

---

# Predicting Credit Card Approval

— Abhinav Saurabh MT20127 —  
Rahul Meena MT20117

---

# Abstract

In the banking system, there is a major task to determine if the person is eligible for the credit card. Profit to the credit card providers largely depend on if person is paying back.

By predicting correct credit card customers ,the banks can maximize the profits.

We here automated this process by applying machine learning techniques. Here we need to make sure to target the right customers.

# Introduction

- The correct assessment for credit card approval is very important for banks and organisations who lend credit card to the people. The recent years has seen huge growth in credit cards and loans.
- The exact judgement of person to be approved for credit cards allows the organisations to minimize losses and same time make suitable credit arrangements as per requirement.
- Due to huge growth in number of applicants there is need for more sophisticated method to automate the process and speed it up.

# Importance of Project

- Credit card approval can be very helpful for organisations that lend credit cards and due to increase in huge number of applicant there is need to automate the task and classify the applicants into if they are eligible for credit card or not.
- This helps to avoid organisation losses by avoiding potential defaulters.
- Here we are not just looking into bank balance but into their personal attributes like gender, married, age, Occupation etc. We account for these personal attributes to evaluate if given is applicant is good customer.
- This gives benefit by cutting down costs on credit analysis and faster credit decisions.

# Literature Survey

- In past various models have been proposed to determine the and evaluate the credit scoring criteria. Techniques can be further classified into parametric and non-parametric models. The most well liked parametric model were logistic regression and Linear discriminant analysis.[1] LDA has been criticized because of categorical nature of data.
- Logistic regression overcomes these problems later turn out to be common credit scoring tool for credit lending organisations. The prediction is taken into consideration after applying sigmoid function into it.[2]

# Literature Survey

- SVM can be combined with backpropagation neural network to get better accuracy. The results displayed SVM's accuracy comparable to that of backpropagation neural network[3]
- Further there have been improvement in various other hybrid data mining techniques to get better results at these kinds of problems. Various ensemble methods have also been used to get improved accuracy by aggregating scores.[4]

# Literature Survey

- Credit card approval is also being done through genetic programming, this paper examines the usage of strong typed genetic programming for automated credit approval. Eight different genetic programming approaches were applied and compared.[5]
- In another paper implementation based on logistic regression has been compared with XGBoost algorithm. It is found that XGBoost algorithm has significantly higher model discrimination and model stability than that of logistic regression.[6]

# Data Description

## Data Source:

The Data Set we have obtained from UC Machine Learning data repository submitted by quinlan'@' cs.su.oz.au. The data contains various variables such as Gender, Age, Education, Employed....

**Data Description:** The data have been encoded to some special text to protect the confidentiality of the persons. But the encoding does not affect our purpose. In the dataset there are 690 instances out of them 70% are used for training and rest 30% used for validation. There are continuous values as well as categorical values. There are 16 columns first 15 are credit application attributes and the last one is the approved columns which contain boolean value either the credit card application is approved or rejected.



# Data Analysis and Preprocessing

The data set contains 690 instances of 16 variable

Which are Gender, Age, Debt, Married, Bank Customer, Education, Ethnicity, Years Employed, Prior Default, Employed, Credit Score, Driving License, Citizenship, Zip Code, Income, **Approved**

Approved is the label of the instance consisting + or - denoting whether the credit card request approved or not.

# Anamolies and Outliers

There were few instances in which some attribute contains abnormal values.

Values which varied a lot.

