapplicantIncome	Additional applicant's salary or income
LoanAmount	Loan Amount approved for the applicant
LoanTerm_Month	Loan repayment period (in days)
Credit_History	Record for credit history (0 for bad credit history, 1 for good history)
Property_Area	Location of property (Rural/Semi-Urban/Urban
Loan_Status	Status of Loan Approval (Y for Approved and N for not accepted)

PROBLEM STATEMENT

Analyse the correlation among variables and their dependency on evaluating Loan Amount

SOLUTION

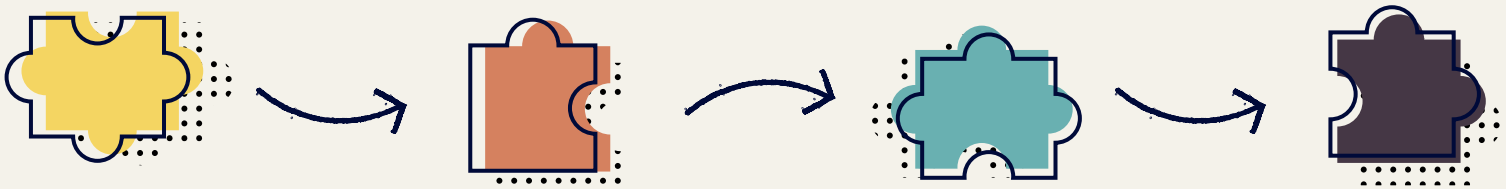
Predict Loan Amount for the Applicant

02

Data Exploration

Data Preparation

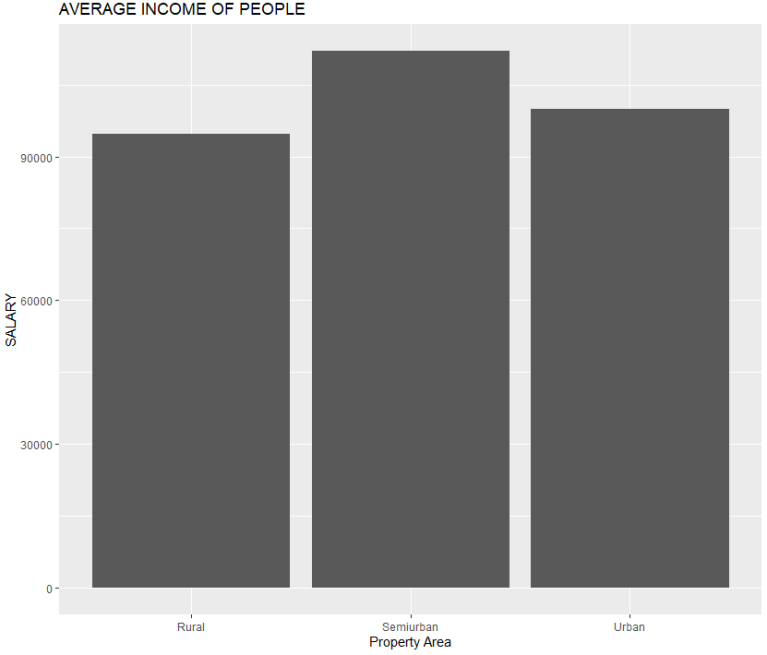
- Tidy data
  - Each variable has its own column
  - Each observation has its own row
  - Each value has its own cell
- Remove rows with Null Values
- Convert categorical data to numeric values

