

Loan Approval

PREDICTION APP

ALEAKHYA PANDETI - 2022A7PS0016U

KHYAATI RUNGTA - 2023A7PS0316U



TABLE OF CONTENTS

- Purpose
- Introduction to dataset
- EDA
- Pre-Processing the data
- Prediction models
- Streamlit application



PURPOSE

A loan approval app can be incredibly useful for the public in several ways:

1.Improved Efficiency: By using predictive models, loan approval apps can quickly assess the likelihood of approval based on various factors. This can streamline the application process, reducing the time it takes for applicants to receive a decision.



2. **Personalization:** Some apps use data analytics to offer personalized loan options tailored to the user's financial situation and credit history.

3. **Better Risk Management:** By analyzing historical data and identifying patterns, loan approval apps can better assess the risk associated with each application. This can lead to more accurate decisions and ultimately reduce the likelihood of default.



INTRODUCTION TO DATASET

The loan approval dataset is from Kaggle. It is a collection of financial records and its associated information used to determine the eligibility of individuals or organizations for obtaining loans from a lending institution.

loan_approval_dataset.csv (384.34 kB)							
Detail	Compact	Column					
	# no_of_dep... ⌵	Δ education ⌵	✓ self_emplo... ⌵	# income_an... ⌵	# loan_amou... ⌵	# loan_term ⌵	# cibil_score ⌵
	2	Graduate	No	9600000	29900000	12	778
	0	Not Graduate	Yes	4100000	12200000	8	417
	3	Graduate	No	9100000	29700000	20	506
	3	Graduate	No	8200000	30700000	8	467
	5	Not Graduate	Yes	9800000	24200000	20	382
	0	Graduate	Yes	4800000	13500000	10	319
	5	Graduate	No	8700000	33000000	4	678
	2	Graduate	Yes	5700000	15000000	20	382
	0	Graduate	Yes	800000	2200000	20	782
	5	Not Graduate	No	1100000	4300000	10	388
	4	Graduate	Yes	2900000	11200000	2	547
	2	Not Graduate	Yes	6700000	22700000	18	538

Source: <https://www.kaggle.com/datasets/architsharma01/loan-approval-prediction-dataset/data>