**These are not expected values**

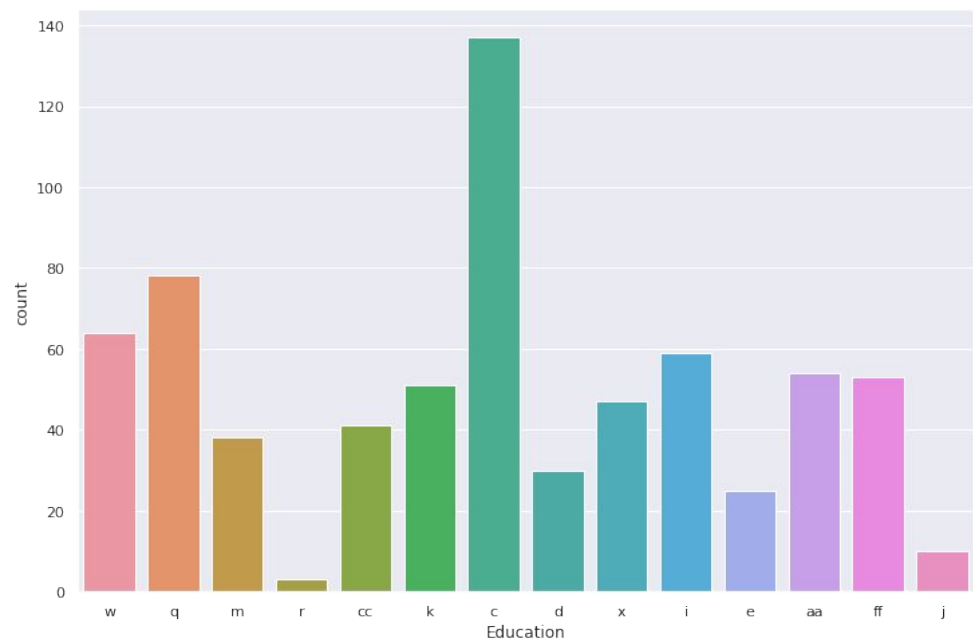
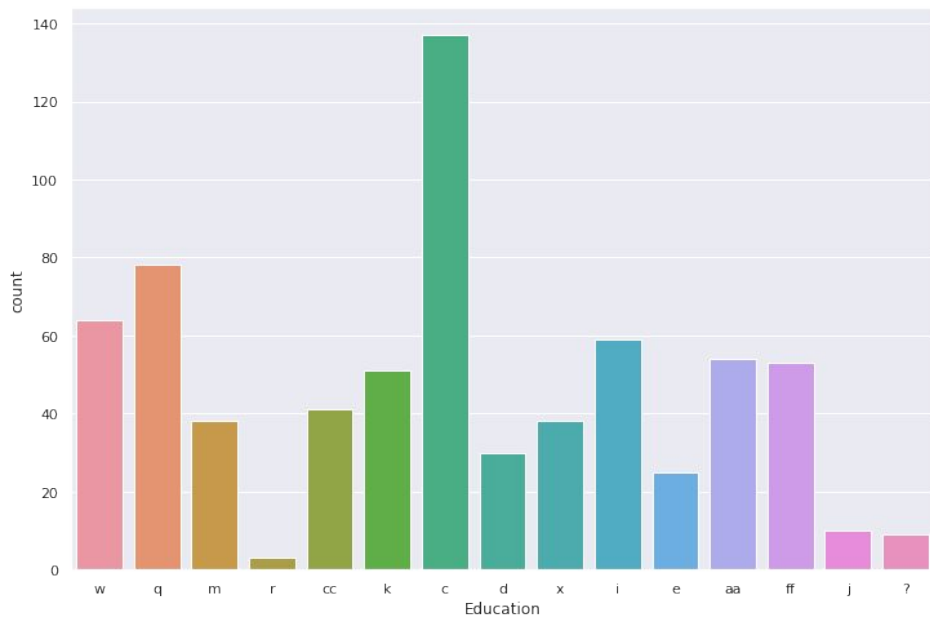
So we replaced these values with the average value of the column.