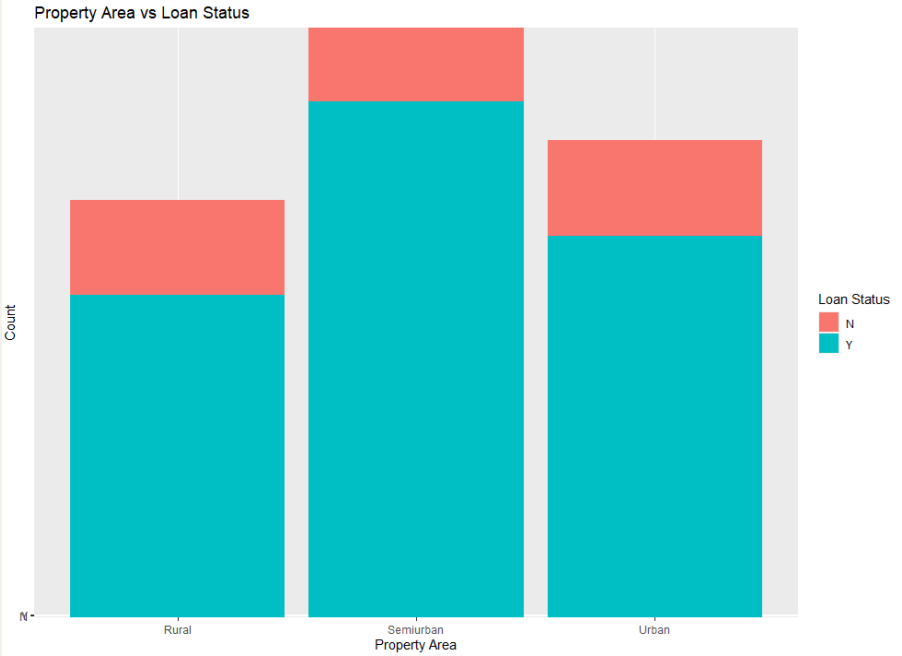


03

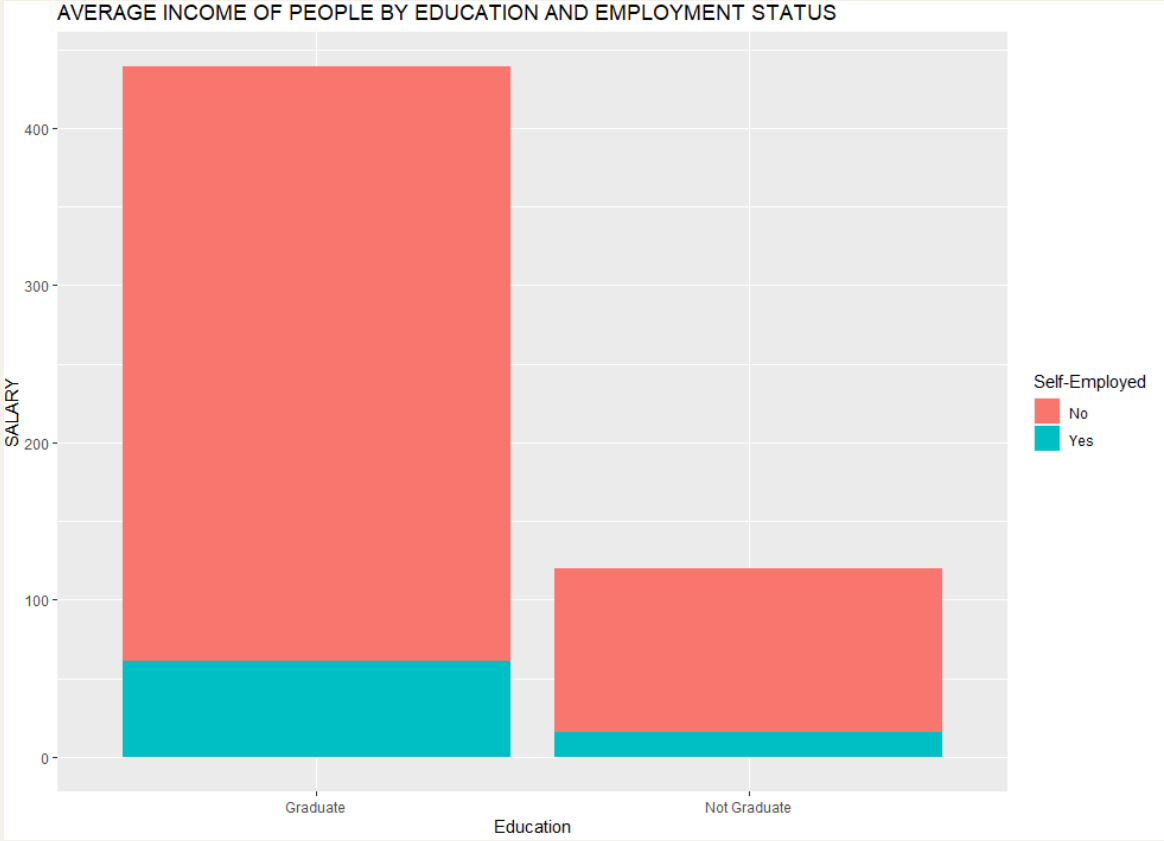
Data Visualisation

Analysis based on Region



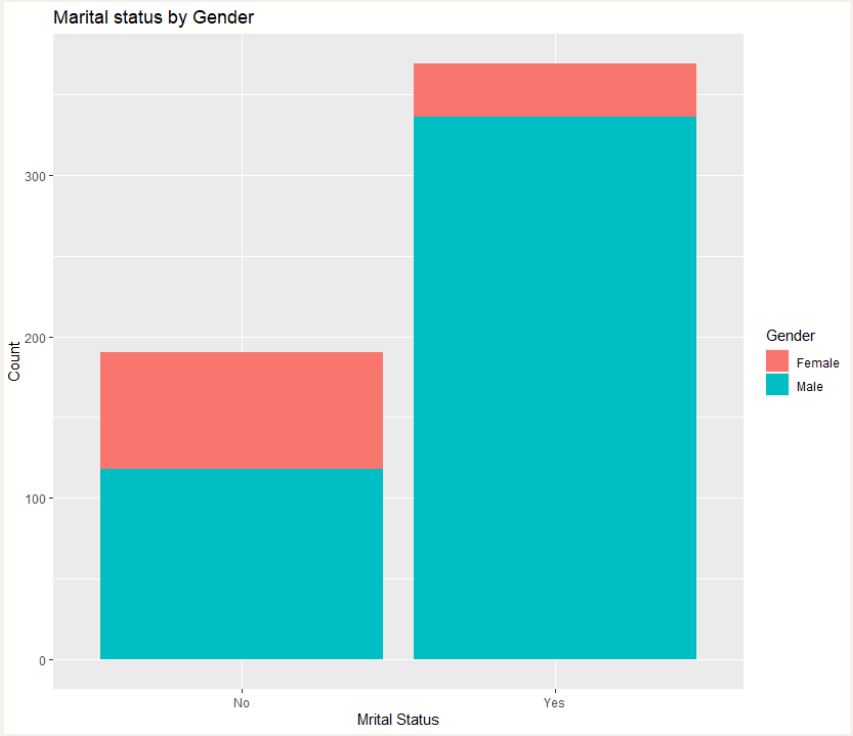