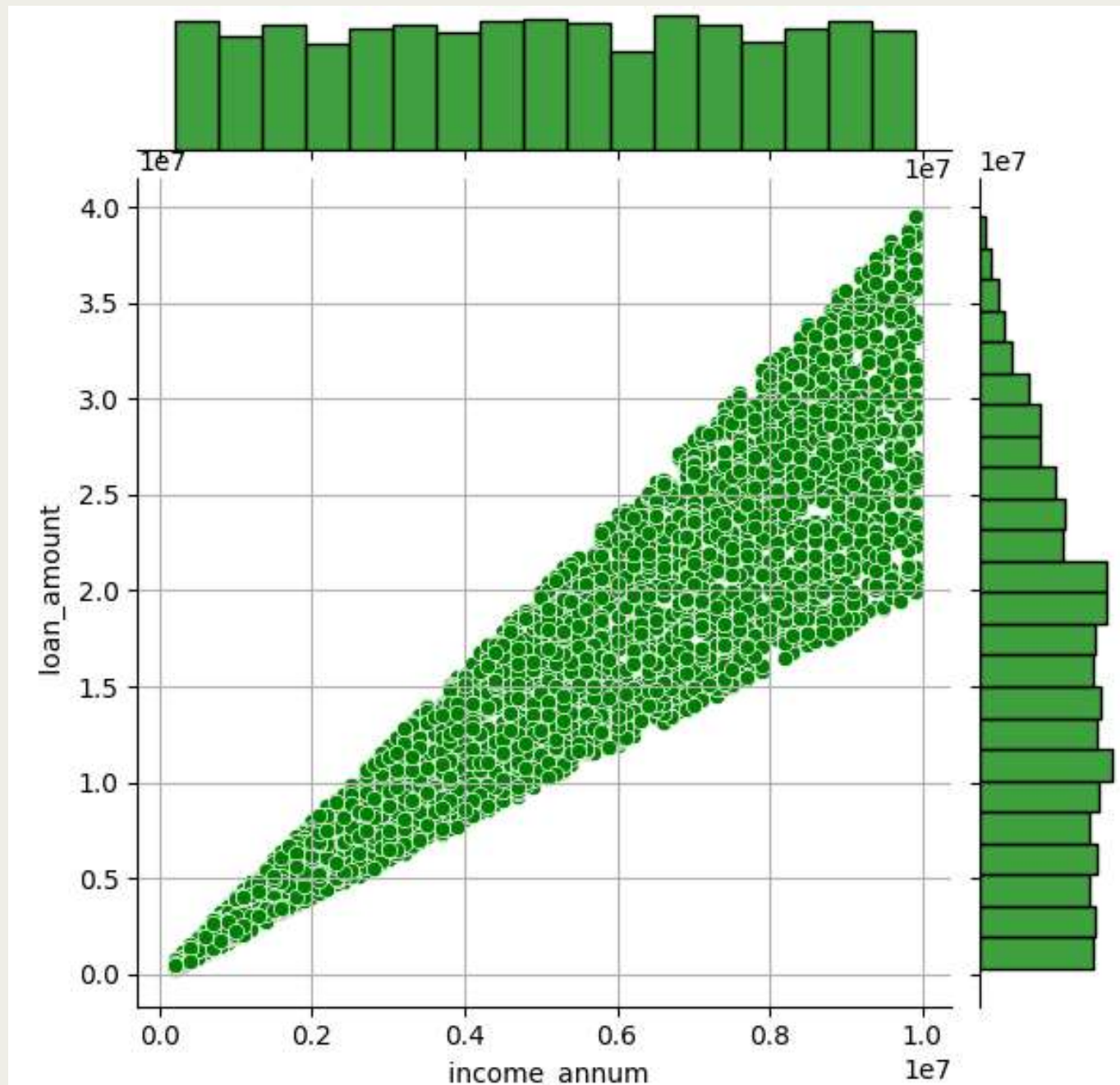


The dataset comprises 4000 rows of data and 13 columns representing the factors affecting the loan status. The factors are as follows :

