

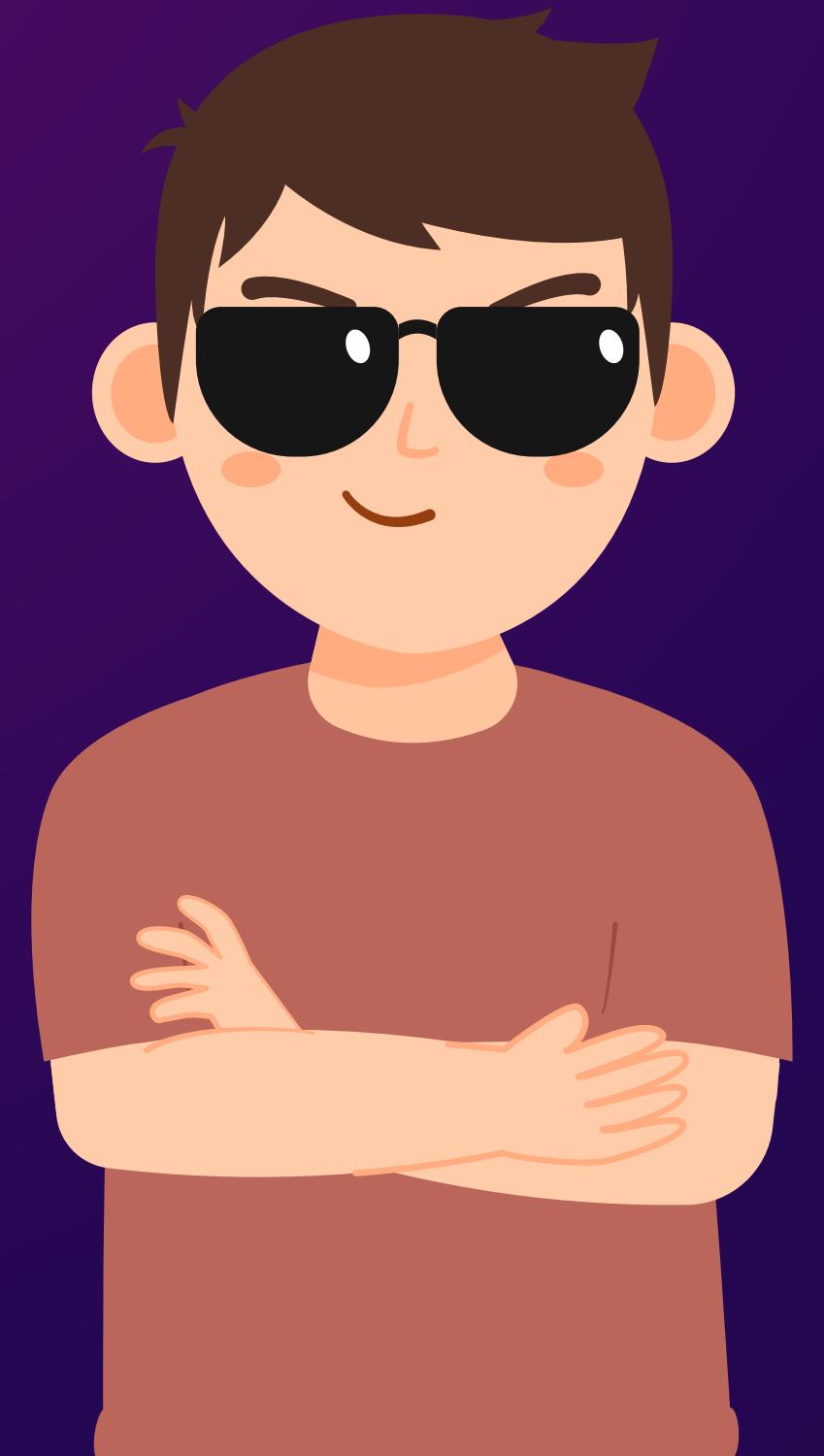
SC1015 MINI PROJECT

LOAN APPROVAL AND LOAN AMOUNT PREDICTION

Group 1 Members: Sim Yang Yue
Lee Yue Hang
Tio Sher Min









Chances of
Loan Approval?

Improvements?

Potential Loan
Amount?

PROBLEM DEFINITION

- How Can We Effectively Predict Applicants' Approval Rate of Receiving a Loan?
- How Can We Predict the Amount of Loan that May Be Obtained by Applicants?





KAI · UPDATED 9 MONTHS AGO

kaggle

◀ 105

New Notebook

Loan-Approval-Prediction-Dataset

Loan Approval Dataset used for Prediction Models

Number of Rows: 4269

Number of Columns: 13

Loan ID
No of Dependents
Education
Self-employed
Income Annum (Annual Income)
Loan Amount
Loan Term
Cibil Score (Credit Score)
Residential Assets Value
Commercial Assets Value
Luxury Assets Value
Bank Asset Value
Loan Status