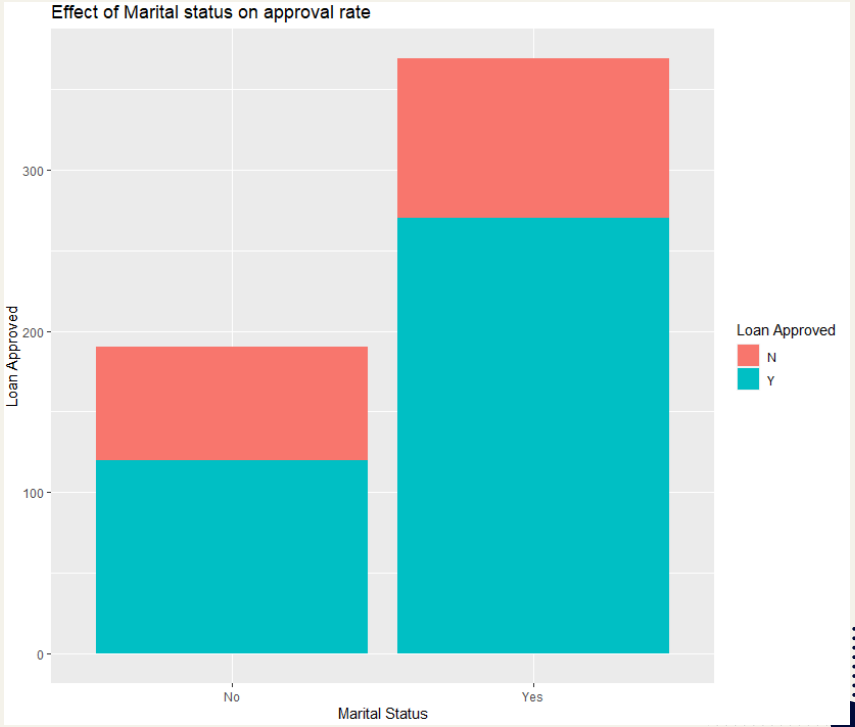
Analysis based on Education and Employment



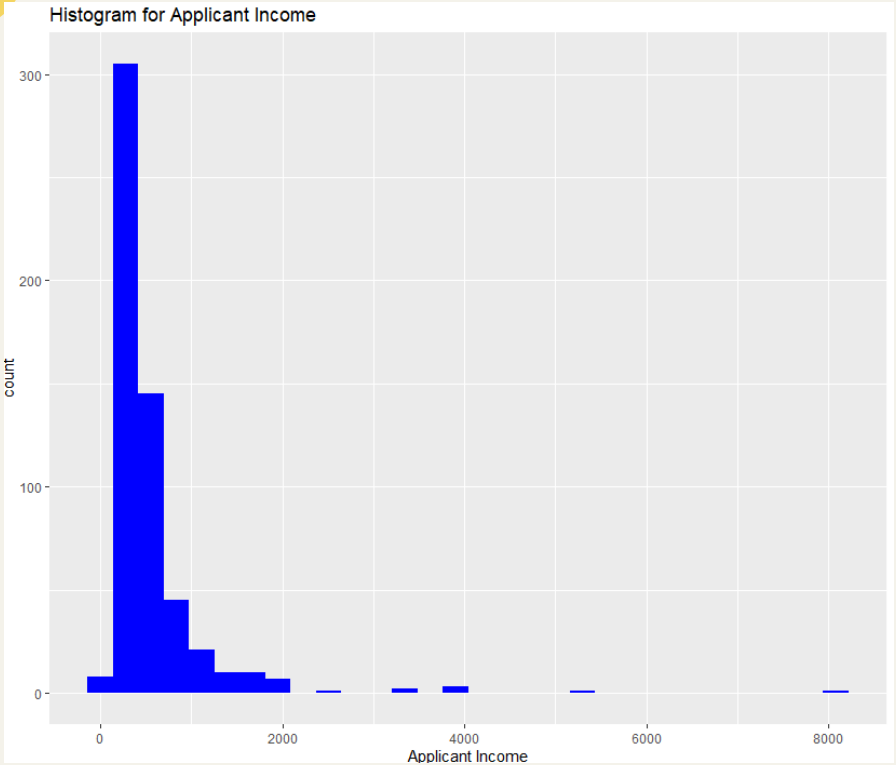
- Majority of graduates as well as non graduates are not self employed.