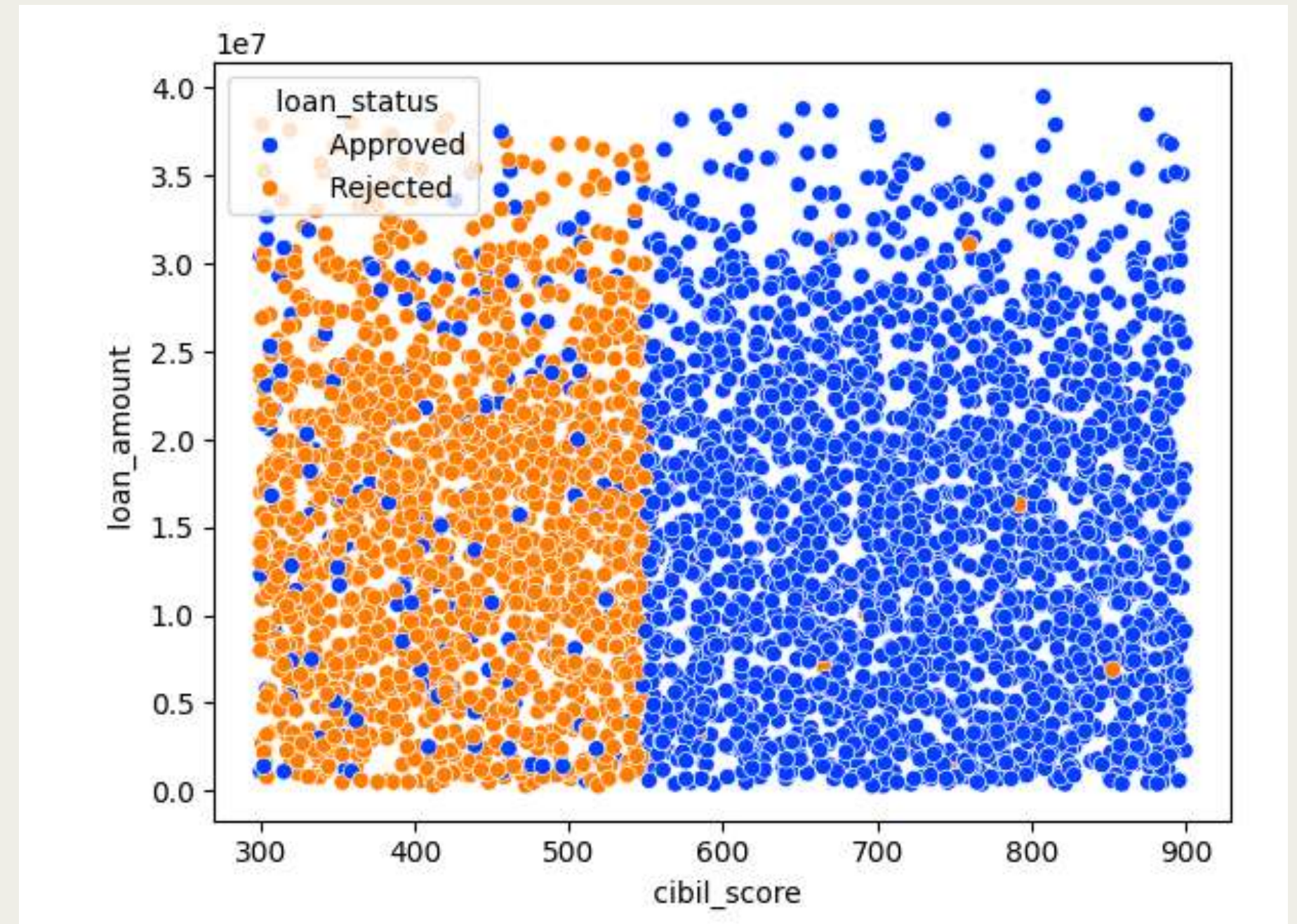
- no of dependents
- education
- self-employed
- income -annum
- loan-amount
- loan - term
- cibil score
- residential asset value
- commercial asset value
- luxury asset value
- bank asset value
- loan-status (Target variable)



