REPORT

LOAN STATUS PREDICTION

Submitted in partial fulfillment of the requirements

of the degree of Bachelor of Engineering in Information Technology by Group 12

Pranav Awhad(02)

Aaryan Mehta(29)

Atharv Parab (38)

under the guidance of

Supervisor (s):

Mrs. Asha Bharambe



**Department of Information Technology**

**Vivekanand Education Society’s Institute of Technology** 

**(Affiliated to University of Mumbai, Approved by AICTE & Recognized by Govt. of Maharashtra)**

**Vivekanand Education Society’s Institute of Technology 2021-2022```````````````**

**2** 

**Vivekanand Education Society’s Institute of Technology**

**(Affiliated to University of Mumbai, Approved by AICTE & Recognized by Govt. of Maharashtra)**

**Department of Information Technology CERTIFICATE**

This is to certify that **Pranav Awhad (02), Aaryan Mehta (29), Atharv Parab (38)** of Third Year Information Technology studying under the University of Mumbai have satisfactorily presented the Mini Project entitled **LOAN STATUS PREDICTION** as a part of the MINI-PROJECT for Semester-VI under the guidance of **Prof. Charusheela Nehete** in the year 2021-2022.

Date:

(Name and sign) (Name and sign) Head of Department Supervisor/Guide

**Vivekanand Education Society’s** 

**Institute of Technology**

**(Affiliated to University of Mumbai, Approved by AICTE & Recognized by Govt. of Maharashtra)**

**DECLARATION**

We, **Pranav Awhad (02), Aaryan Mehta (29), Atharva Parab (38)** from **D15B**, declare that this project represents our ideas in our own words without plagiarism and wherever others' ideas or words have been included, we have adequately cited and referenced the original sources.

We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our project work.

We declare that we have maintained a minimum 75% attendance, as per the University of Mumbai norms.

We understand that any violation of the above will be cause for disciplinary action by the Institute.

Yours Faithfully,

1.Pranav Awhad

2.Aaryan Mehta

3. Atharv Parab

(Name & Signature of Students with Date)

**Acknowledgement**

**Vivekanand Education Society’s** 

**Institute of Technology**

**(Affiliated to University of Mumbai, Approved by AICTE & Recognized by Govt. of Maharashtra) Vivekanand Education Society’s**

**Institute of Technology**

**(Affiliated to University of Mumbai, Approved by AICTE & Recognized by Govt. of Maharashtra)**

Table of Contents

**List of Tables**

| **Introduction** | **1.1** |
| --- | --- |
| **Problem Statement** | **1.2** |
| **Objectives** | **1.3** |
| **Scope** | **1.4** |
| **Literature Survey** | **2** |
| **Analysis** | **3.1** |
| **Need of the project** | **3.2** |
| **Proposed System** | **3.3** |
| **Data Methodology** | **3.4** |
| **Results and Hypothesis** | **4** |
| **Technology Used** | **4.1** |
| **Conclusion** | **5** |

**1.1. Introduction**

* The Banking Industry always needs a more accurate predictive modeling system for many issues.
* Predicting credit defaulters is a difficult task for the banking industry.
* The loan status is one of the quality indicators of the loan.
* The loan status is used for creating a credit scoring model.
* The credit scoring model is used for accurate analysis of credit data to find defaulters and valid customers.

**1.2. Problem Statement**

* The two most pressing issues in the banking sector are: 1) How risky is the borrower? 2) Should we lend to the borrower given the risk?
* The response to the first question dictates the borrower's interest rate. Interest rate, among other things (such as time, value of money), tests the riskiness of the borrower, i.e. the higher the interest rate, the riskier the borrower. We will then decide whether the applicant is suitable for the loan based on the interest rate.
* Banking processes use manual procedures to determine whether or not a borrower is suitable for a loan based on results.
* Manual procedures were mostly effective, but they were insufficient when there were a large number of loan applications. As a result, the loan prediction machine learning model can be used to assess a customer's loan status and build strategies.
* This model extracts and introduces the essential features of a borrower that influence the customer's loan status. Finally, it produces the planned performance (loan status).