Analysis based on marital status

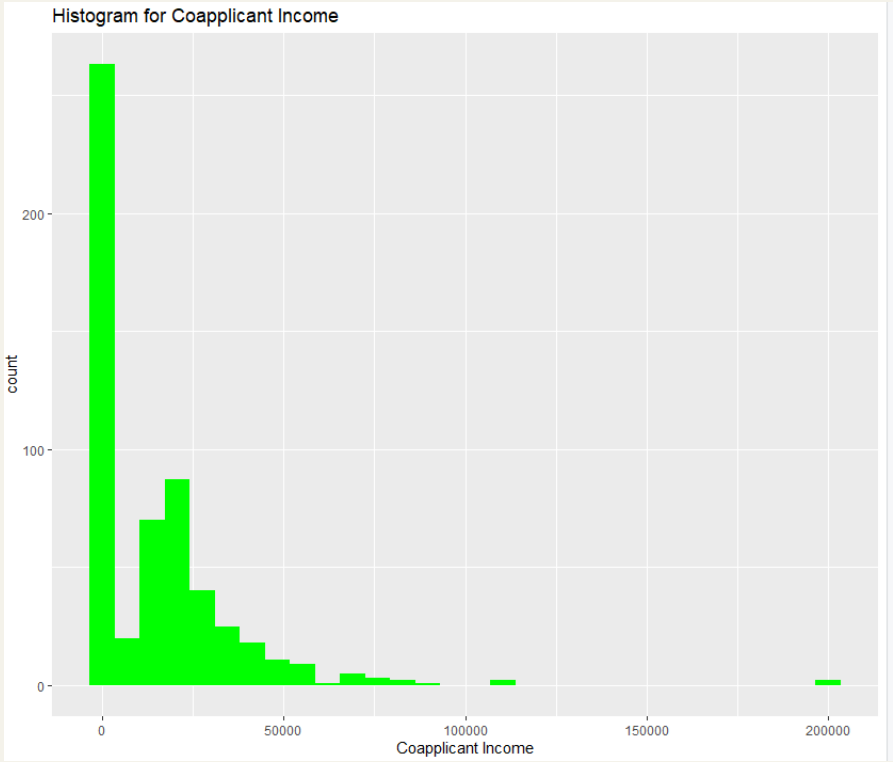




Analysis based on Income

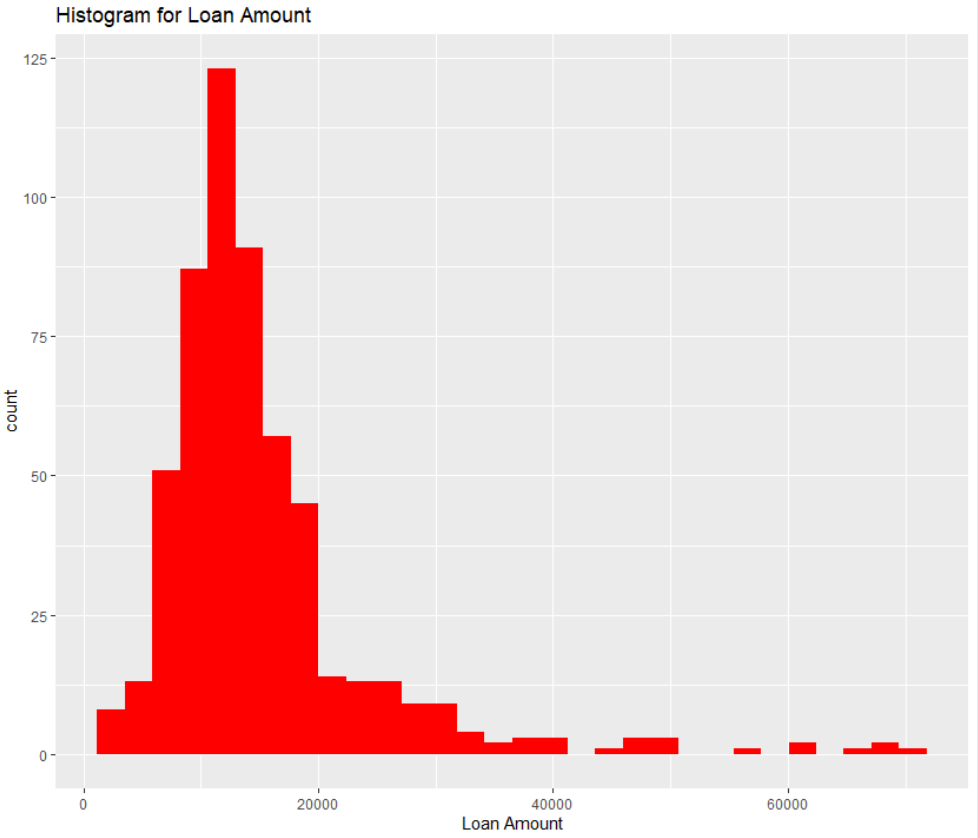


- Majority of applicants have income less than 100000



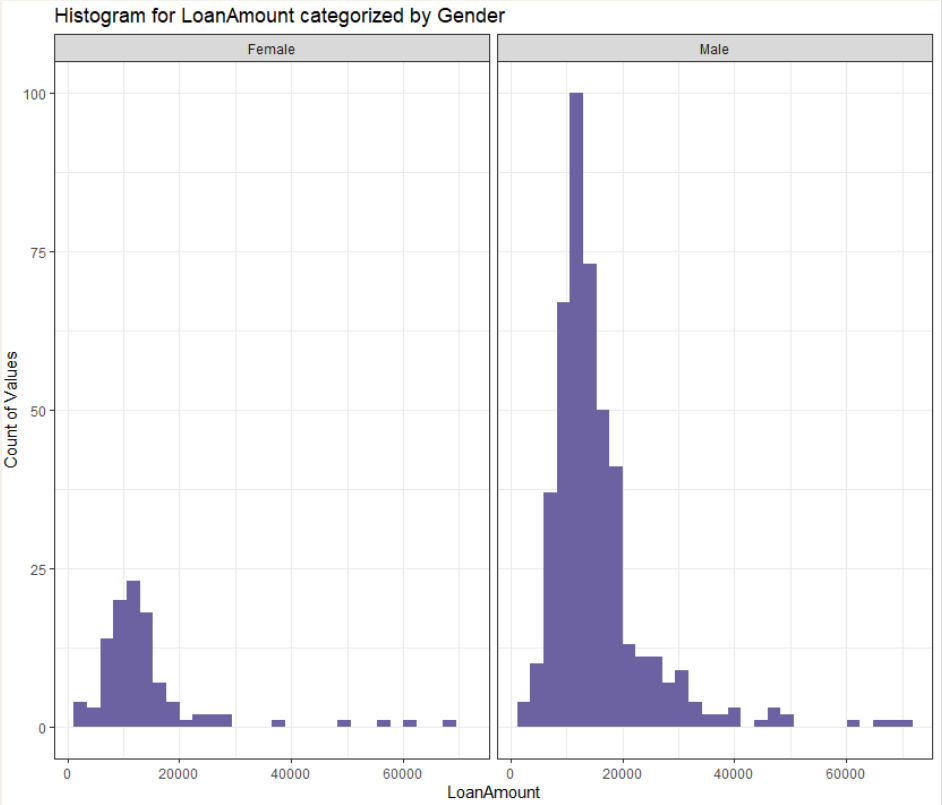