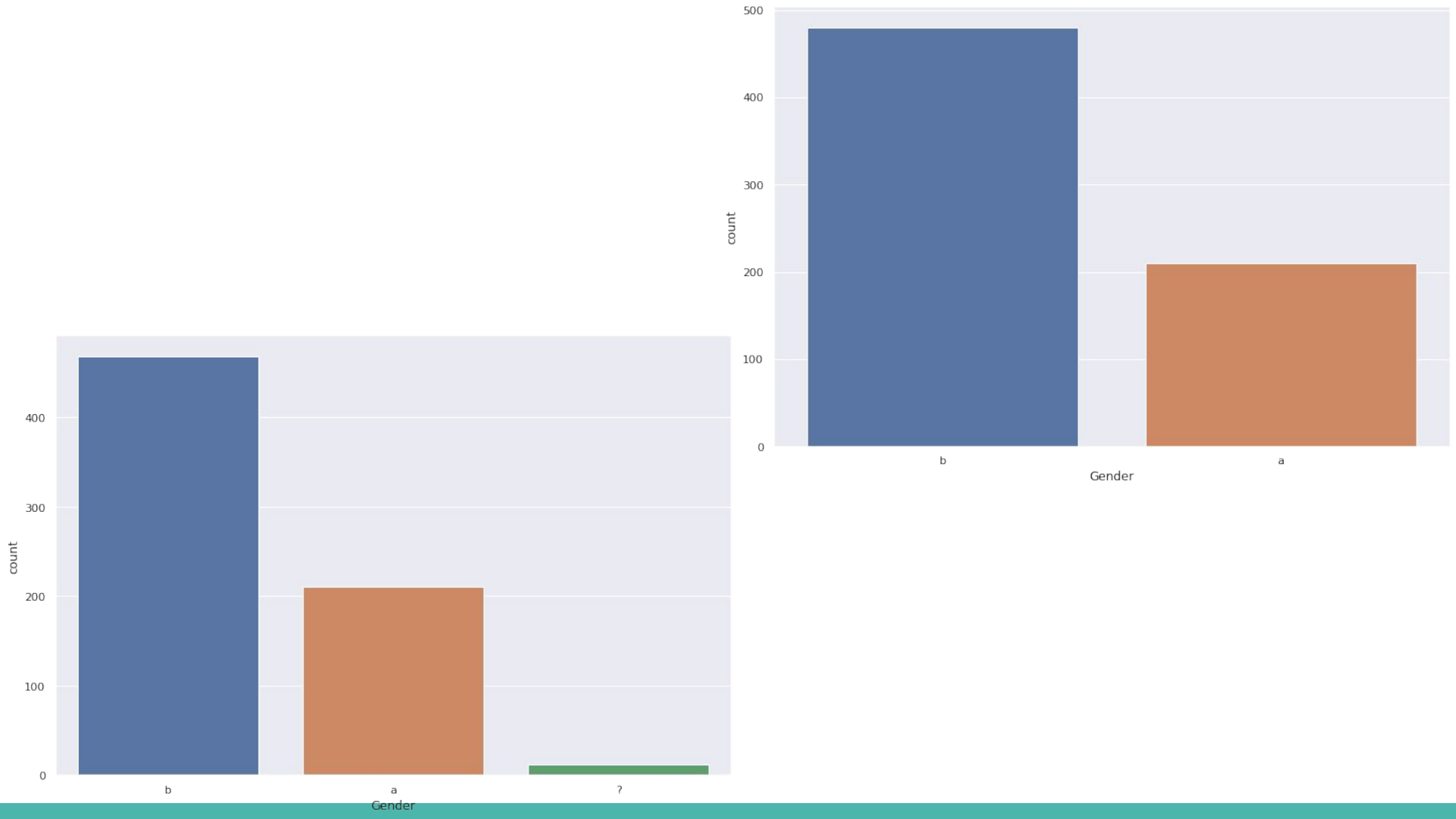
# Missing Data

Some data were missing and are represented by ?