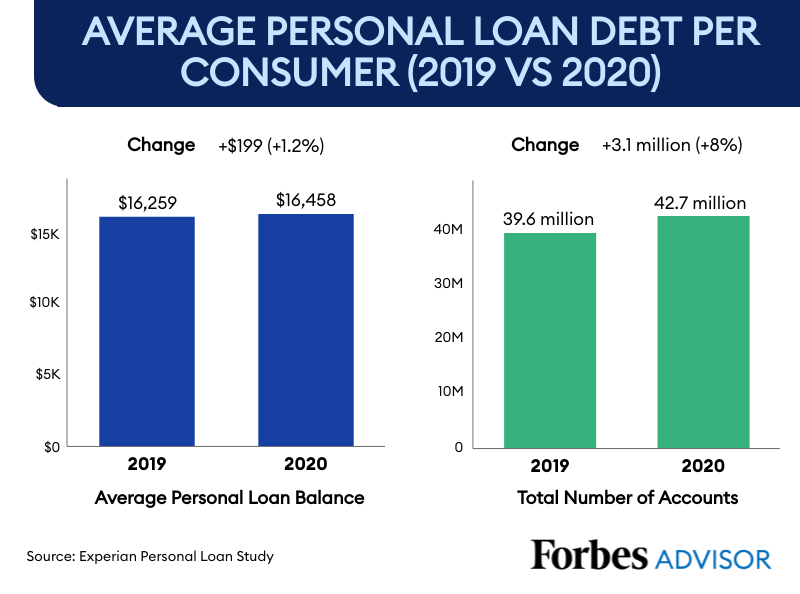
**1.3. Objectives**

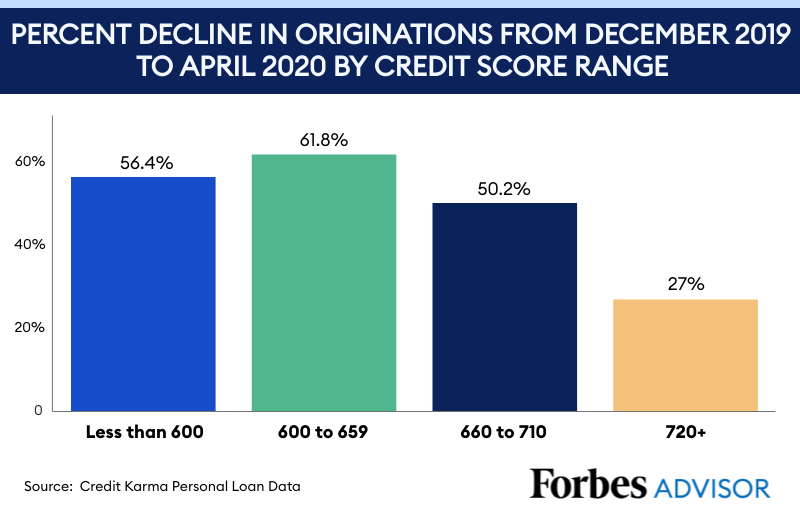
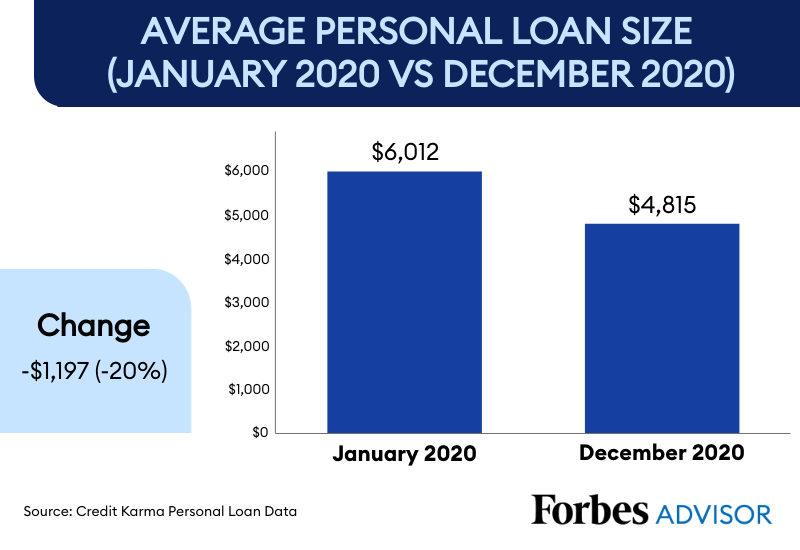
* The main objective of the system is to pick out which customer will be able to pay the debt and which customer will likely not pay their debts
* Algorithms like logistic regression,decision tree or random forest
* The main objective of the system is to predict if the person will return his or her debt
* A classification model is run on data attempting to classify whether the person is eligible to get loan from bank with good accuracy of statement
* We predict the loan data using this system