- Majority of co-applicants have low income

Analysis based on Loan Amount



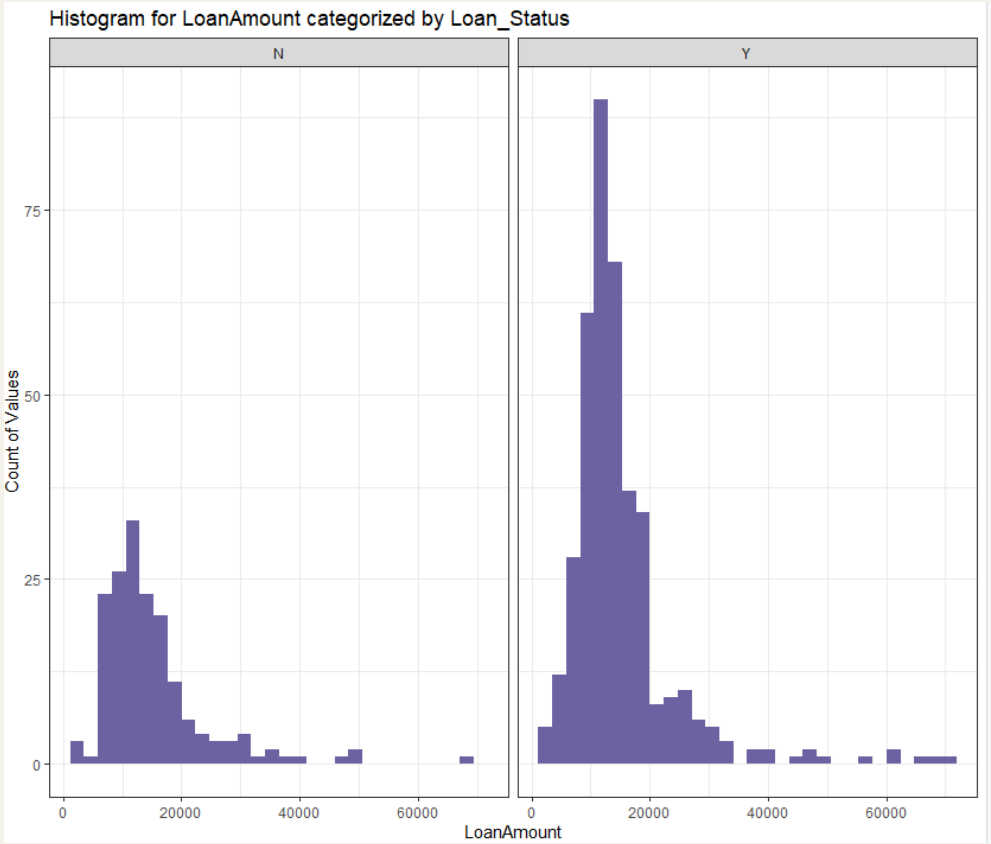
- Histogram for loan amount an applicant can get.
- Very few people got loan amount higher than 30000.