EDA



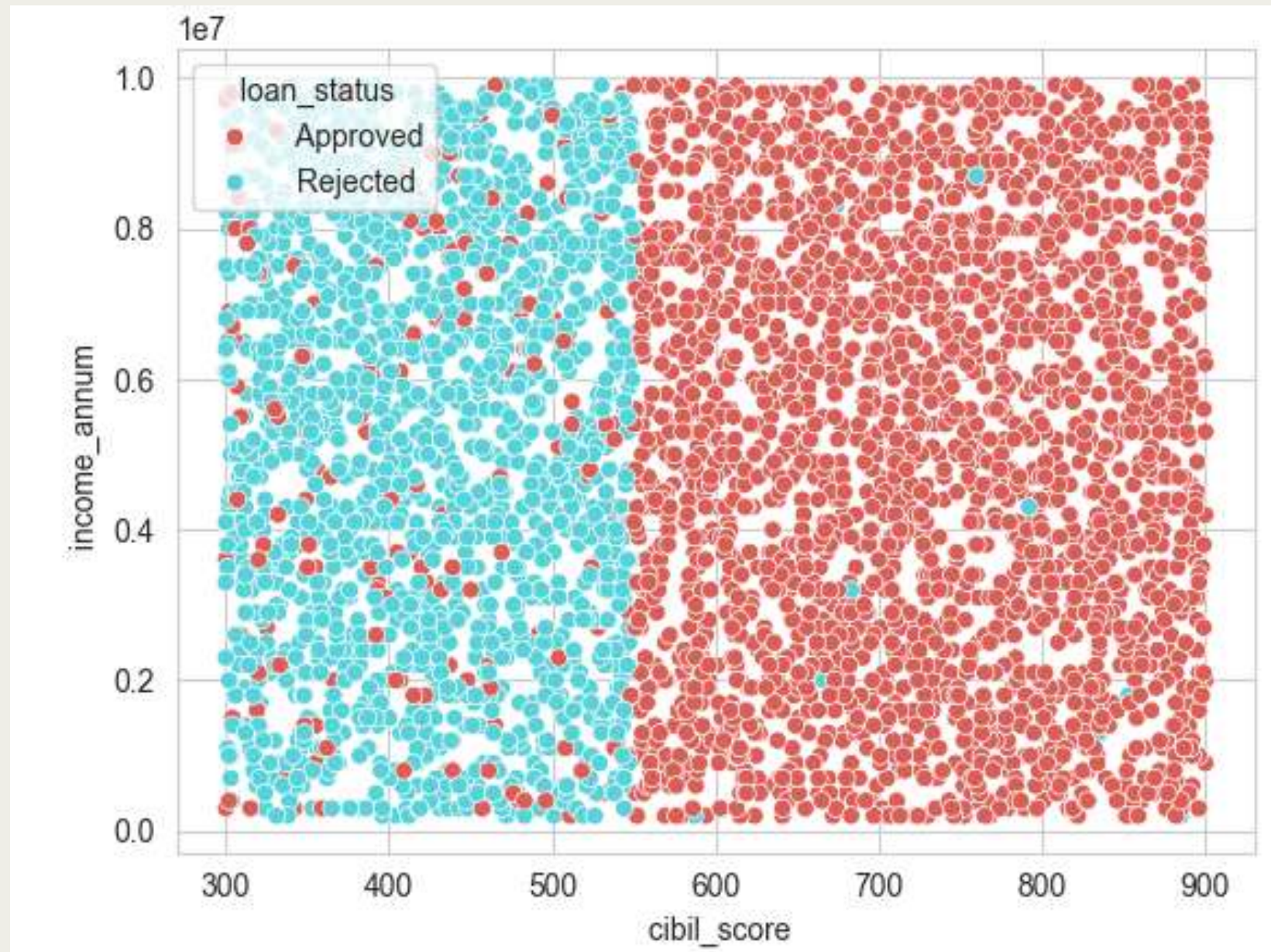
Loan amount is directly proportional to Income annum



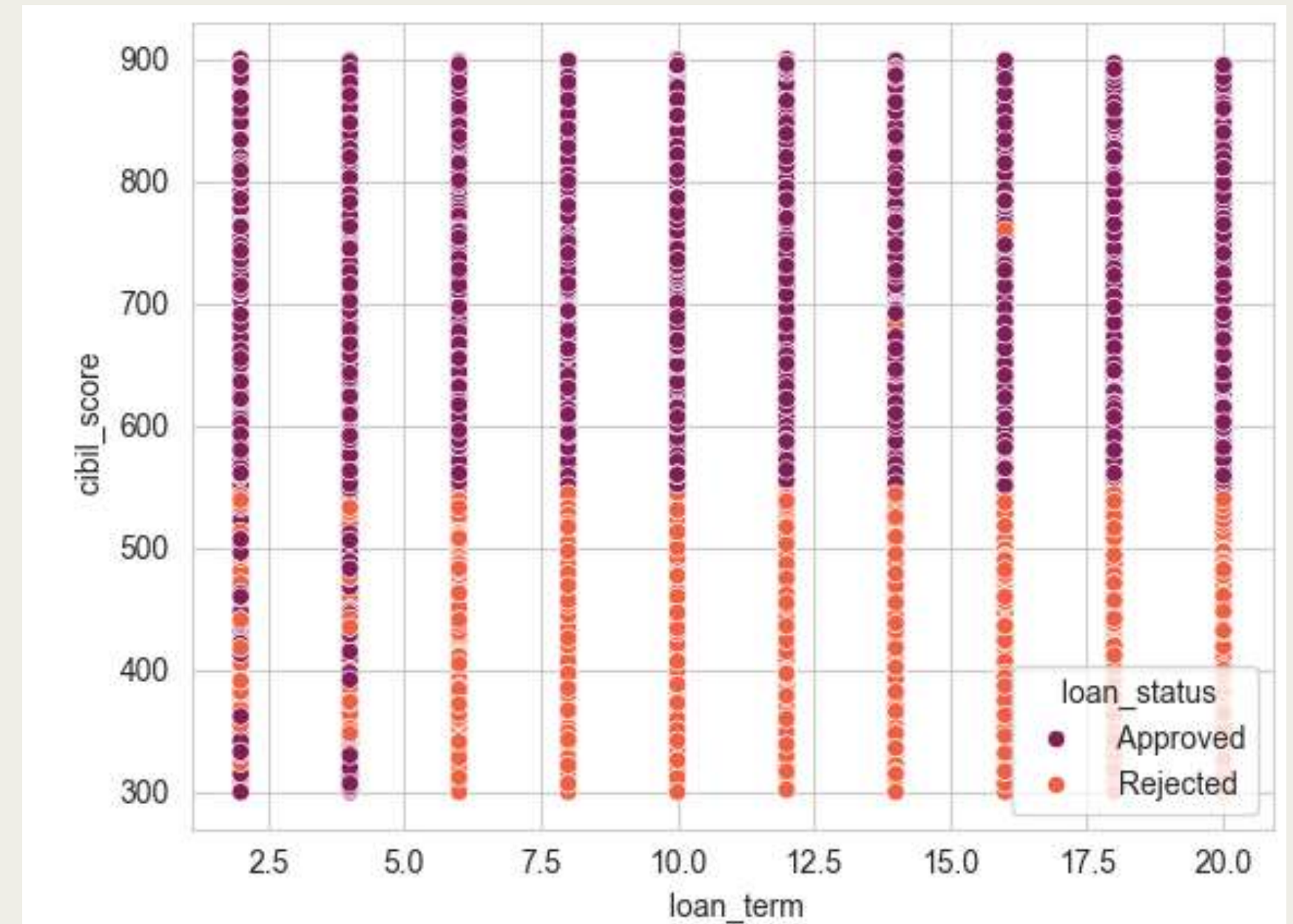
From approximately 550 cibil score, most of the loan amount begins to get approved