The categorical type data is replaced with the most occurred data

While the numerical data replaced with the mean

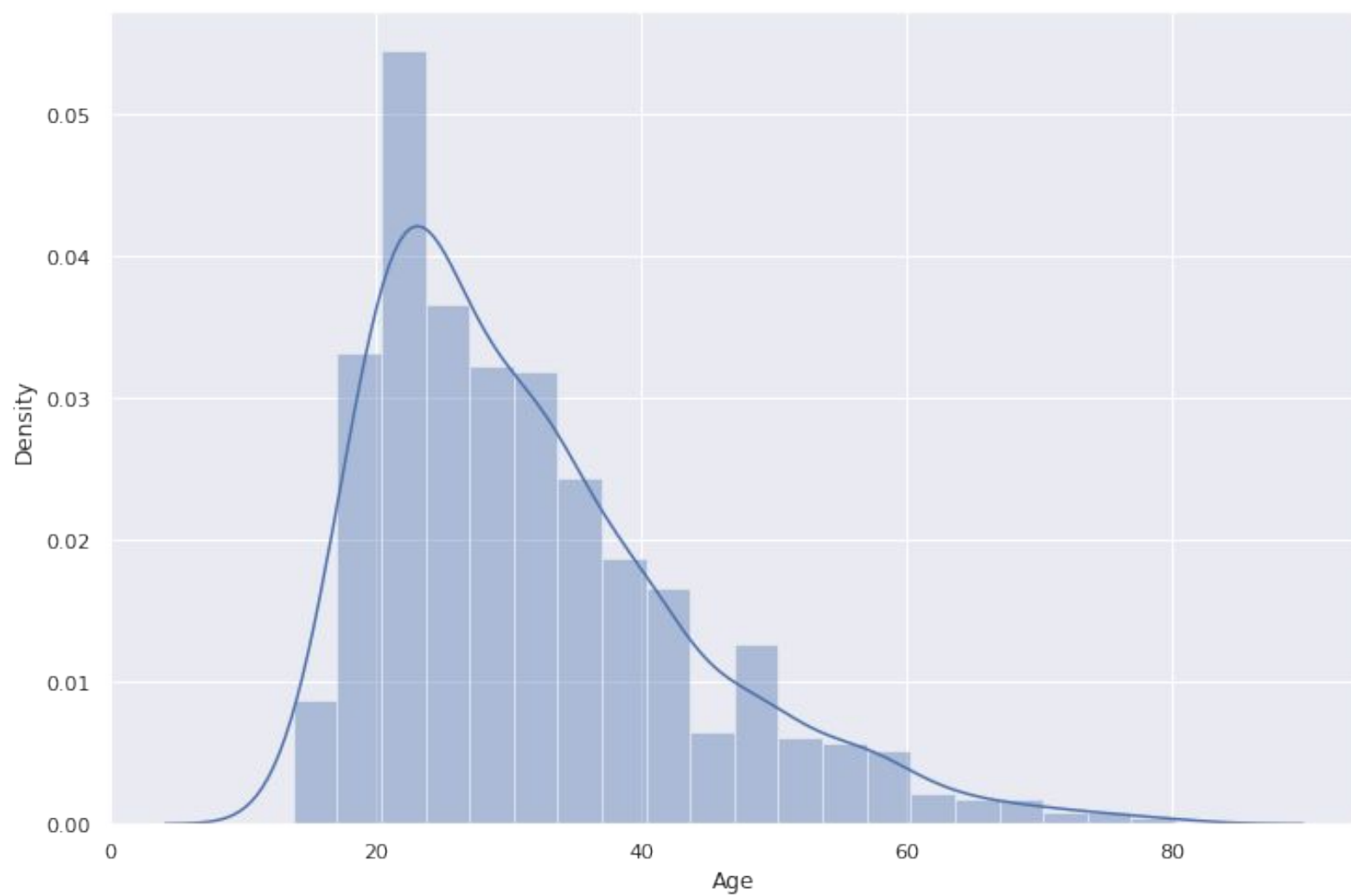


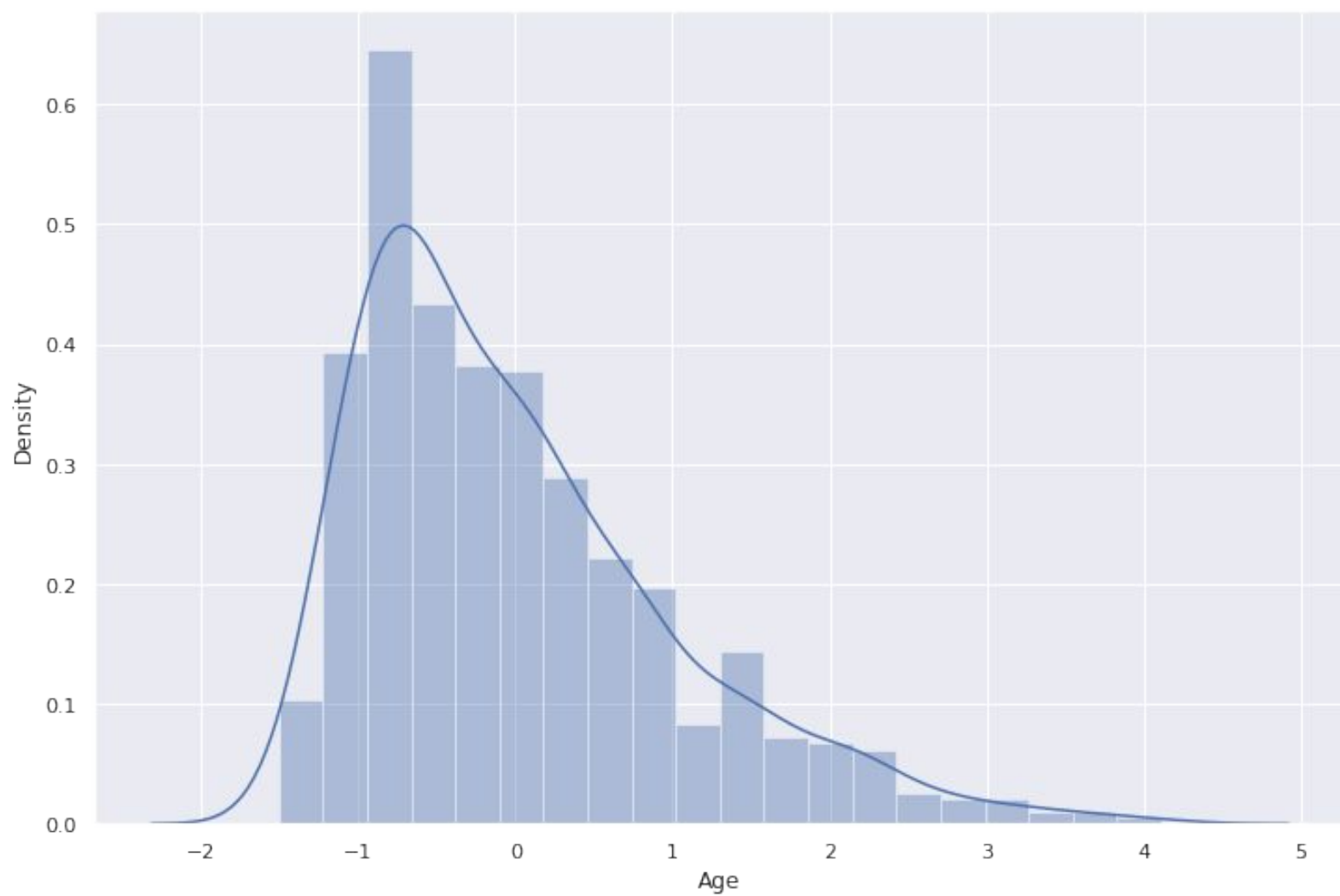