→ \$200,000 - \$9,900,000

→ 300 - 900

TABLE OF CONTENTS

01 Data Cleaning and Exploratory Analysis

02 Machine Learning

03 Outcome and Insights

DATA CLEANING AND EXPLORATORY ANALYSIS



Our Data

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4269 entries, 0 to 4268
Data columns (total 13 columns):
 #   Column           Non-Null Count  Dtype  
 ---  -- 
 0   loan_id          4269 non-null    int64  
 1   no_of_dependents 4269 non-null    int64  
 2   education        4269 non-null    object  
 3   self_employed     4269 non-null    object  
 4   income_annum     4269 non-null    int64  
 5   loan_amount       4269 non-null    int64  
 6   loan_term         4269 non-null    int64  
 7   cibil_score       4269 non-null    int64  
 8   residential_assets_value 4269 non-null    int64  
 9   commercial_assets_value 4269 non-null    int64  
 10  luxury_assets_value 4269 non-null    int64  
 11  bank_asset_value  4269 non-null    int64  
 12  loan_status       4269 non-null    object  
dtypes: int64(10), object(3)
memory usage: 433.7+ KB
```

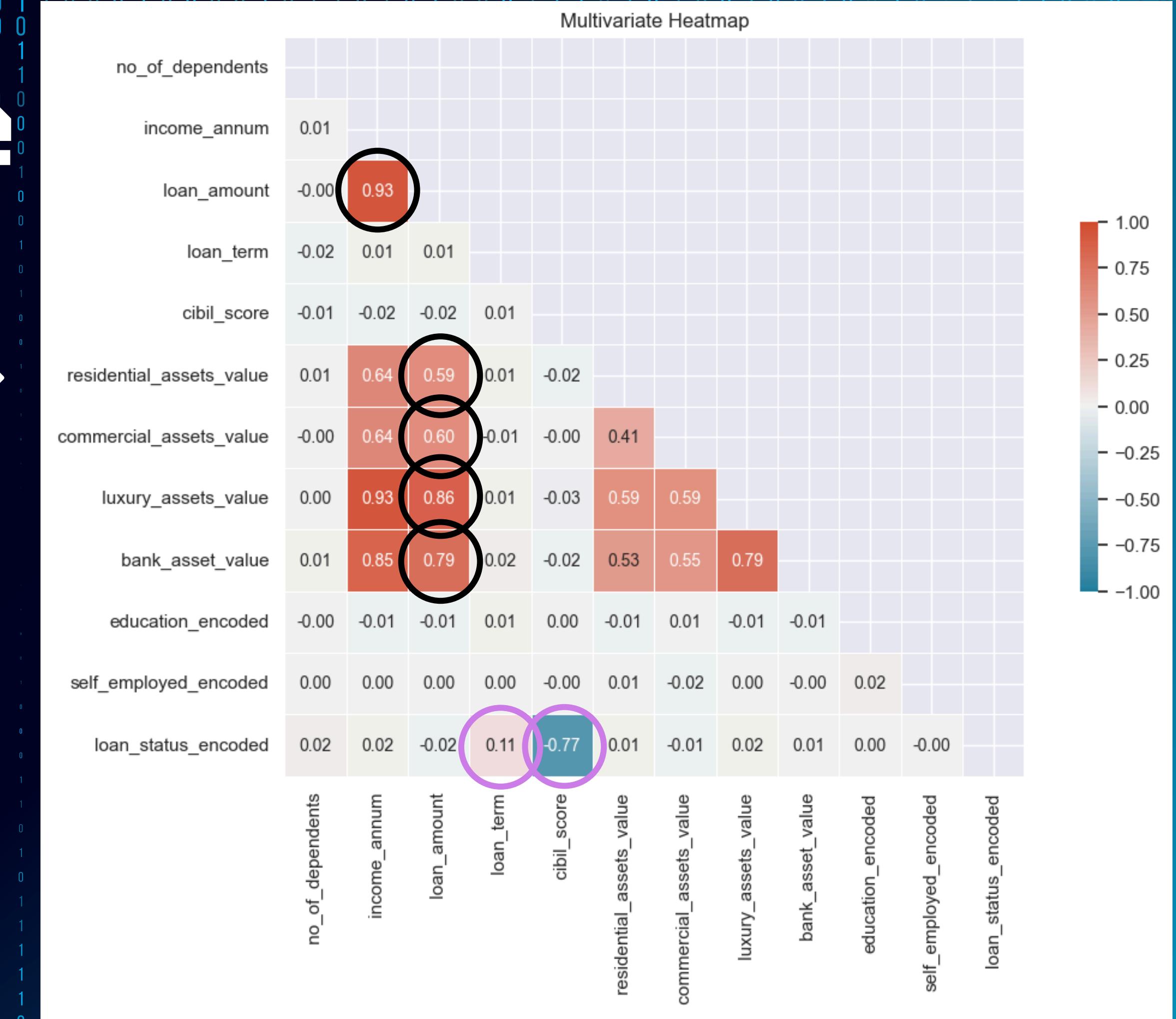
Which Variables Correlate the
Most with Loan Status and
Loan Amount Respectively?

Multivariate Heat Map

Little to No Correlation

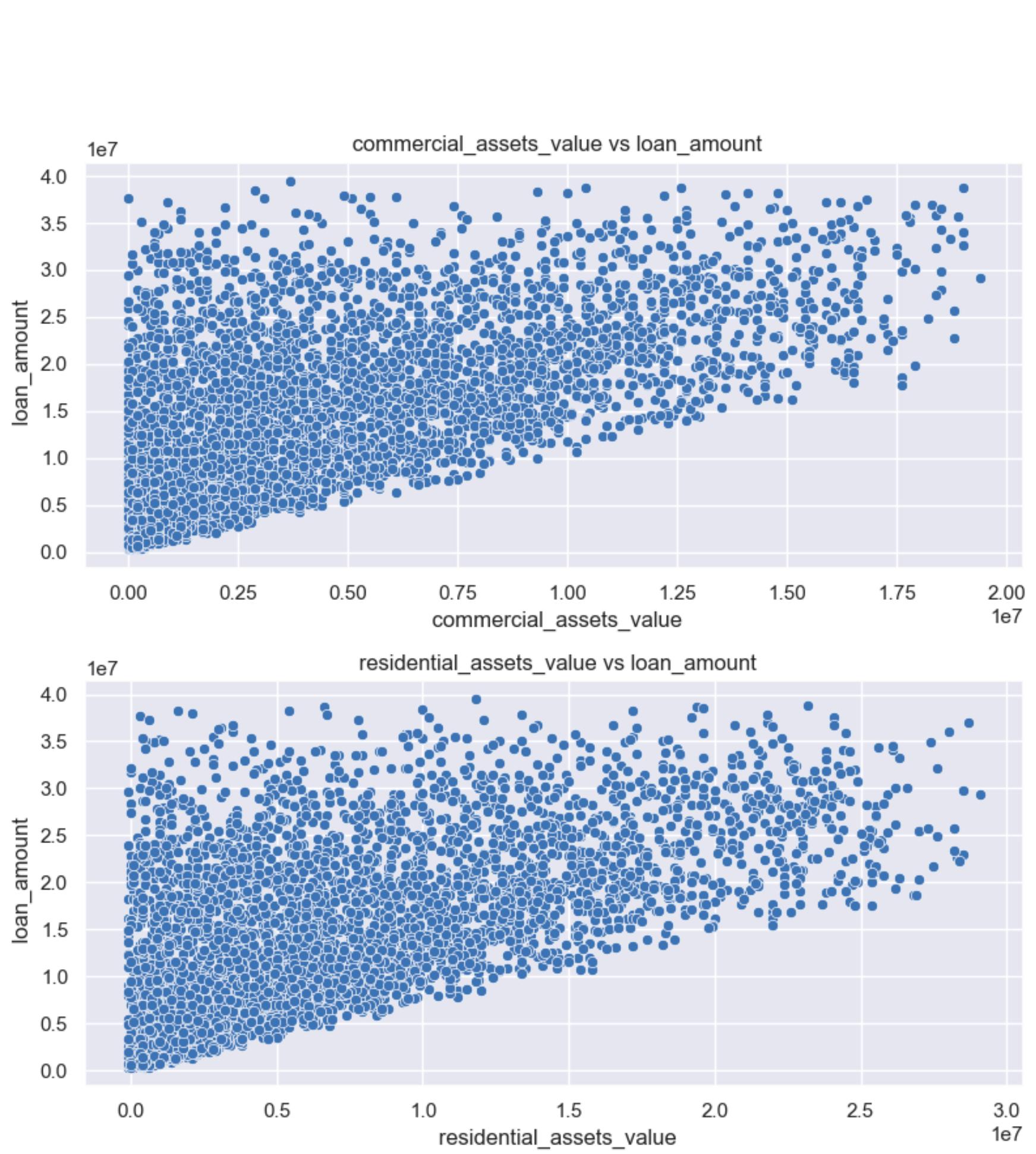
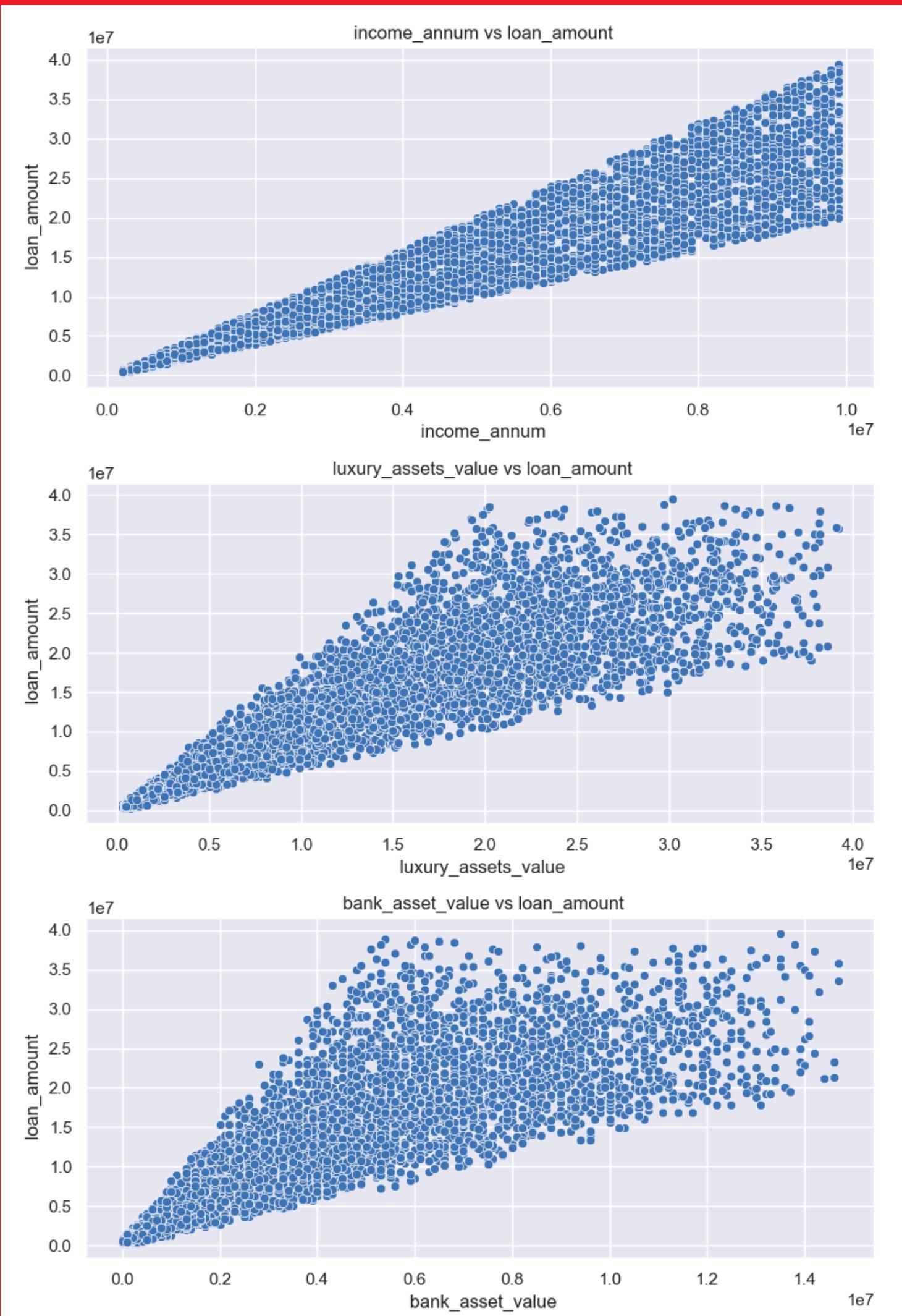
- No of Dependents
- Self-employed
- Loan ID

Filtered Out



Expand on Correlation of Loan Amount

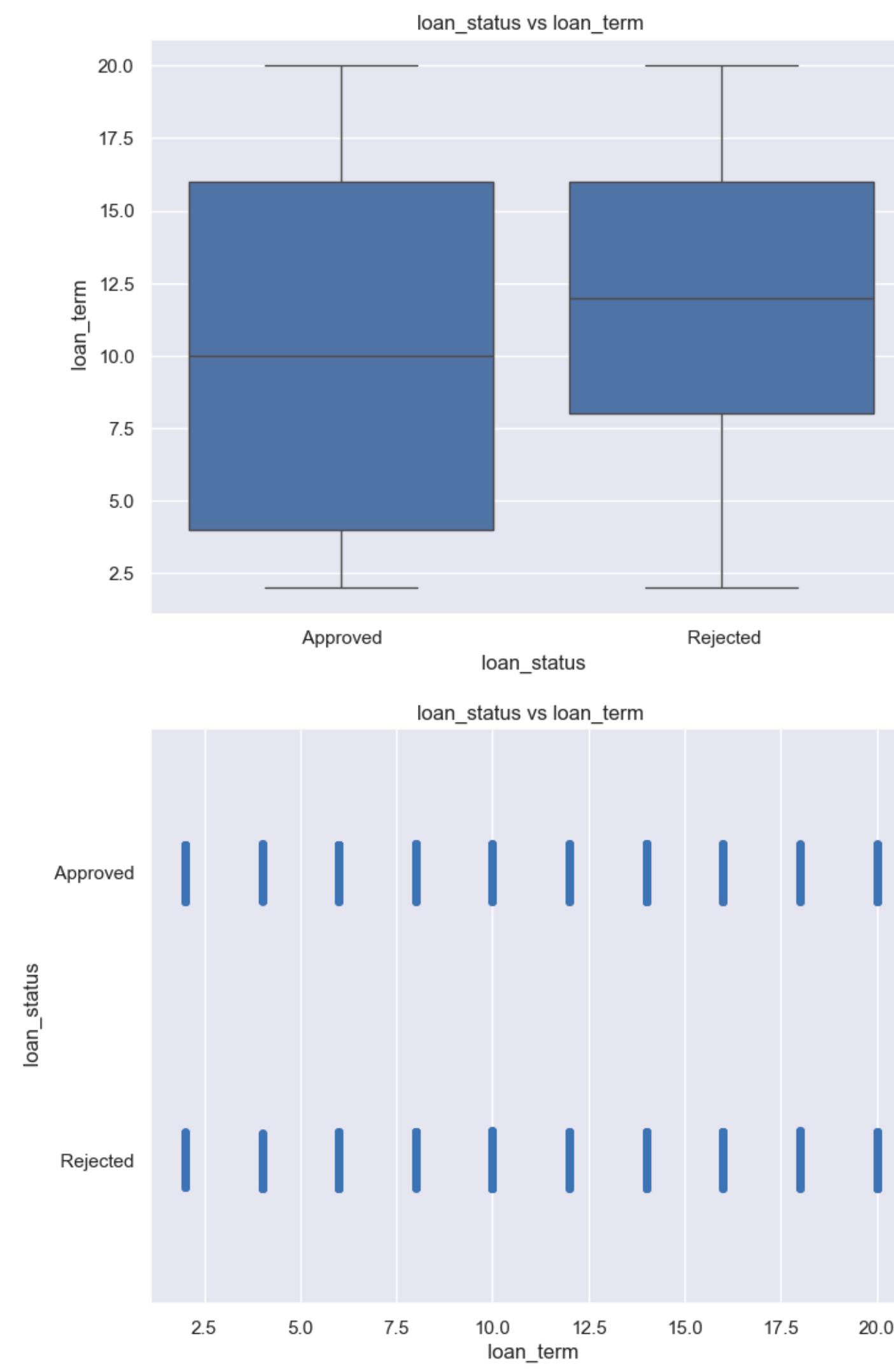
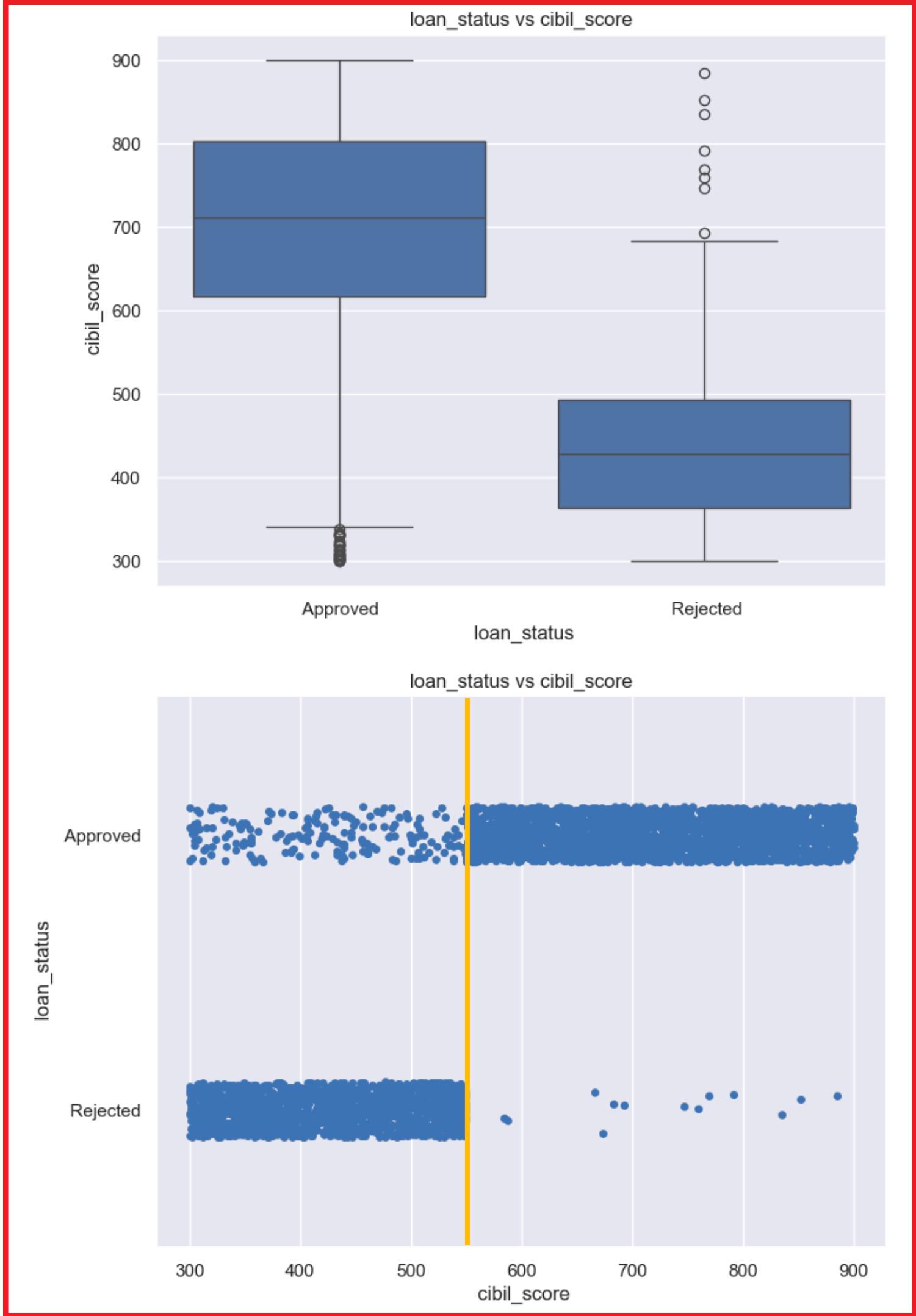
- Income Annum
- Luxury Assets Value
- Bank Asset Value
- Commercial Assets Value
- Residential Assets Value

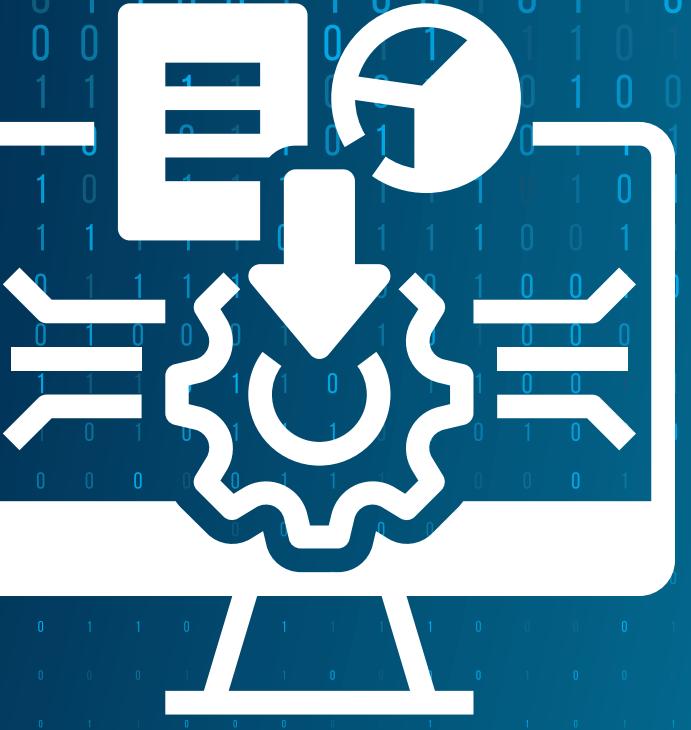


To use in Linear Regression Model Later

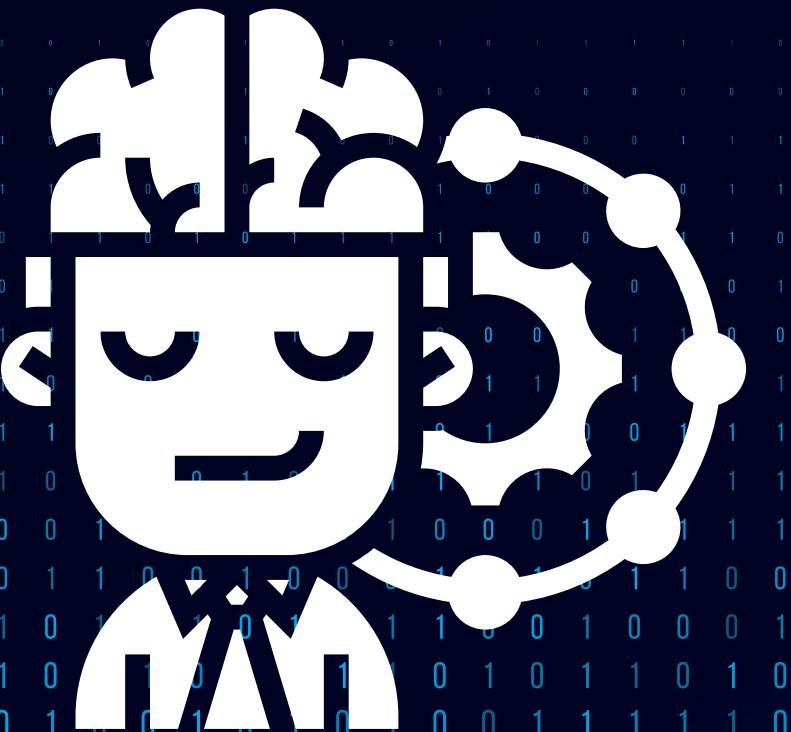
Expand on Correlation of Loan Status

- Credit Score
- Loan Term

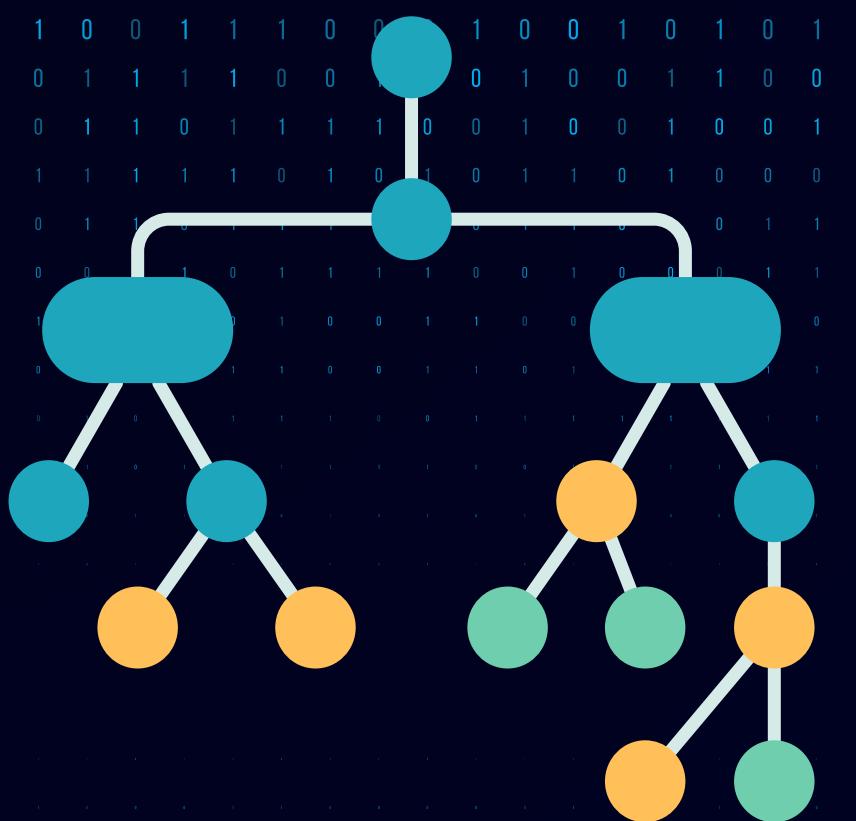




MACHINE LEARNING

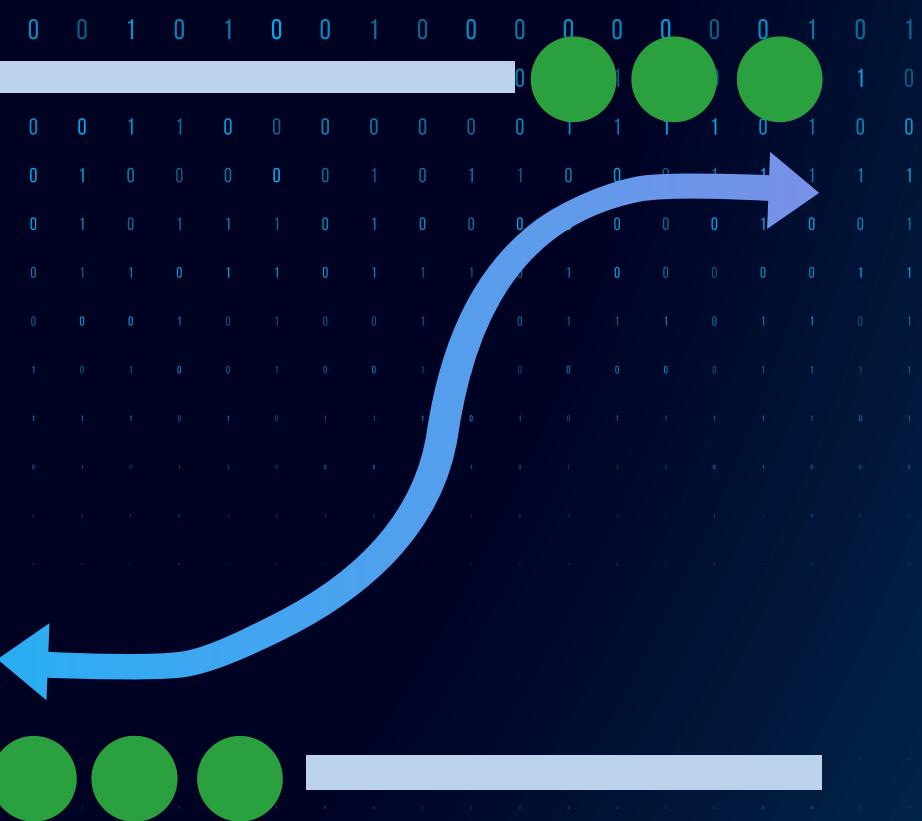


MODELS USED



Decision Tree Model

- Univariate
- Multivariate

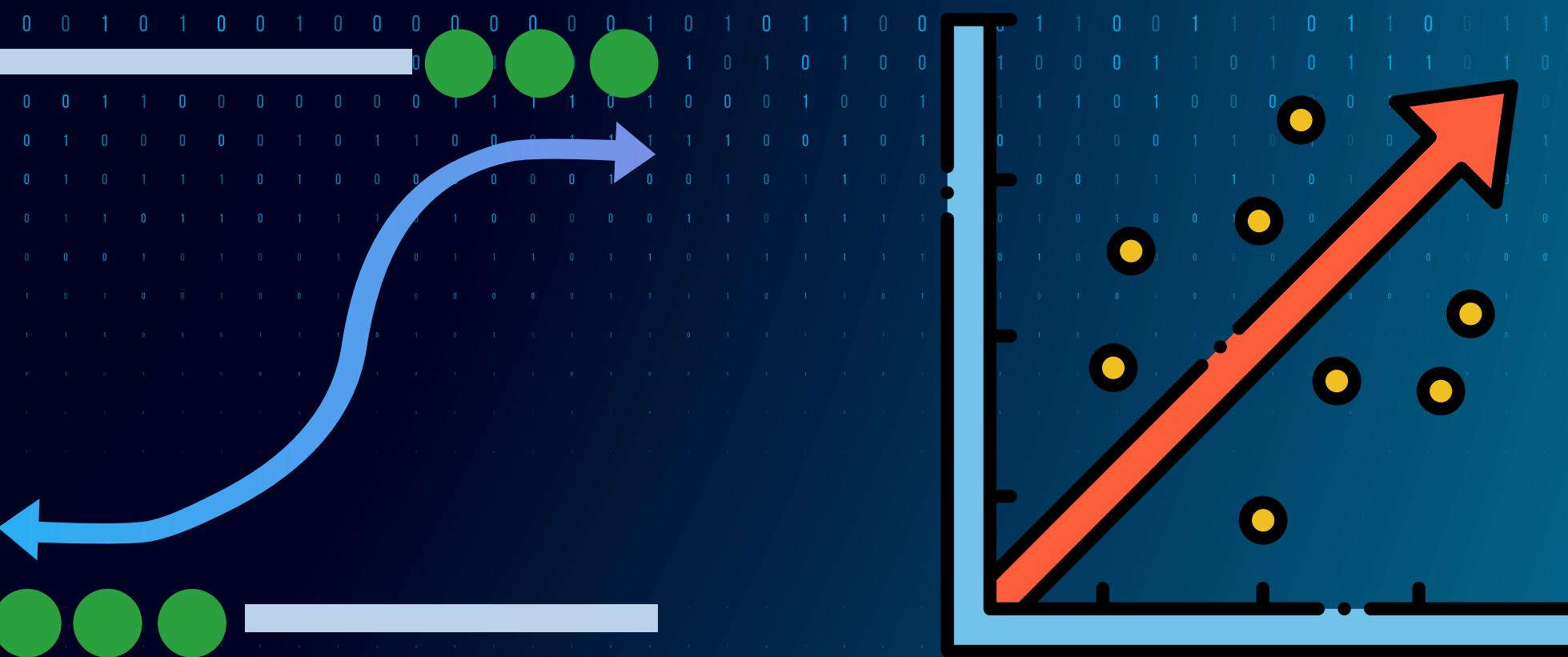


Logistic Regression Model

Linear Regression Model

- Univariate
- Multivariate

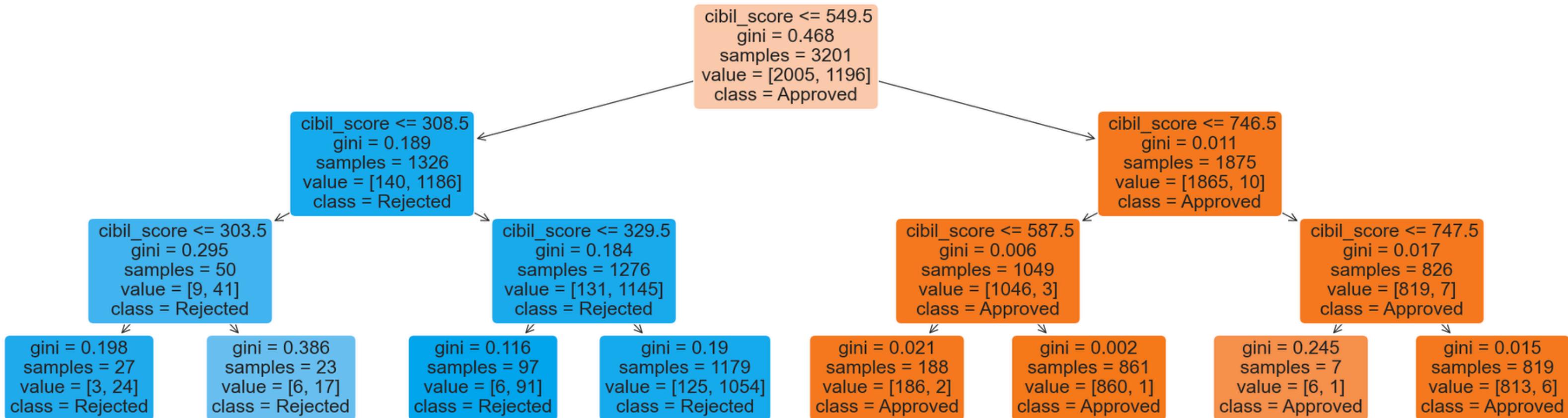