EDA



Approximately 550 of cibil score onwards
approved loan applications increase irrespective
of income



Loan requests with (low cibil score and
less loan terms) and high cibil score
are approved



PRE - PROCESSING

```
[9]: hd_data['education'].unique()

[9]: array([' Graduate', ' Not Graduate'], dtype=object)

[10]: from sklearn import preprocessing
      label_encoder = preprocessing.LabelEncoder()

      hd_data['education'] = label_encoder.fit_transform(hd_data['education'])

      hd_data['education'].unique()

[10]: array([0, 1])

[11]: hd_data['self_employed'].unique()

[11]: array([' No', ' Yes'], dtype=object)

[12]: hd_data['self_employed'] = label_encoder.fit_transform(hd_data['self_employed'])
      hd_data['self_employed'].unique()

[12]: array([0, 1])

[13]: hd_data['loan_status'].unique()

[13]: array([' Approved', ' Rejected'], dtype=object)

[14]: hd_data['loan_status'] = label_encoder.fit_transform(hd_data['loan_status'])
      hd_data['loan_status'].unique()

[14]: array([0, 1])
```

Label Encoding is a technique used in machine learning and data analysis to convert categorical variables into numerical format. It is particularly useful when working with algorithms .



MODEL TRAINING

Models used to obtain highest accuracy :

Random Forest : It is a widely-used machine learning algorithm which combines the output of multiple decision trees to reach a single result

Decision Tree: Decision trees are used to solve classification problems and categorize objects depending on their learning features.