# Standardization of Data

For the health of our model, we need to Standardization the data

By subtracting each value by its mean and dividing by SD

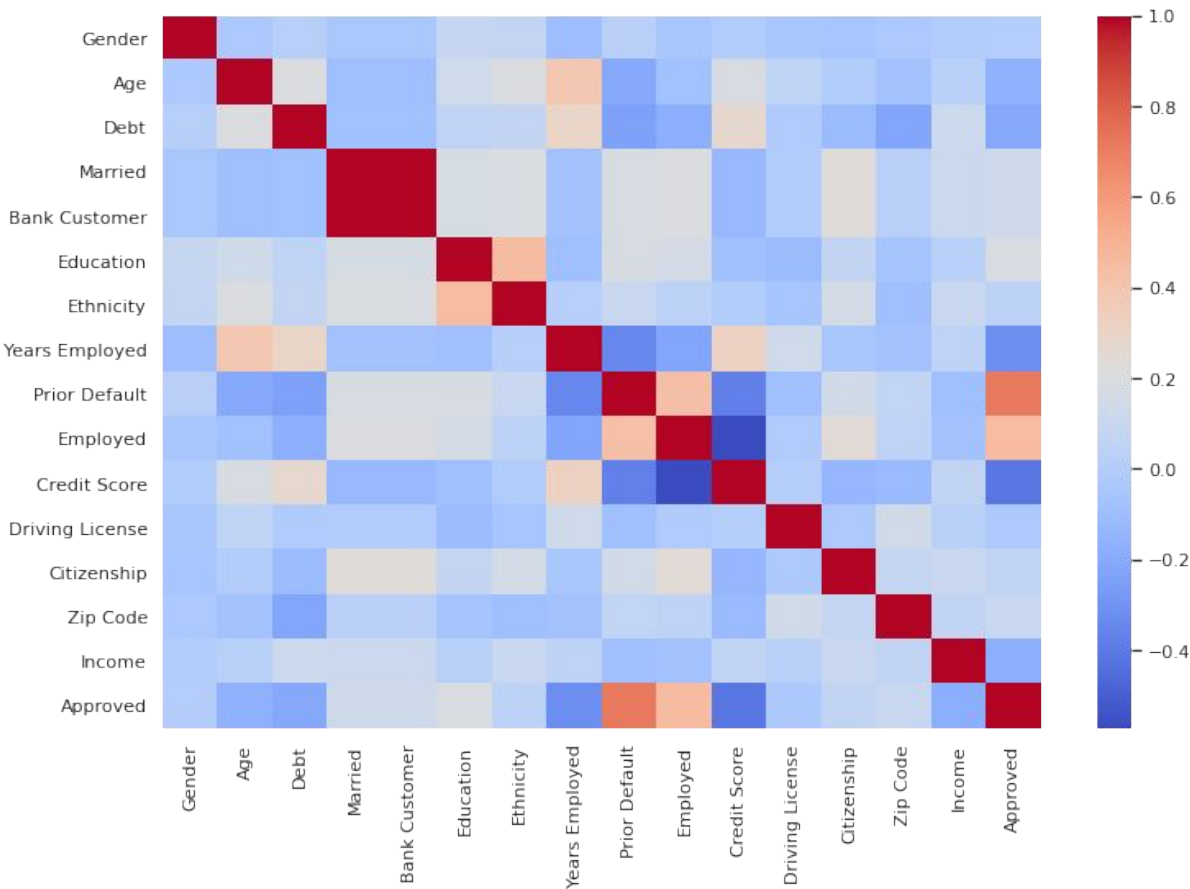






The correlation matrix tells us that some features like Married, Gender, Citizenship does not play any important role in the credit card approval.