Loan Amount Analysis based on Gender



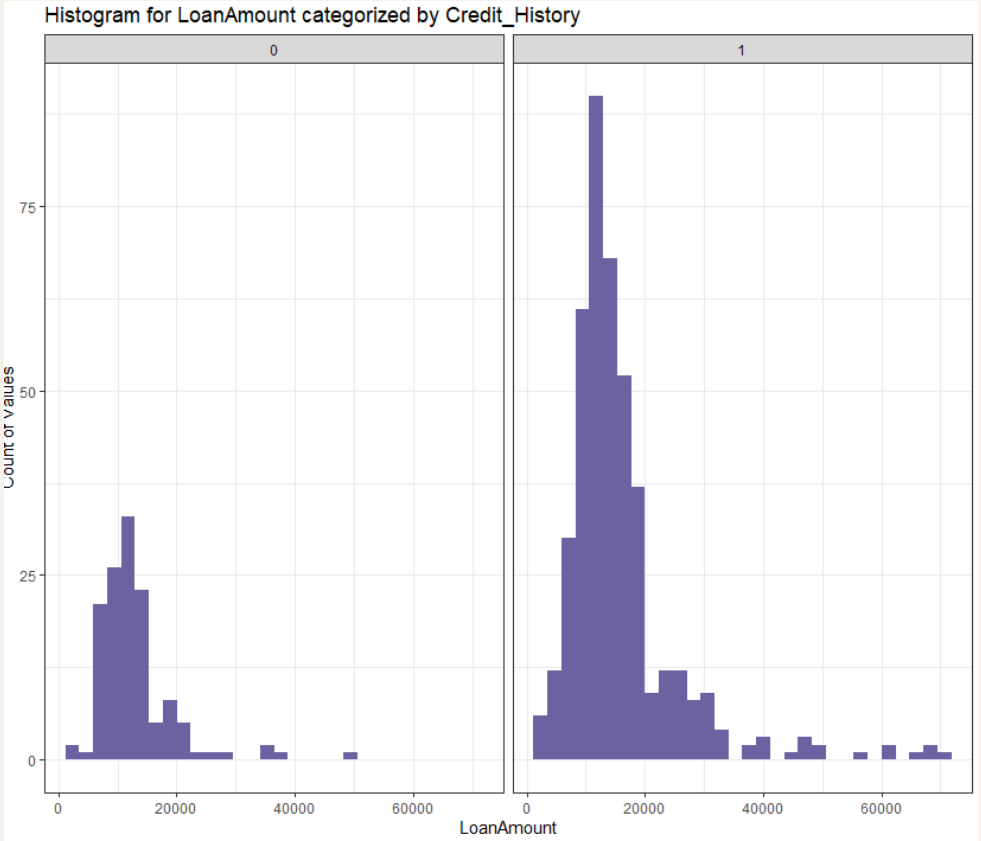
- Loan amount for an applicant based on the gender male or female.
- As per this dataset, male can get higher loan amounts.

Loan Status Analysis based on Amount



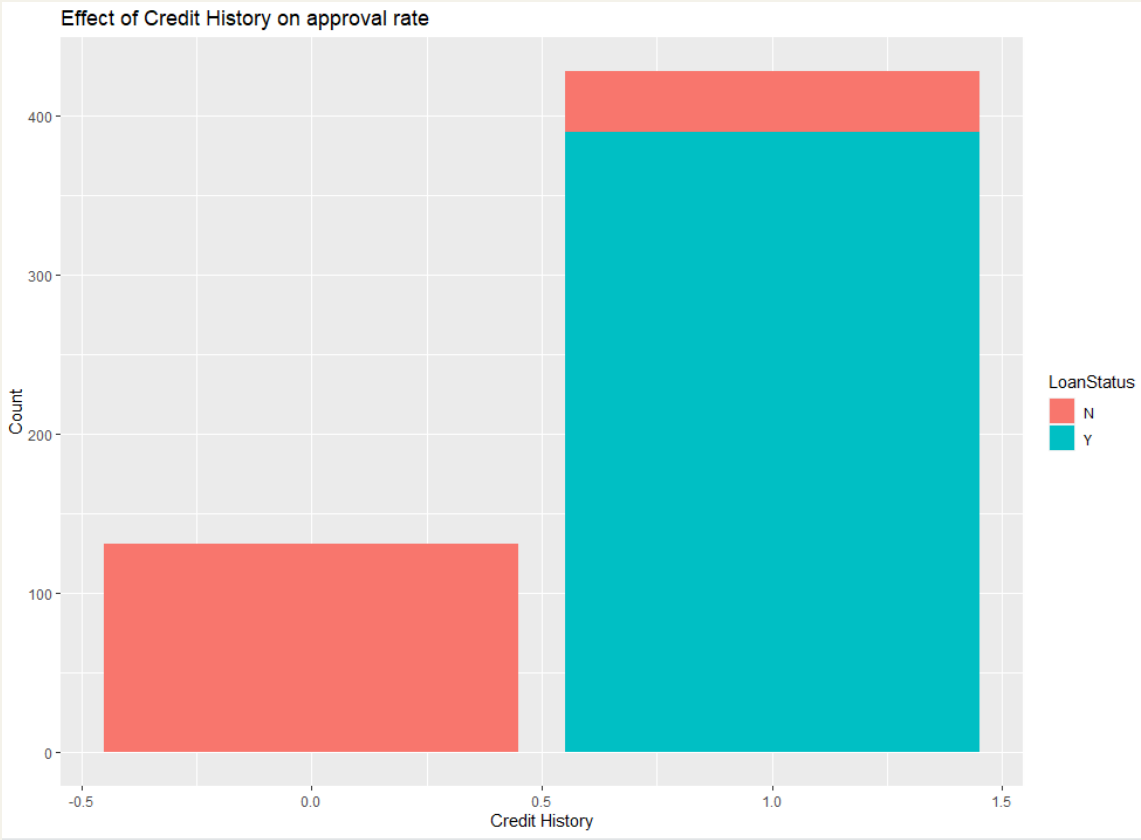
- Loan Approval rate with respect to the loan amount.