Predict Loan Amount



Predict Loan Approval Rate

UNIVARIATE DECISION TREE

(Loan Approval Rate)



Train set accuracy: 0.9531396438612934

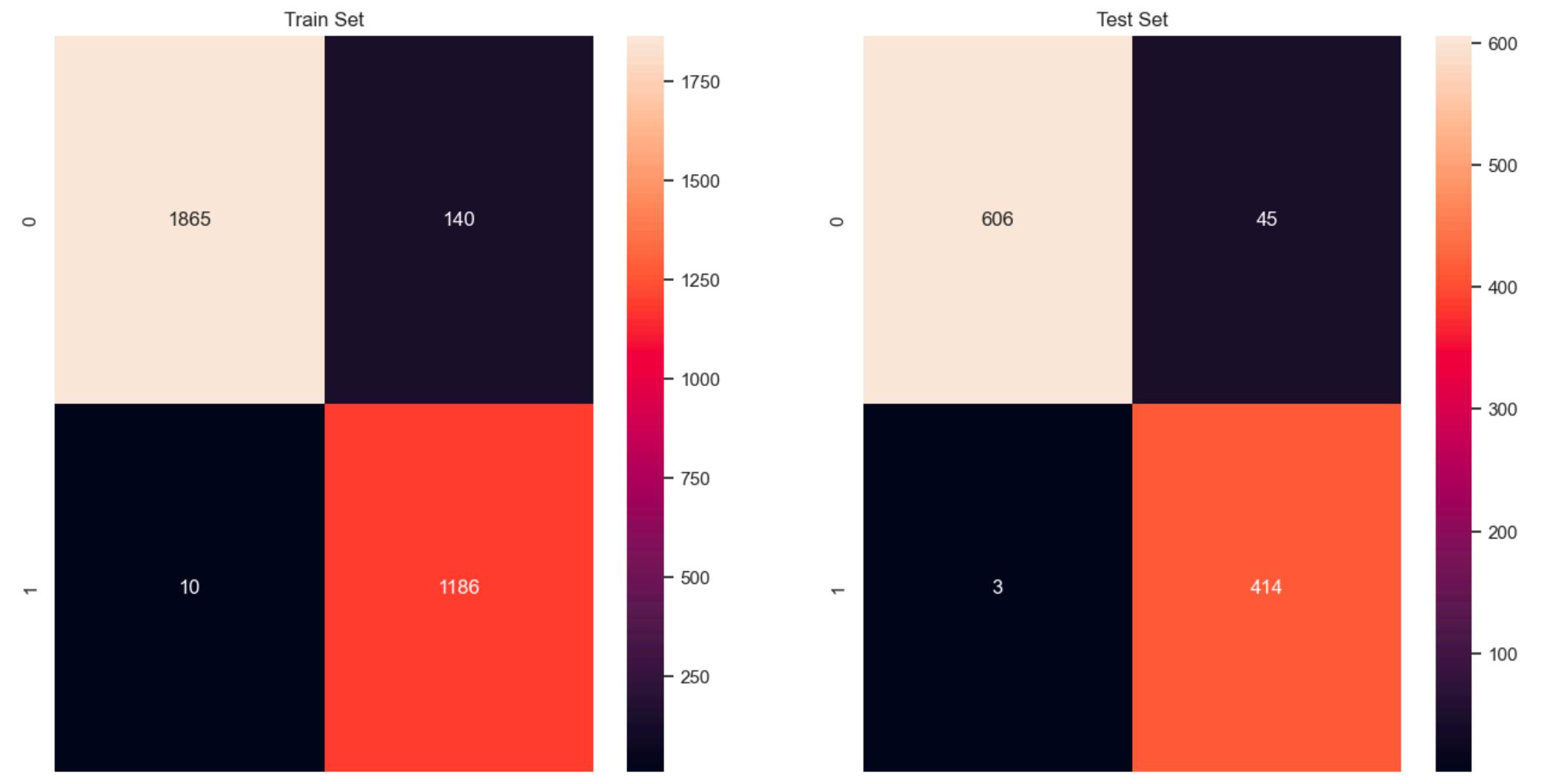
Test set accuracy: 0.9550561797752809

Train set true positive rate: 0.9916387959866221

Train set false positive rate: 0.06982543640897755

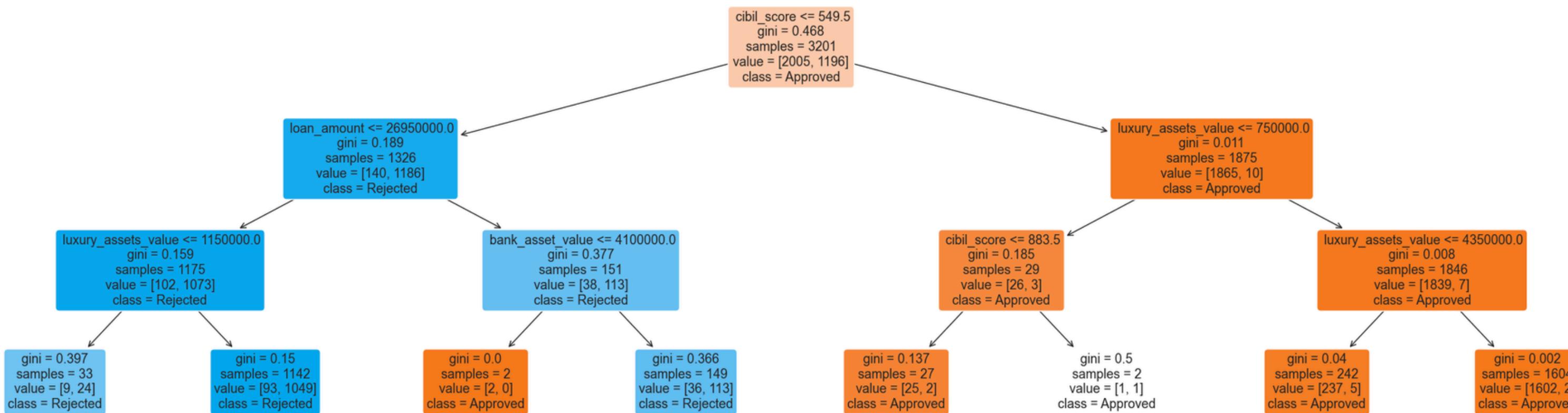
Test set true positive rate: 0.9928057553956835

Test set false positive rate: 0.06912442396313365



MULTIVARIATE DECISION TREE

(Loan Approval Rate)



Train set accuracy: 0.9537644486098095

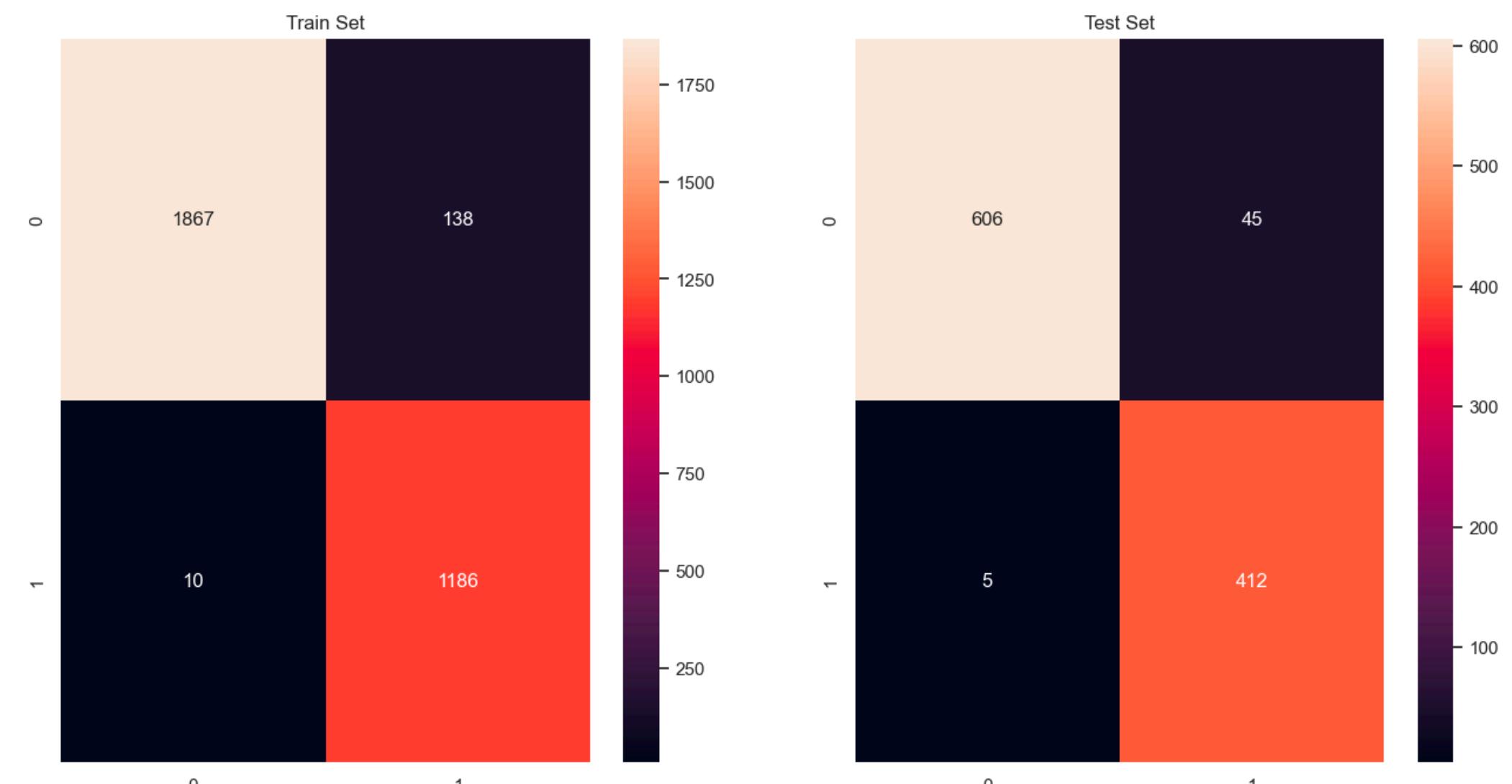
Test set accuracy: 0.9531835205992509

Train set true positive rate: 0.9916387959866221

Train set false positive rate: 0.0688279301745636

Test set true positive rate: 0.988009592326139

Test set false positive rate: 0.06912442396313365



Univariate Decision Tree

Train set accuracy: 0.9531396438612934

Test set accuracy: 0.9550561797752809

Train set true positive rate: 0.9916387959866221

Train set false positive rate: 0.06982543640897755

Test set true positive rate: 0.9928057553956835