We can simply drop these attributes

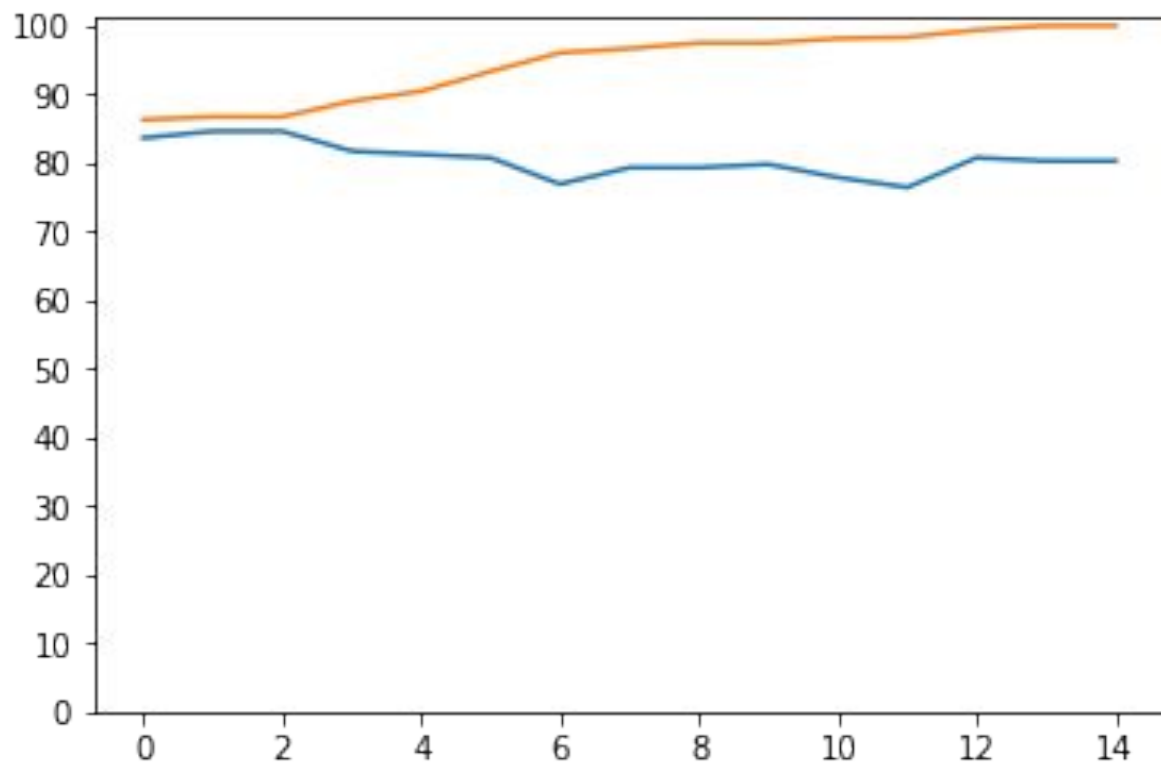


# Model Selection

Because the output is classification type so we can't use Linear Regression here, but can use Decision Tree Classifier, Logistic Regression

Using Decision Tree Classifier we got the following accuracies

Depth	Training	Testing
1	0.8631	0.8029
2	0.8672	0.8365
3	0.8672	0.8462
4	0.89	0.8462
5	0.9046	0.8173
6	0.9336	0.8125
7	0.9606	0.8077
8	0.9668	0.7692
9	0.9751	0.7933
10	0.9751	0.7933
11	0.9813	0.7981
12	0.9834	0.7788
13	0.9938	0.7644
14	1	0.8077
15	1	0.8029



# Using Logistic Regression

We also used Logistic Regression but the accuracy is not that good as DTC.

Its accuracy was 72% to 80%

# Summary

- 1) Data Source
- 2) Data Analysis
- 3) Data Cleaning
- 4) Feature Selection
- 5) Model Selection

# Conclusion

Humans who approve request of credit card sometimes they issue for those persons which might not pay the bill on time which is called bankruptcy and is perfectly legal, then bank clear their debts.

This cause huge loss to bank, as banks are backbones of our country so it is loss for our economy so for us also.

With our system it is possible to pre-determine the eligible person whom to issue credit card whom to not which saves our country's economy and human time.