Analysis based on Loan Amount and Credit history



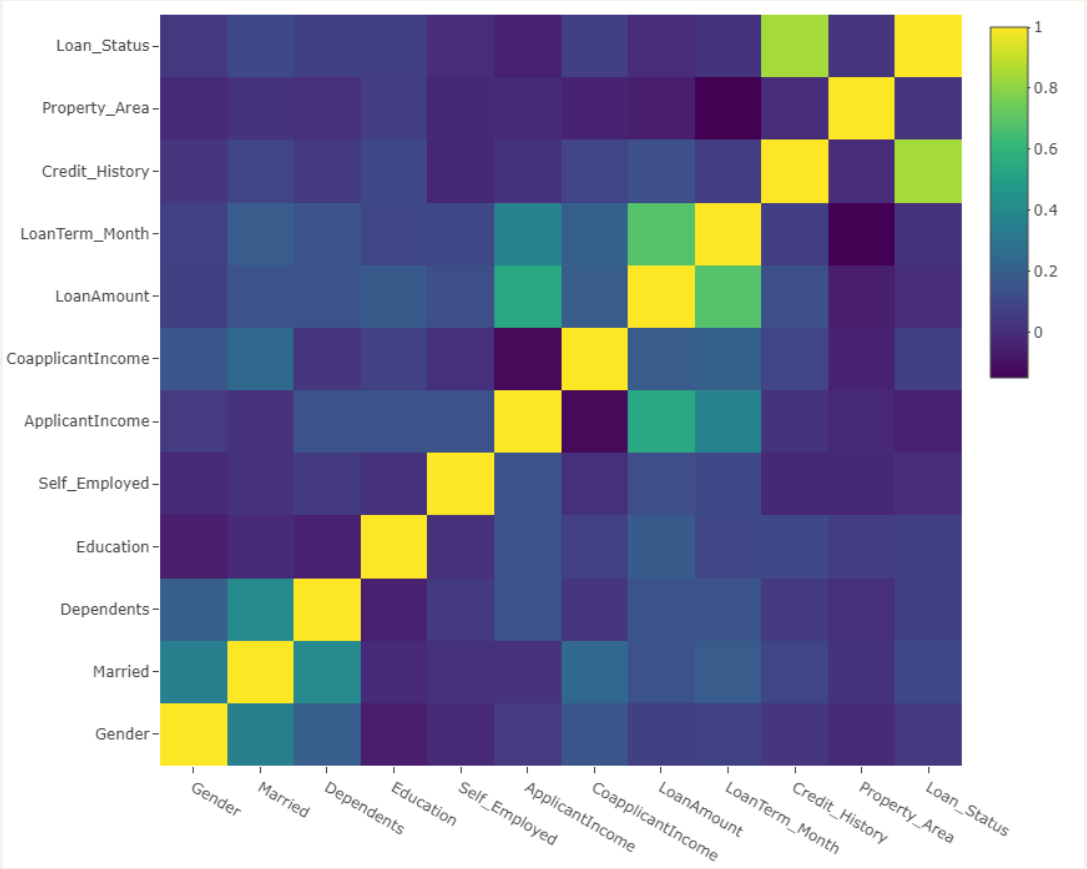
- People with credit history can get high loan amounts.

Analysis based on Credit History vs. Loan Approval



- Loan Application is rejected if applicant doesn't have credit history
- Higher chances of loan approval if applicant has credit history

Heat Map



Plotting heatmap to analyse correlation among the variables.

1 depicts high correlation and 0 interprets no relation at all.