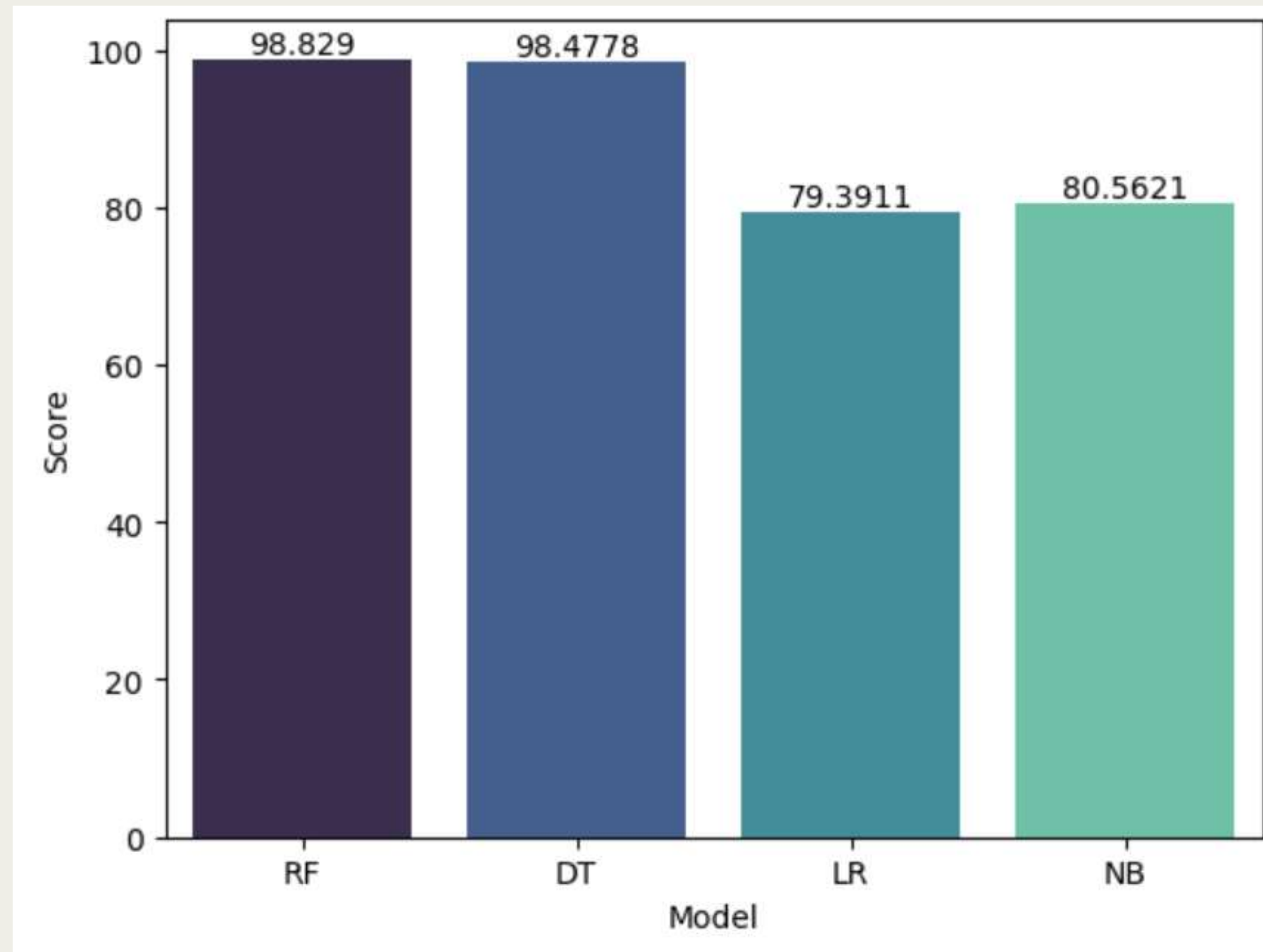
MODEL TRAINING

Logistic Regression: Logistic regression is a statistical method used for binary classification, which means it's used to predict the probability of a binary outcome based on one or more predictor variables.

Naive Bayes: Naive Bayes is a simple but powerful probabilistic classifier based on applying Bayes' theorem with strong (naive) independence assumptions between the features.



MODEL TRAINING



Comparison plot for the accuracy of the models chosen show that random forest is the best performing algorithm to use for model training



STREAMLIT APP



Loan Approval Prediction App

This app predicts whether the loan will be approved or not...

Data obtained from kaggle datasets(<https://www.kaggle.com/datasets/architsharma01/loan-approval-prediction-dataset/data>)



Thank you!

Github link: <https://github.com/AleakhyaCS/ACM-W-Project>