Test set false positive rate: 0.06912442396313365

Multivariate Decision Tree

Train set accuracy: 0.9537644486098095

Test set accuracy: 0.9531835205992509

Train set true positive rate: 0.9916387959866221

Train set false positive rate: 0.0688279301745636

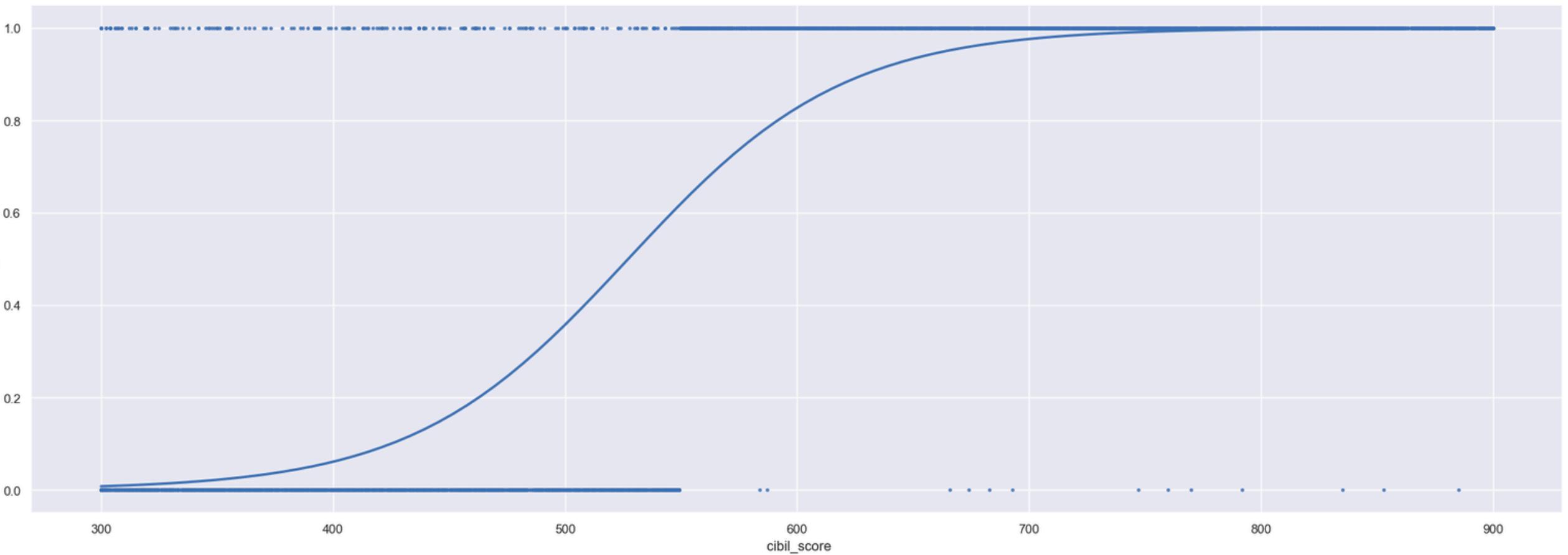
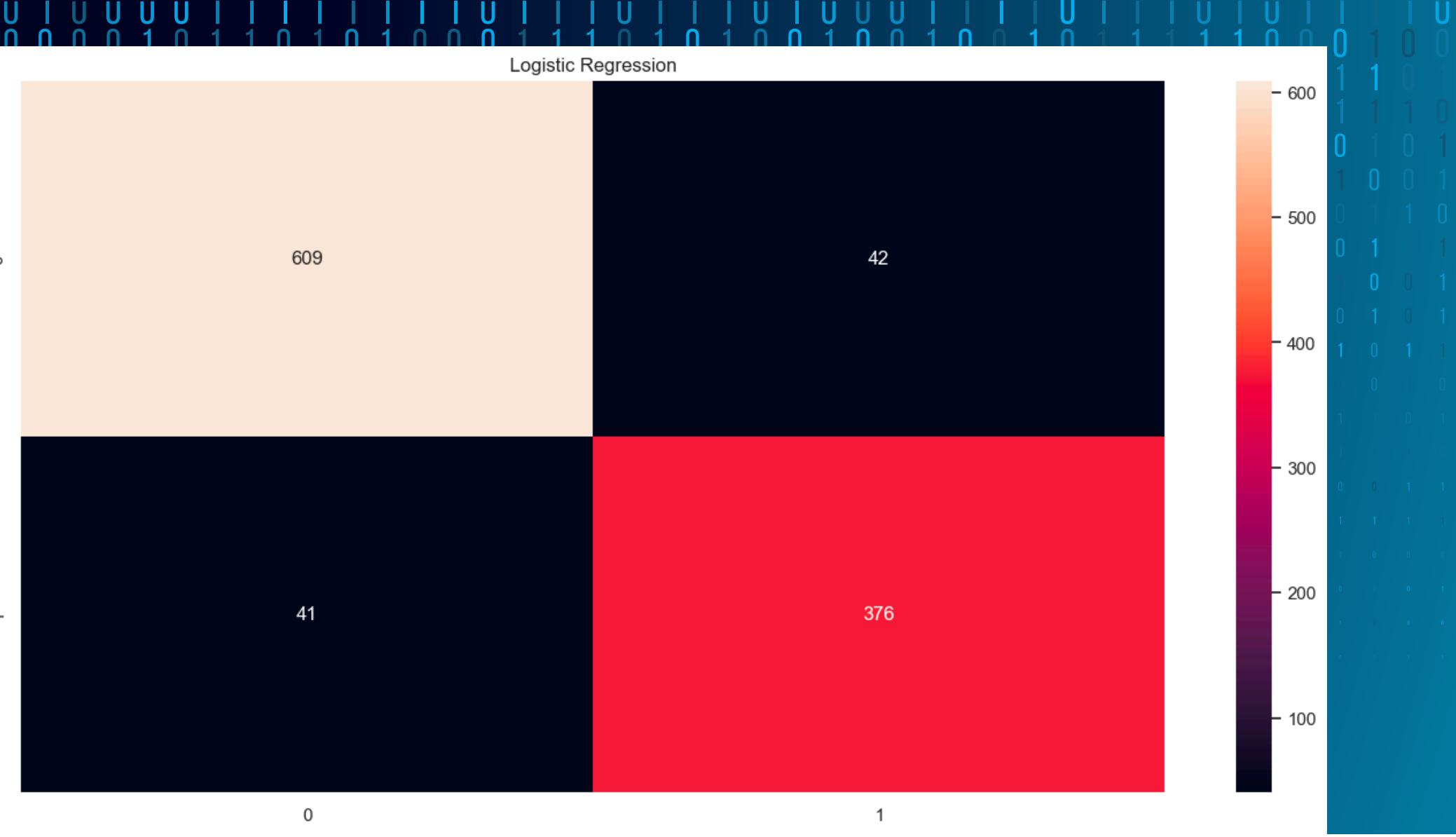
Test set true positive rate: 0.988009592326139

Test set false positive rate: 0.06912442396313365

LOGISTIC REGRESSION MODEL

(Loan Approval Rate)

Accuracy on Test Set: 0.92



BUILDING LINEAR REGRESSION MODELS

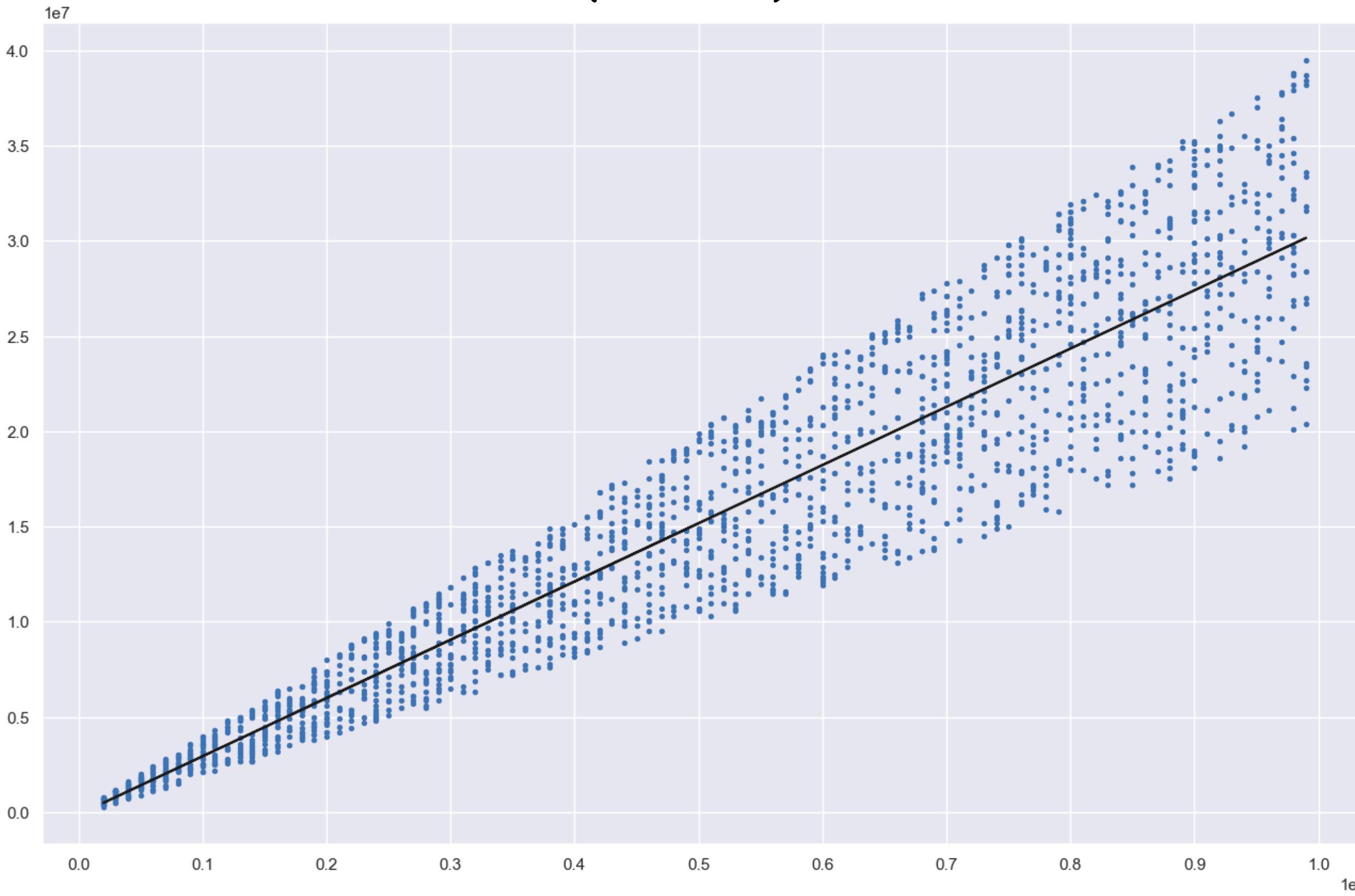
(Loan Amount)

Filter Out Rejected Data

- To increase reliability of loan amount to be received

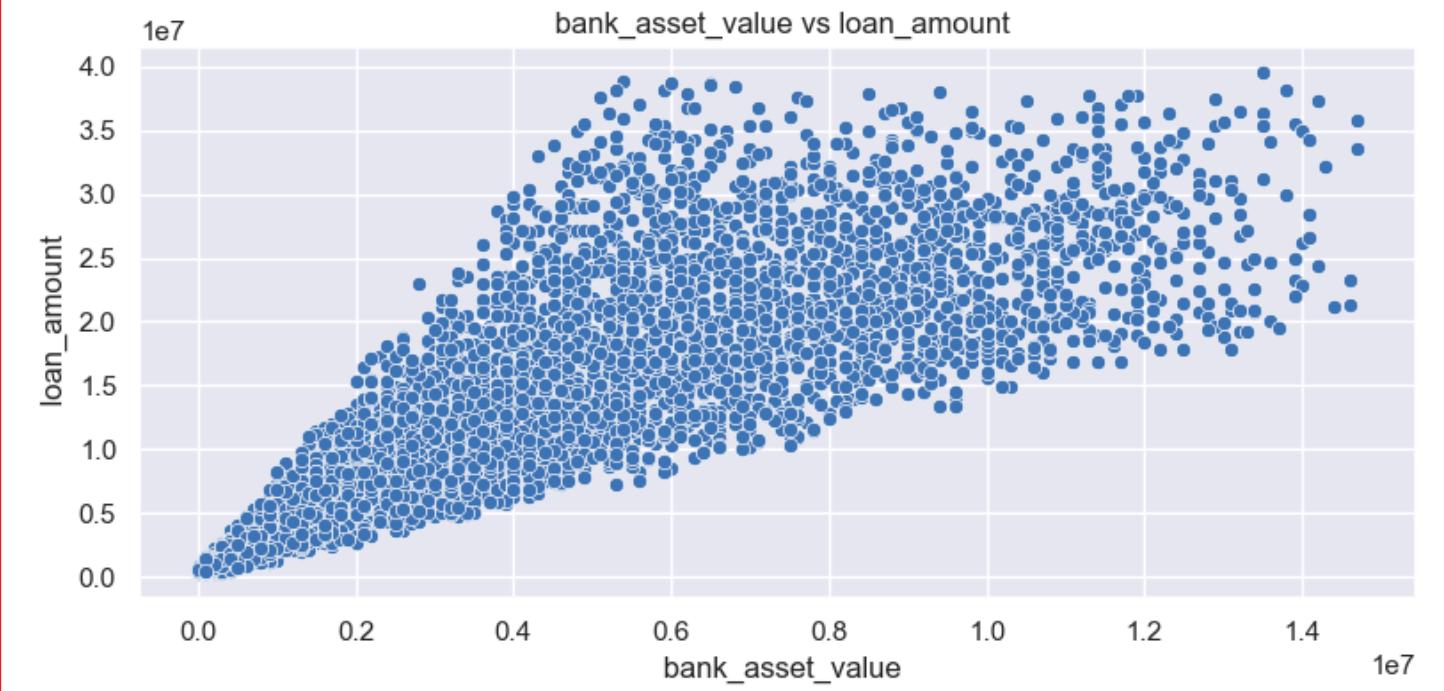
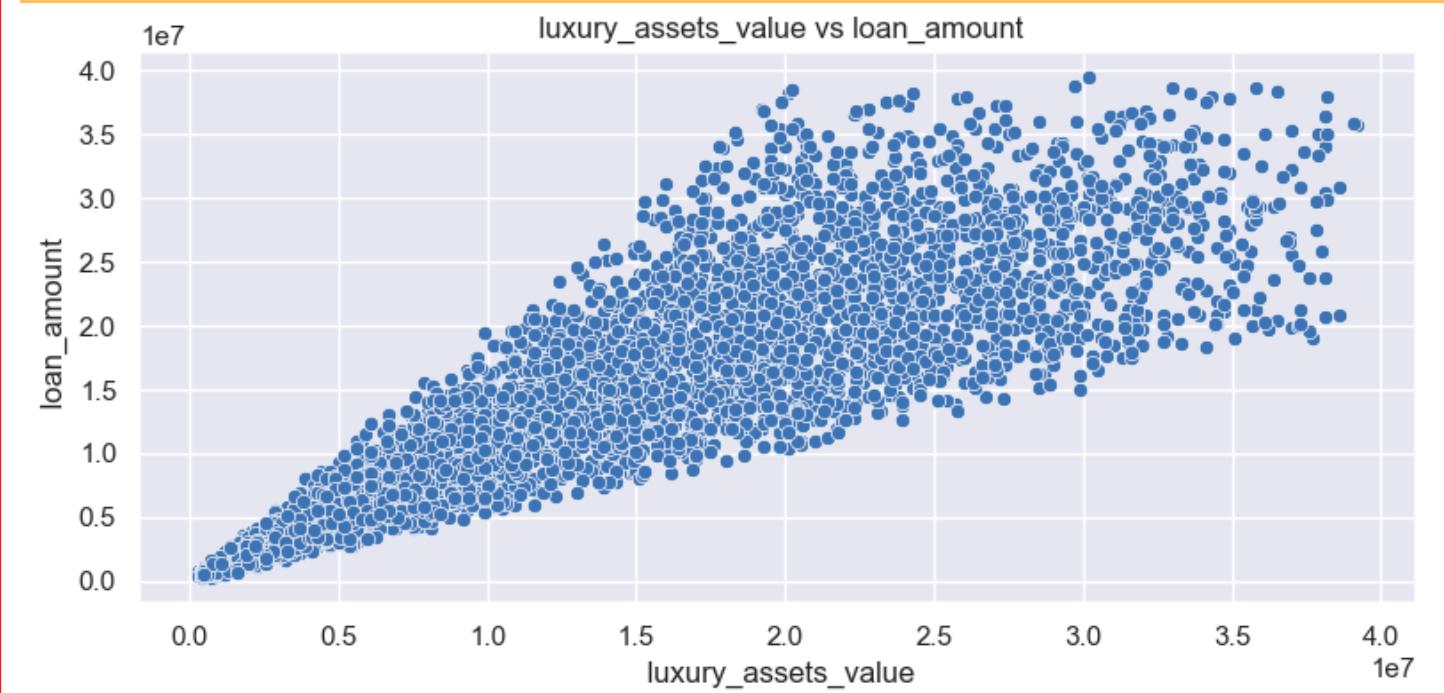
UNIVARIATE LINEAR REGRESSION MODEL

(Loan Amount)



Mean squared error(MSE): 11665958956266.385

Explained Variance on the test dataset is: 0.8589414839594748



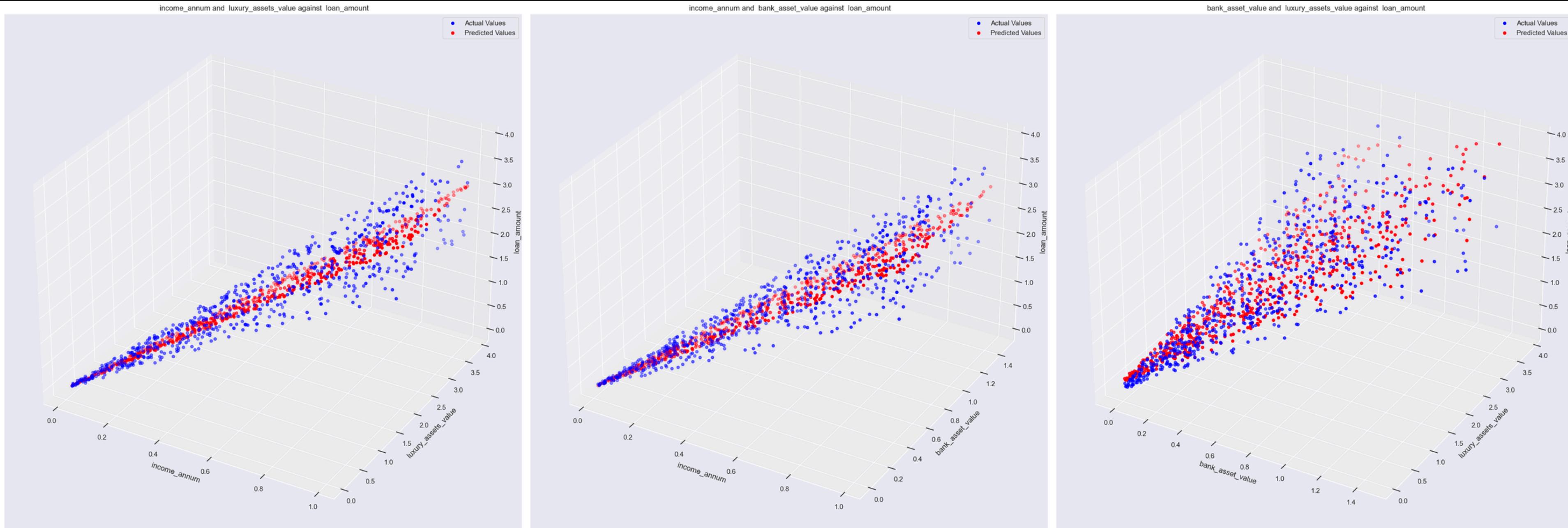
Most Correlated to Loan Amount

Previously...

To use in Linear Regression Model Later