**1.4. Scope**

* The main scope of the Loan status prediction system is to help the banks beware of frauds
* This system is very useful for the government as well since no one will be able to con them and flee to another country

**2. Literature Survey**

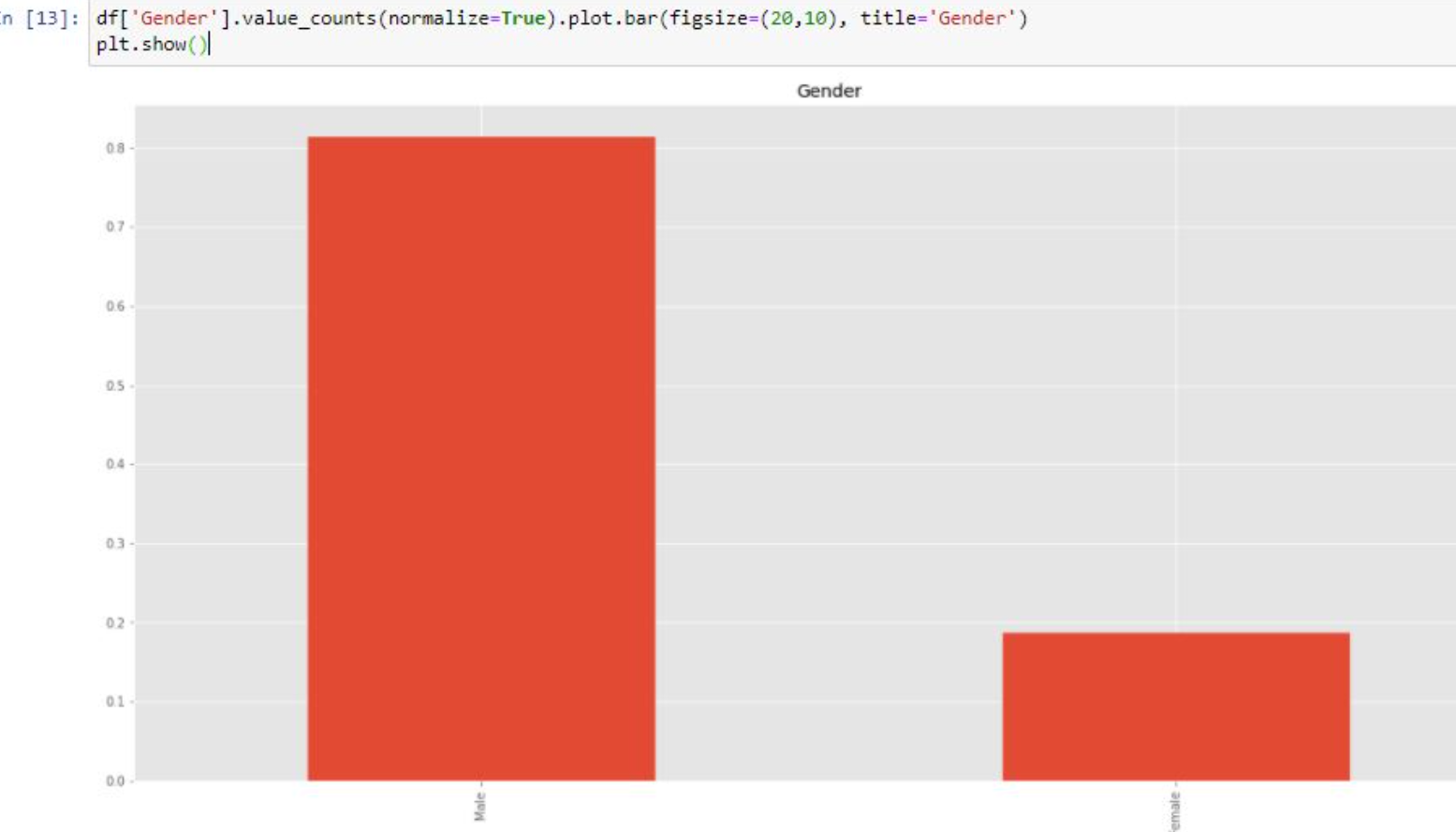
****A major portion of risk management is the approval of loans to promising candidates. But the black-box nature of Machine learning algorithms makes many loan providers vary the result