Correlation

Loan Amount

0.5 ApplicantIncome

Greater the Applicant income, higher the loan amount

0.2 CoapplicantIncome

Greater the co-applicant income, higher the loan amount

0.6 LoanTerm_Month

If loan amount is high, loan repayment period is longer

Loan Status

0.8 Credit_History

Applicant needs to have credit history to get loan approval

	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount	LoanTerm_Month	Credit_History	Property_Area	Loan_Status
Gender	1.0	0.4	0.2	-0.1	0.0	0.0	0.2	0.1	0.1	0.0	0.0	0.0
Married	0.4	1.0	0.4	0.0	0.0	0.0	0.2	0.1	0.2	0.1	0.0	0.1
Dependents	0.2	0.4	1.0	0.0	0.0	0.1	0.0	0.1	0.1	0.0	0.0	0.1
Education	-0.1	0.0	0.0	1.0	0.0	0.1	0.1	0.2	0.1	0.1	0.1	0.1
Self_Employed	0.0	0.0	0.0	0.0	1.0	0.1	0.0	0.1	0.1	0.0	0.0	0.0
ApplicantIncome	0.0	0.0	0.1	0.1	0.1	1.0	-0.1	0.5	0.4	0.0	0.0	-0.1
CoapplicantIncome	0.2	0.2	0.0	0.1	0.0	-0.1	1.0	0.2	0.2	0.1	0.0	0.1
LoanAmount	0.1	0.1	0.1	0.2	0.1	0.5	0.2	1.0	0.7	0.1	-0.1	0.0
LoanTerm_Month	0.1	0.2	0.1	0.1	0.1	0.4	0.2	0.7	1.0	0.1	-0.2	0.0
Credit_History	0.0	0.1	0.0	0.1	0.0	0.0	0.1	0.1	0.1	1.0	0.0	0.8
Property_Area	0.0	0.0	0.0	0.1	0.0	0.0	0.0	-0.1	-0.2	0.0	1.0	0.0
Loan_Status	0.0	0.1	0.1	0.1	0.0	-0.1	0.1	0.0	0.0	0.8	0.0	1.0

>

04

Predicting the Loan Amount

Multiple Regression

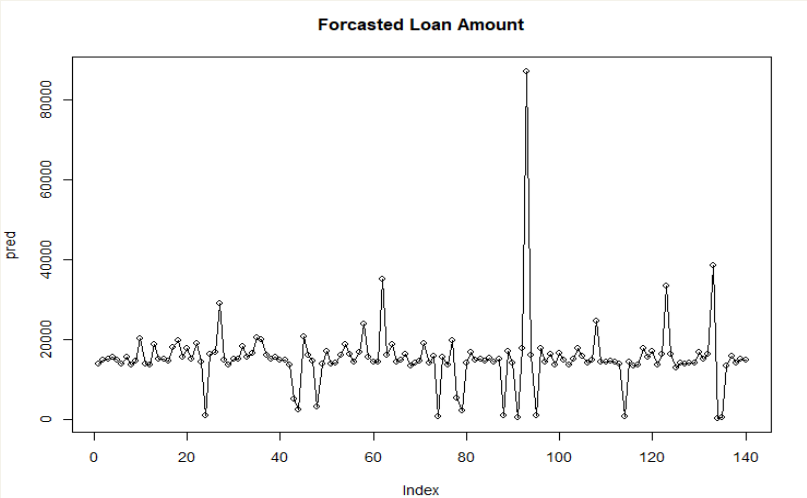
- Dataset split into **Training**(80%) & **Testing** (20%)
- Performed Multiple Regression to predict Loan Amount (**Ŷ)** using:

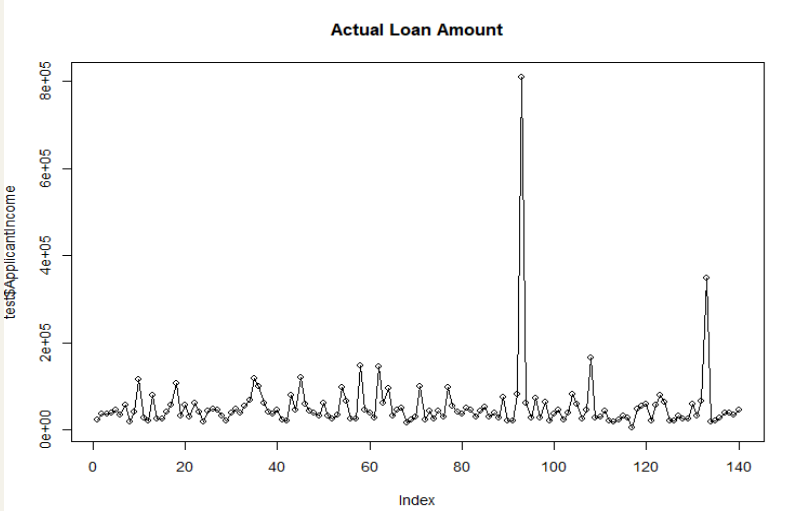
	Coefficients
Intercept	-1.35E+04
ApplicantIncome (x1)	7.84E-02
CoapplicantIncome (x2)	4.95E-02
LoanTerm_Month (x3)	1.03E+03

Loan amount (Ŷ) = a0 + a1* Applicant Income + a2* Co-applicant Income + a3* Loan Term

Conclusion

- Analysis on Testing data (20% of the dataset)
  - MAPE = 31%
  - KPI Prediction = 69% Accuracy





Thus we observe that families having higher income are granted higher loan amounts and higher loan amounts are observed to have longer terms of repayment.

Thank You!


WHEN YOU'VE FINALLY
LOCKED ALL YOUR LOANS
AND RATES DROP AGAIN
top-mind

References

1. https://www.kaggle.com/datasets/burak3ergun/loan-data-set
2. https://www.statmethods.net/stats/regression.html
3. http://www.sthda.com/english/articles/40-regression-analysis/168-multiple-linear-regression-in-r/