MULTIVARIATE LINEAR REGRESSION MODEL

(Loan Amount)



income_annum and luxury_assets_value against loan_amount:
Mean Squared Error (MSE): 11668608843159.533
Explained Variance (R²): 0.8589094429490252

income_annum and bank_asset_value against loan_amount:
Mean Squared Error (MSE): 11659580139332.31
Explained Variance (R²): 0.8590186131911297

bank_asset_value and luxury_assets_value against loan_amount:
Mean Squared Error (MSE): 18480953412531.926
Explained Variance (R²): 0.7765382277480466

Univariate Linear Regression Model

Mean squared error(MSE): 11665958956266.385

Explained Variance on the test dataset is: 0.8589414839594748

Multivariate Linear Regression Model

income_annum and luxury_assets_value against loan_amount:

Mean Squared Error (MSE): 11668608843159.533

Explained Variance (R^2): 0.8589094429490252

income_annum and bank_asset_value against loan_amount:

Mean Squared Error (MSE): 11659580139332.31

Explained Variance (R^2): 0.8590186131911297

bank_asset_value and luxury_assets_value against loan_amount:

Mean Squared Error (MSE): 18480953412531.926

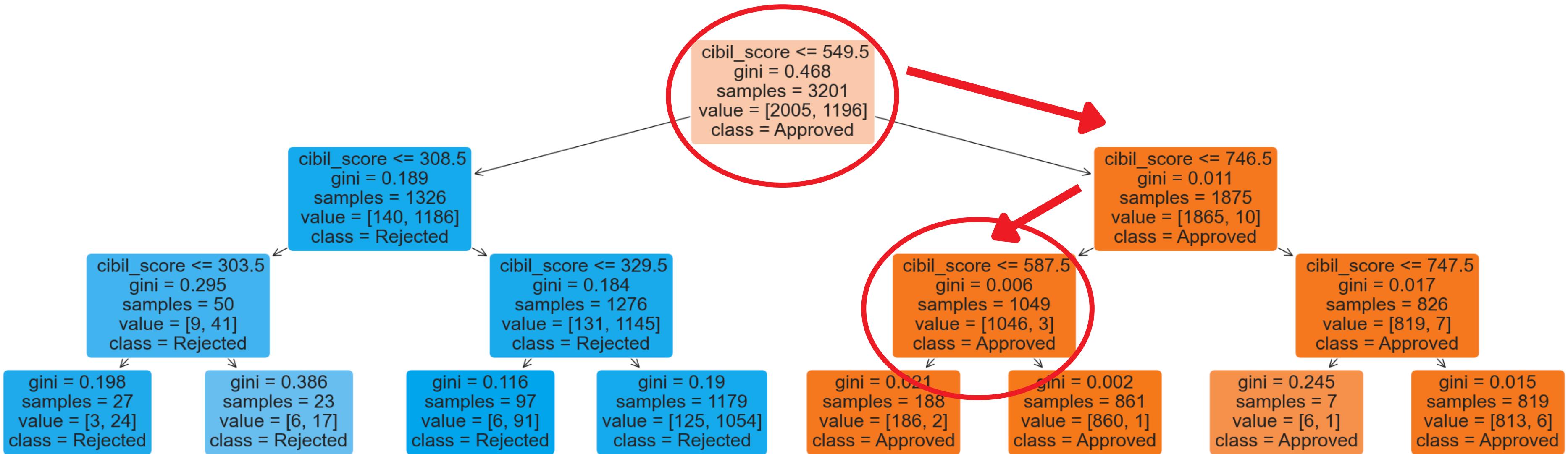
Explained Variance (R^2): 0.7765382277480466

TAKE IN APPLICANT'S CREDIT SCORE

(Loan Approval Rate)

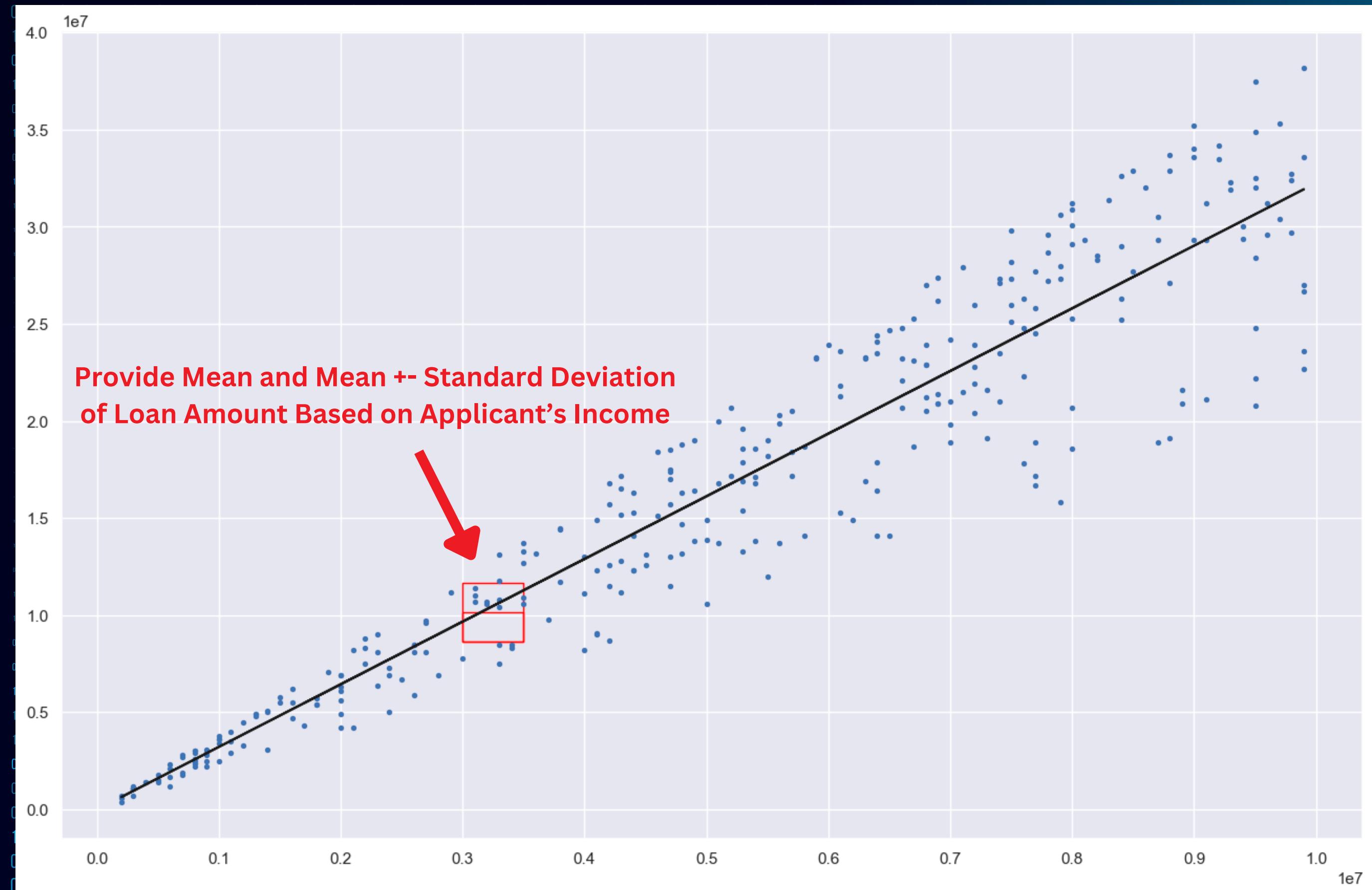
- Generate another linear regression line, taking credit score into account
- Range of credit score used -> from credit score of leaves of univariate decision tree

Applicant's Credit Score: 580



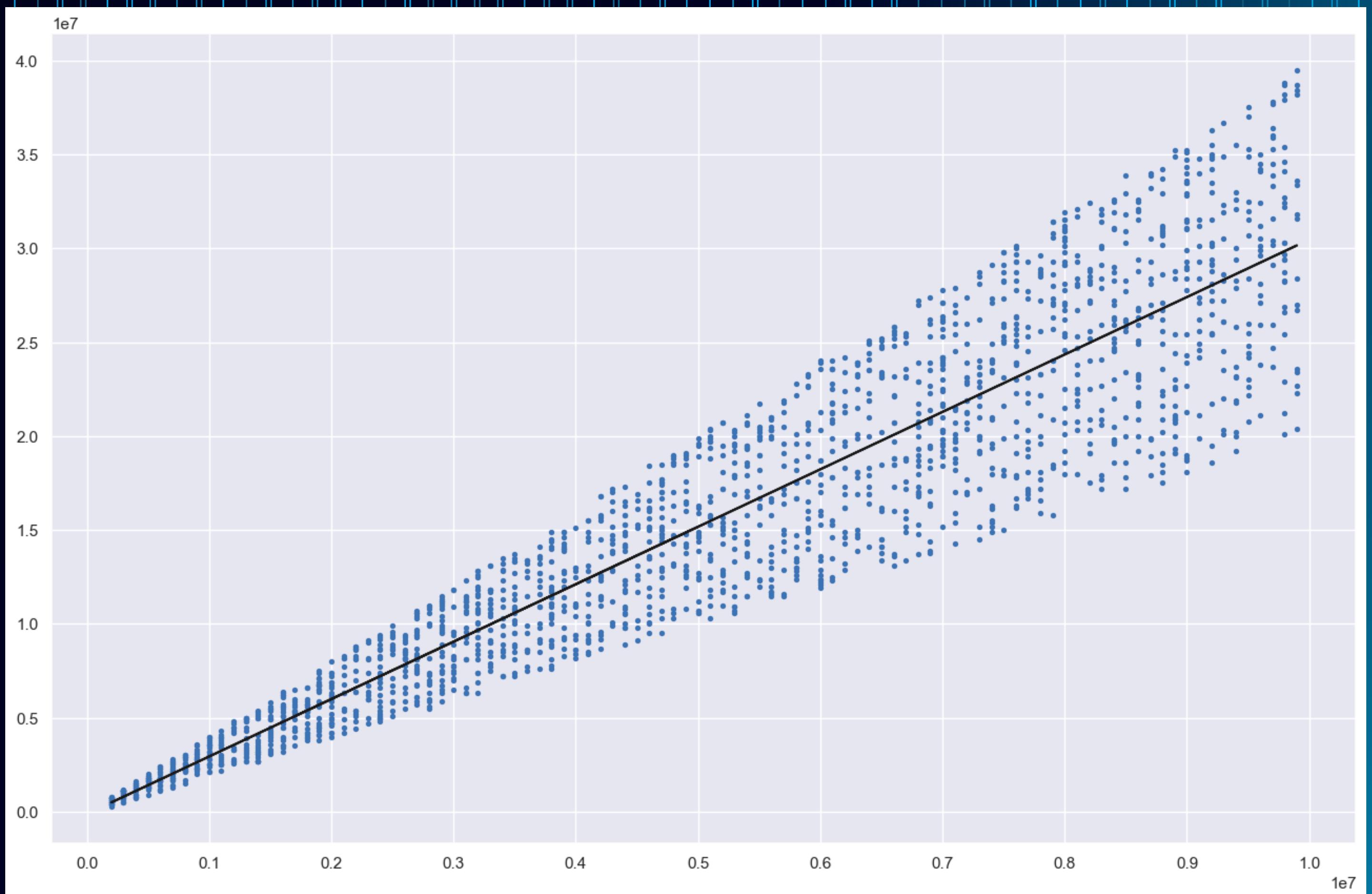
Credit Score Range: 549.5 - 587.5

LINEAR REGRESSION MODEL (AFTER TAKING CREDIT SCORE INTO ACCOUNT)



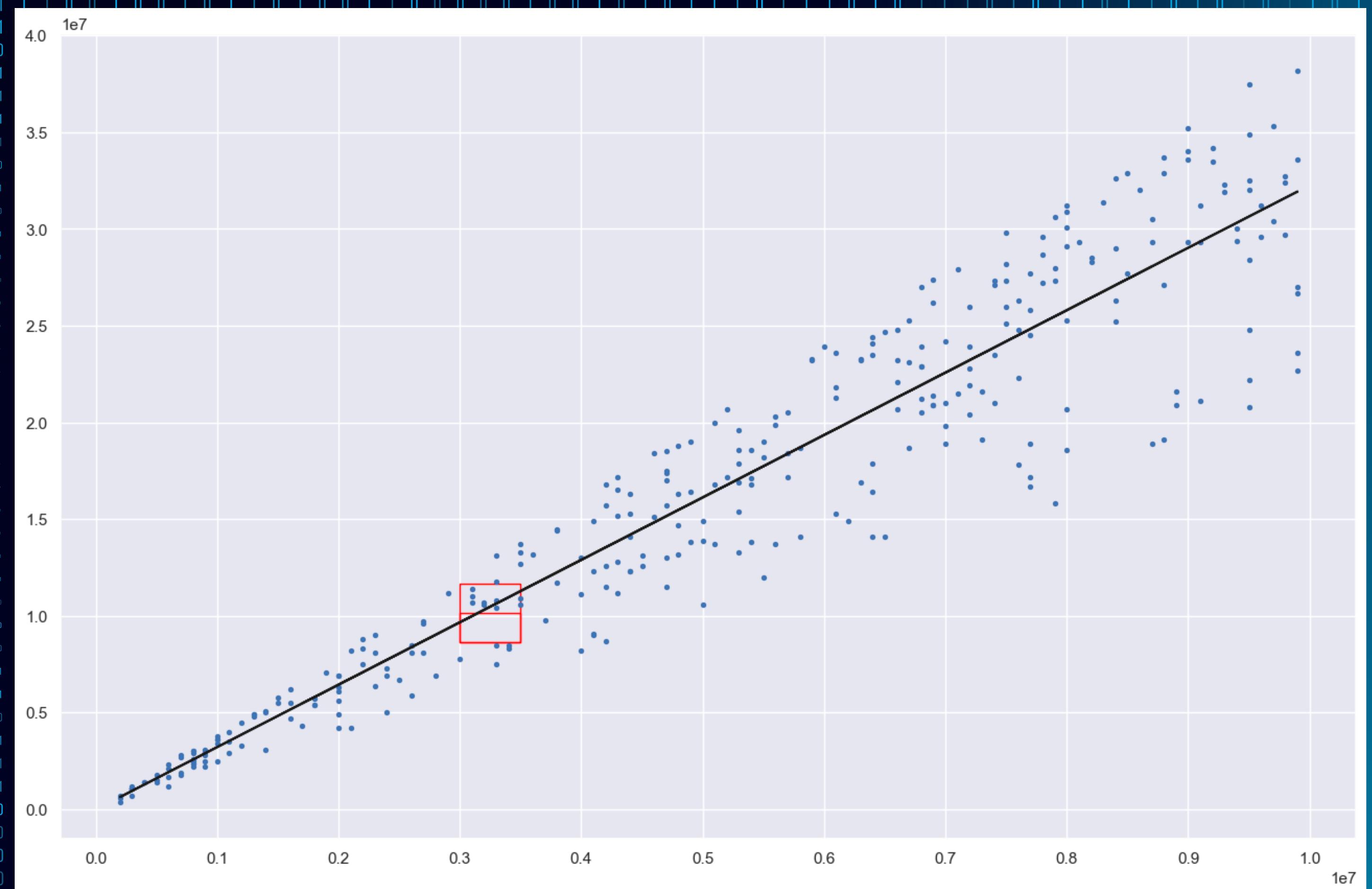
Applicant's Credit Score: 580

Applicant's Annual Income: \$3,000,000



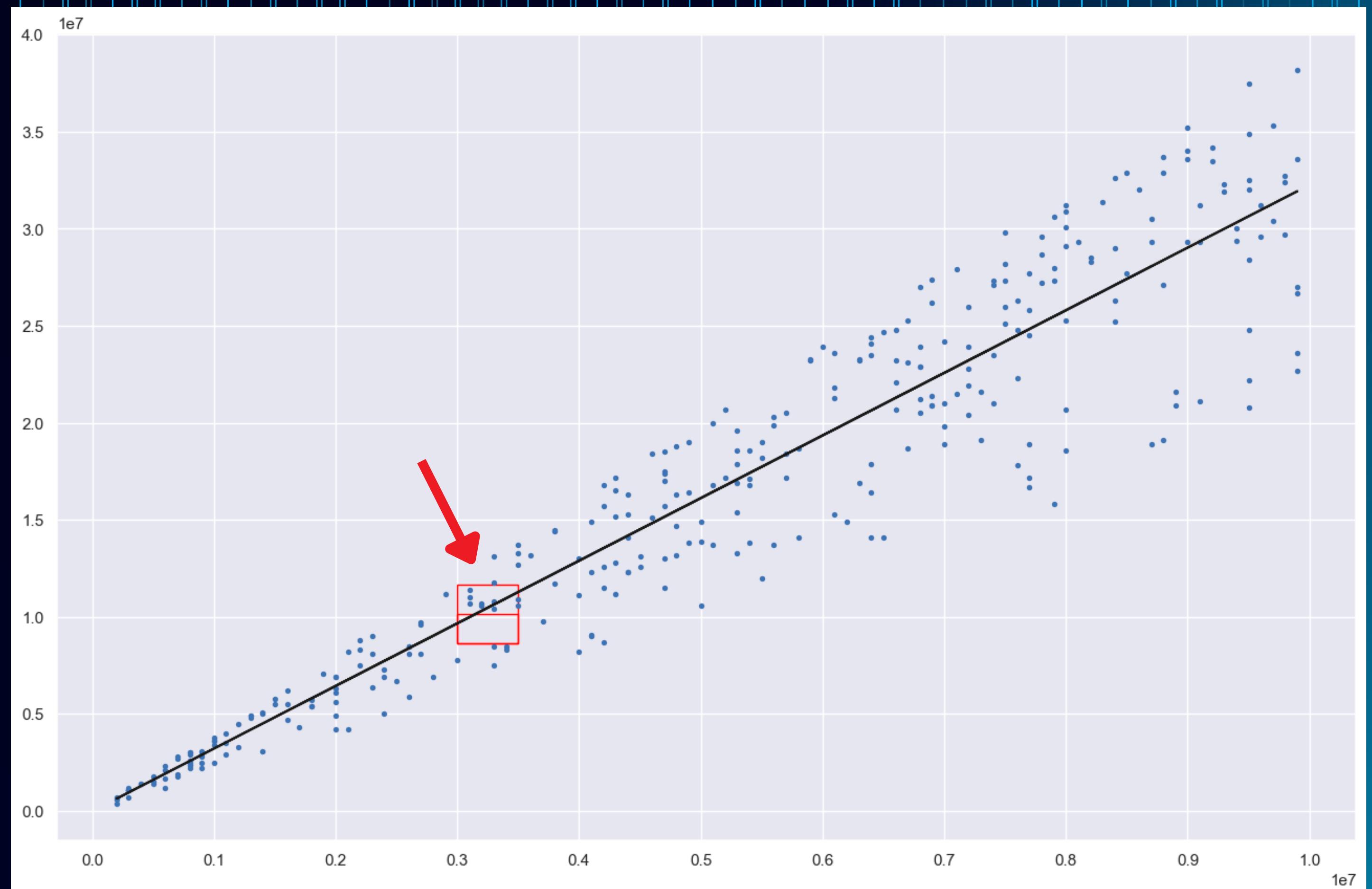
Applicant's Credit Score: 580

Applicant's Annual Income: \$3,000,000

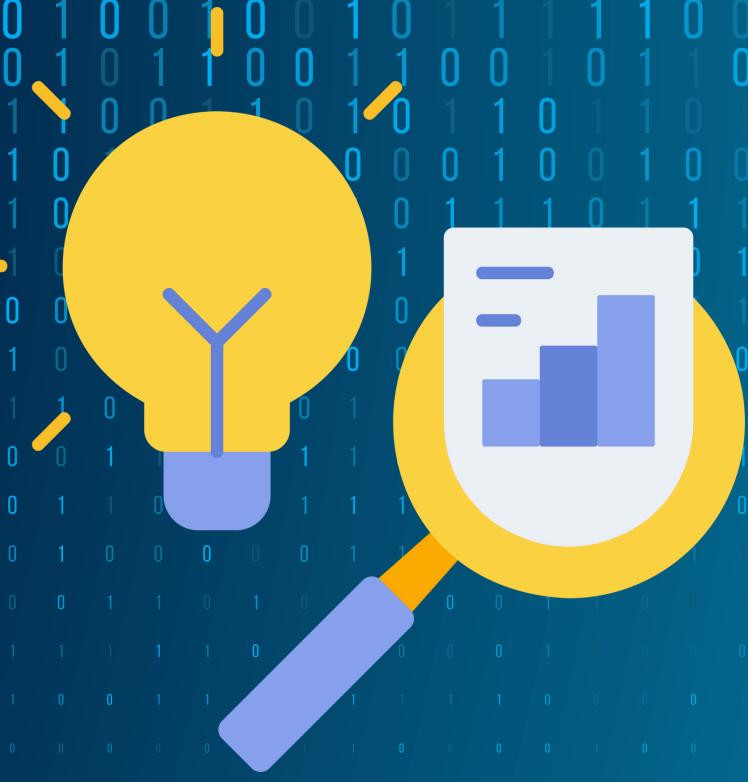


Applicant's Credit Score: 580

Applicant's Annual Income: \$3,000,000



OUTCOME AND INSIGHTS



OUTCOME

According to the information you provided as followed:

Credit Score = 580

Annual Income = \$ 3000000

We provide the following analysis:

Your loan approval prediction according to Decision Tree Model is: Approved

The probability of your loan being Approved is 75.56 % according to Logistic Regression Model

Your recommended loan amount by Linear Regression Model is: \$ 9055582.09

After filtering by your credit score range: \$ 9671891.07

Most of the loans that are approved in your income range as well as your credit score range lie in the range: \$ 8630428.55 - \$ 11665223.62

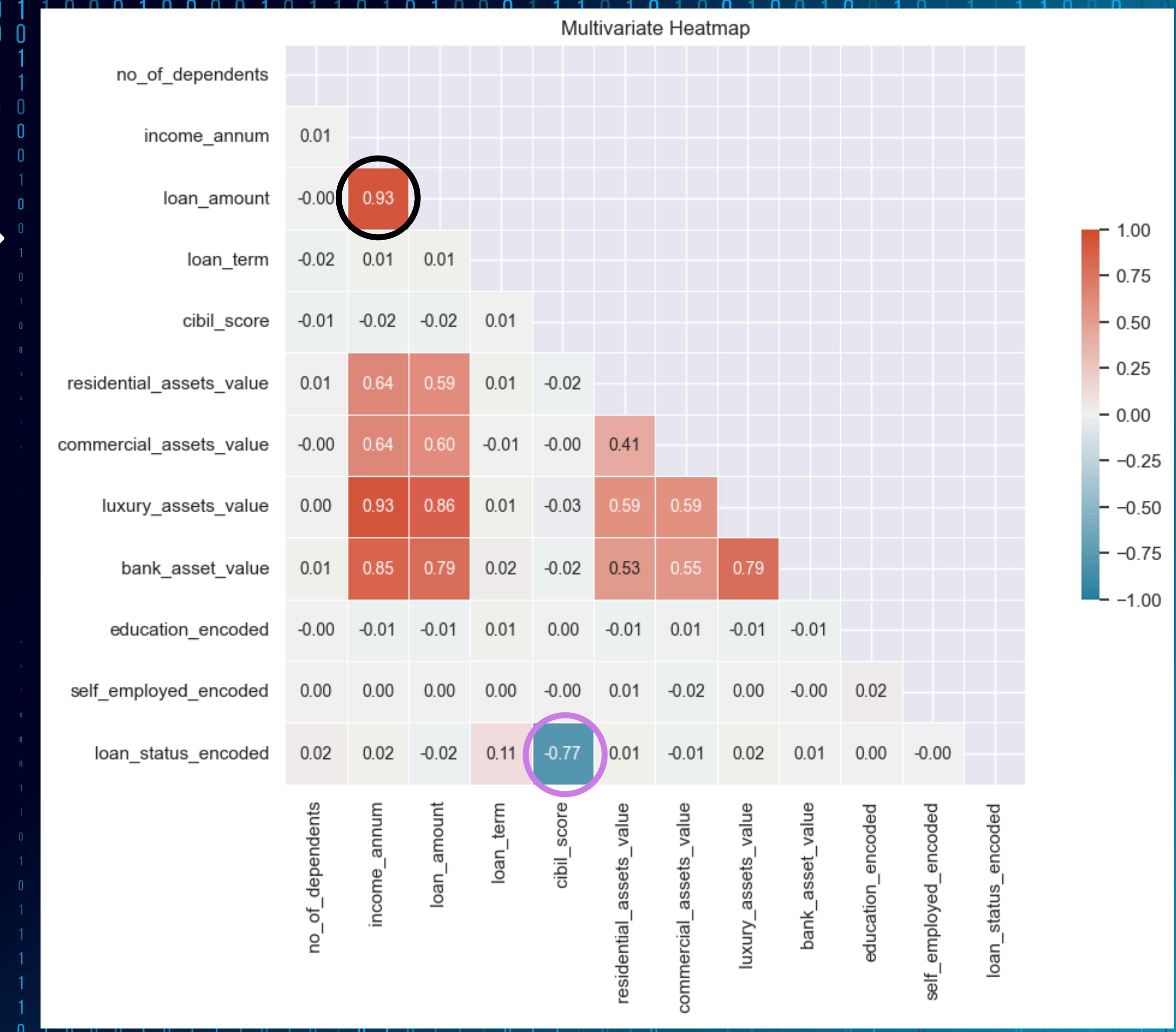
Mean + Standard Deviation \$ 11665223.62

The average(mean) of the loan you can take is \$ 10147826.09

Mean - Standard Deviation \$ 8630428.55

Insights

- Strong correlation of credit score and annual income significantly influences loan approval rate and loan amount
- Using multivariate models may not always increase accuracy of data
- Finding the most influential variables that affect the targeted results is crucial



CONCLUSION

Our project clearly informs applicants of their loan approval rate and loan amount they can receive with their current conditions





THANK YOU