* Ashlesha Vaidya used logistic regression as a probabilistic and predictive approach to loan approval prediction. The author pointed out how Artificial neural networks and Logistic regression are most used for loan prediction as they are easier comparatively develop and provide the most accurate predictive analysis
* But the nonlinear effect and power terms are easily handled by Logistic regression as there is no need for the independent variables on which the prediction takes place to be normally distributed..
* Classification and Regression Trees ar referred to as CART (in short) introduced by Leo Breiman. It best suits both predictive and decision modeling problems. This Binary Tree methodology is the greedy method is used for the selection of the best splitting. Although Decision trees gave us a similar accuracy. The benefits of Decision Trees, in this case, were due to the latter giving equal importance to both accuracy and prediction. This model became successful in making a lower number of False Predictions to reduce the risk factor.
* Some machine learning models give different weights to each factor but in practice sometimes loans can be sanctioned based on a single strong factor only. To eliminate this problem J. Tejaswini and T. Mohana Kavya in their research paper have built a loan prediction system that automatically calculates the weight of each feature taking part in loan processing and on new test data the same features are processed concerning their associated weight. They have implemented six machine learning classification models using R for choosing the deserving loan applicants. The models include Decision Trees, Random Forest, Support Vector Machine, Linear Models, Neural Network and Adaboost.
* Anchal Goyal and Ranpreet Kaur discuss various ensemble algorithms. Ensemble algorithm is a supervised machine learning algorithm that is a combination of two or more algorithms to get better predictive performance. They carried out a systematic literature review to compare ensemble models with various stand-alone models such as neural network, SVM, regression, etc. The authors after reviewing different literature reviews concluded that the Ensemble Model performs better than the stand-alone models. Finally, they concluded that the concept of combined algorithms also improves the accuracy of the model.
* 

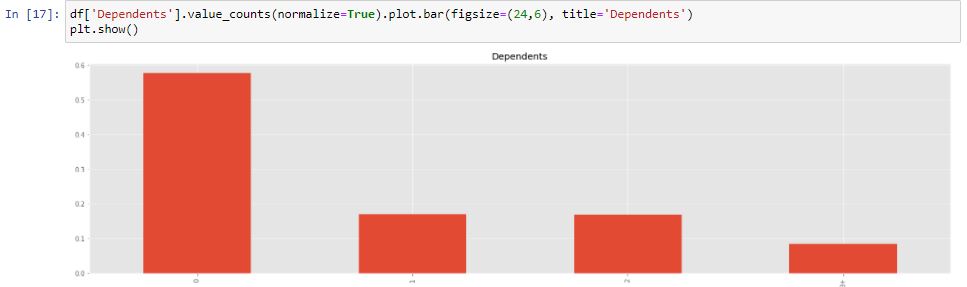
**3. Analysis of the system**

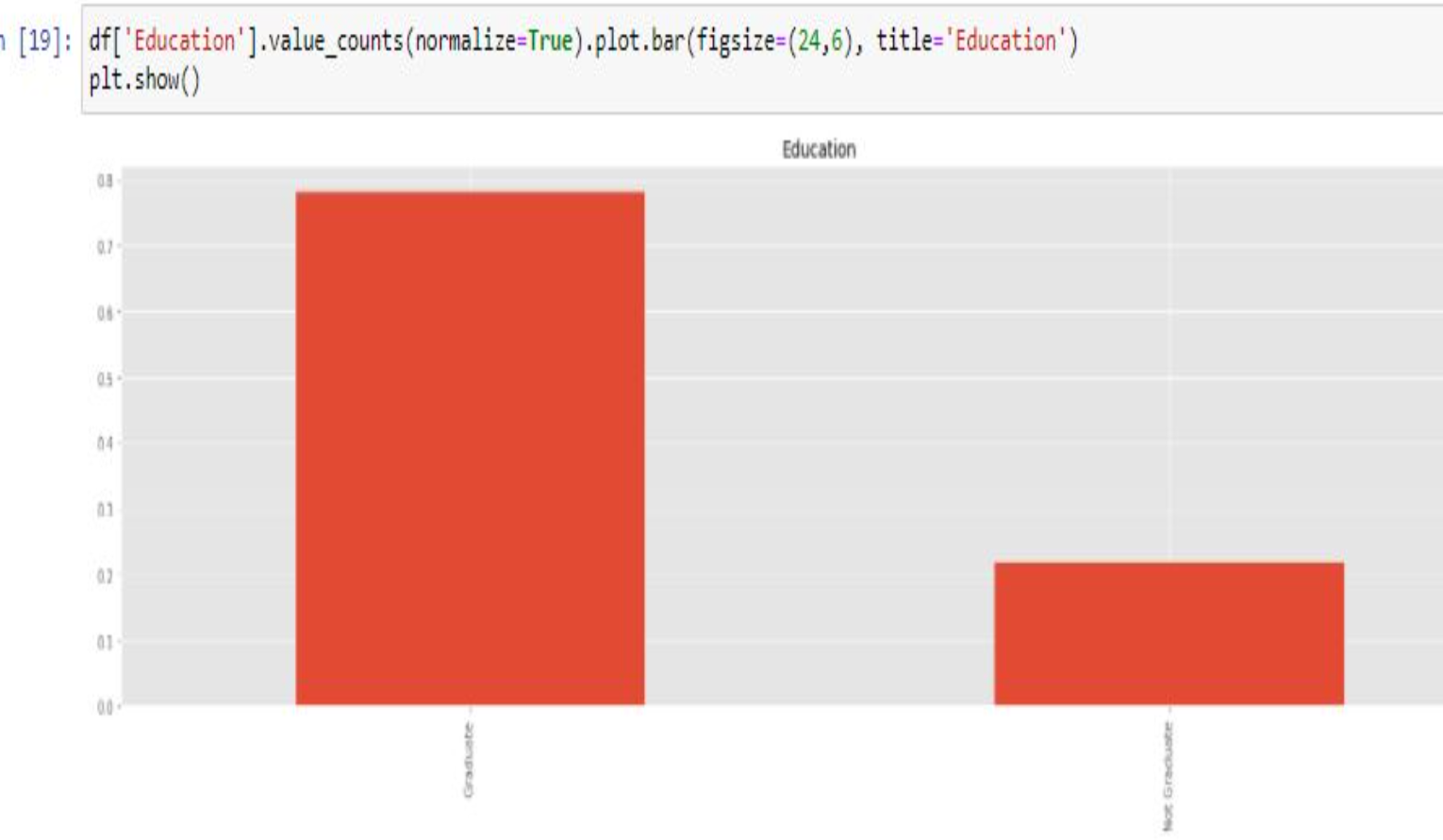
**Univariate analysis**

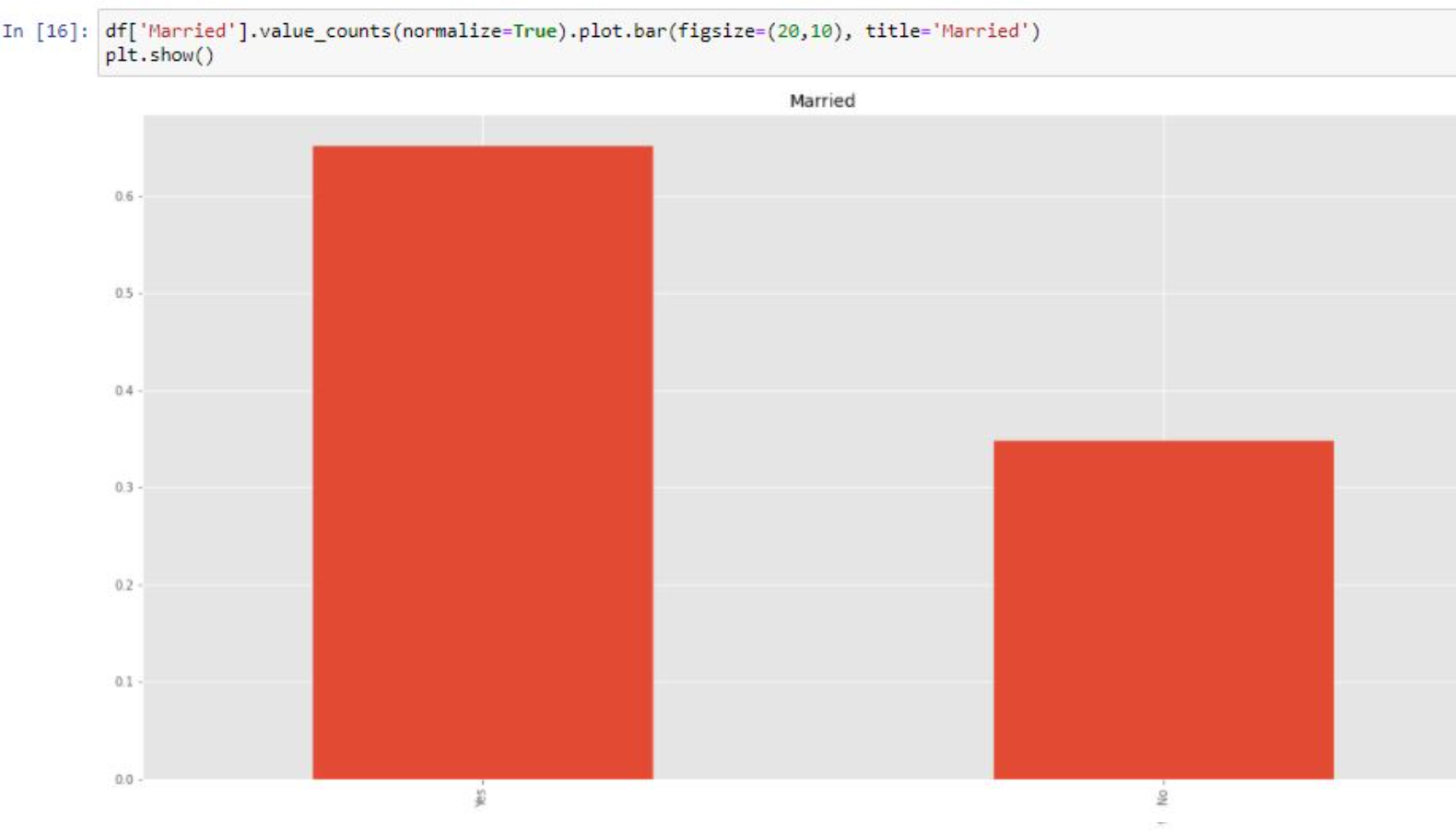
**a)**

****

**b)**

****

**c)**

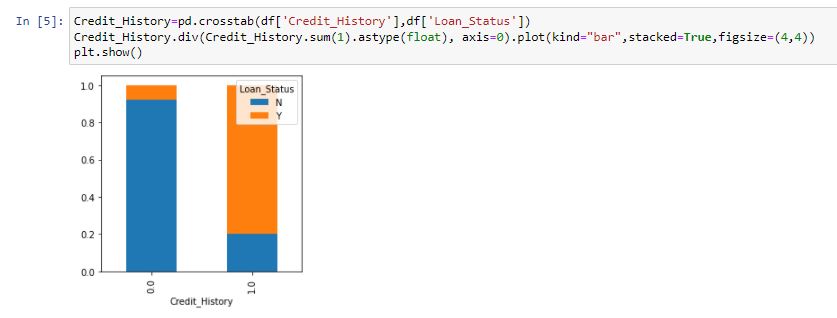
**d)**

**Bivariate analysis**

**a)**

**b)**

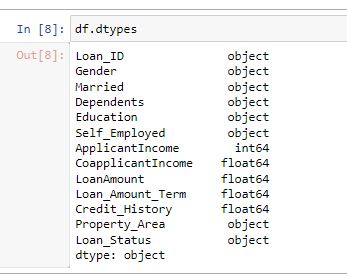
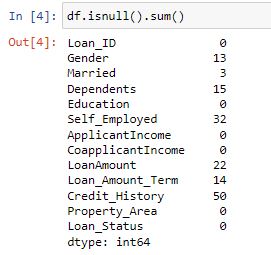
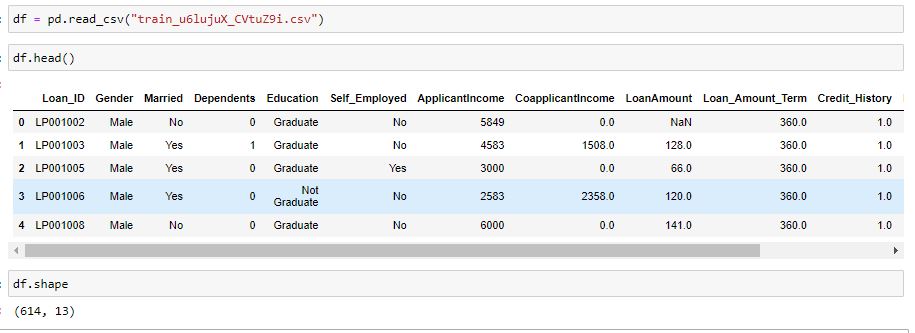
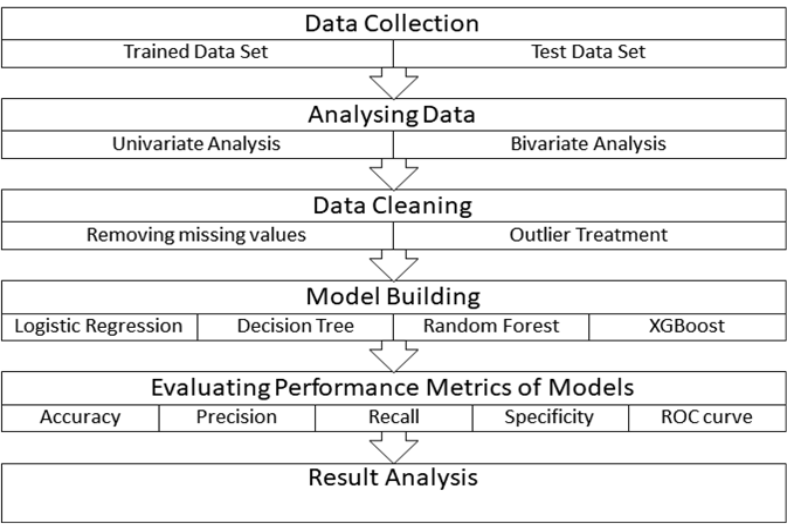
**c)**

**3.1. Need of the Project**

* In Today’s world , the prices of everything have increased whether one has to buy a new house , start a new life at a new place or to go abroad for further education.
* Most of the people take out loans and fulfill their dreams and ambitions.
* So, it’s important that banks have an efficient system to predict loan status and approve the valid customers
* To automate this process, it has now become a necessity to have a system that reviews the application details and predicts the status of the loan.
* Loan Status Prediction System would help banks to make the process quick and error free.

**3.3. Proposed System**

**3.4.Data Methodology**

****

**4.1.Results and Hypothesis**

* Applicants with high incomes should have more chances of loan approval.
* Applicants who have repaid their previous debts should have higher chances of loan approval.
* Loan approval should also depend on the loan amount. If the loan amount is less, the chances of loan approval should be high.
* The Less the amount to be paid monthly to repay the loan, the higher the chances of approval.

**4.2.Technology Used**

* Web Development
* HTML
* CSS
* Javascript
* PHP
* Visualization/Machine Learning
* Jupyter Notebook
* Python
* Pandas (Data Analysis Library)
* Seaborn (Data Analysis Library)
* Scikit Learn (Data Analysis Library)

**5.Conclusion**

##### We did Exploratory data Analysis on the features of this dataset and saw how each feature is distributed.We did bivariate and univariate analysis to see the impact of one another on their features using charts.We analyzed each variable to check if data is cleaned and normally distributed.We cleaned the data and remove NA valuesWe also generated hypotheses to prove an association among the Independent variables and the Target variable. And based on the results, we assumed whether or not there is an association.We calculated correlation between independent variables and found that applicant income and loan amount have significant relation.We created dummy variables for constructing the model

##### Thus we developed a loan status predictor system

\

**References**[**https://www.kaggle.com/dansbecker/classification**](https://www.kaggle.com/dansbecker/classification)

**https://www.kaggle.com/yonatanrabinovich/loan-prediction-dataset-ml-project**

E-Commerce- The Cutting Edge of Business, Bajaj & Nag- New Delhi 2000 9. P.

Malhotra and B. Singh, "Determinants of internet banking adoption by banks in

India", Internet Research, vol. 17, no. 3, (2007), pp. 323-339.

Electronic Commerce from Wikipedia- the free encyclopedia

http://en.wikipedia.org/wiki/Electronic\_commerce, 22/1/2014.

.