College Admission (R Project)

Background of the problem

Every year thousands of applications are being submitted by international students for admission in colleges of the USA. It becomes an iterative task for the Education Department to know the total number of applications received and then compare that data with the total number of applications successfully accepted and visas processed. Hence to make the entire process easy, the education department in the US analyse the factors that influence the admission of a student into colleges. The objective of this exercise is to analyse the same.

Objective

Our objective here with this sample of 400 students is to find out what type of students are more likely to get selected. It is a classification problem where we have to predict whether the student made it or didn't make it to the University.

Features and variables

Attribute	Description
GRE	Graduate Record Exam Scores
GPA	Grade Point Average
Rank	It refers to the prestige of the undergraduate institution. The variable rank takes on the values 1 through 4. Institutions with a rank of 1 have the highest prestige, while those with a rank of 4 have the lowest.
Admit	It is a response variable; admit/don't admit is a binary variable where 1 indicates that student is admitted and 0 indicates that student is not admitted.
SES	SES refers to socioeconomic status: 1 - low, 2 - medium, 3 - high.
Gender_male	Gender_male (0, 1) = 0 -> Female, 1 -> Male
Race	Race – 1, 2, and 3 represent Hispanic, Asian, and African-American

Steps we followed:

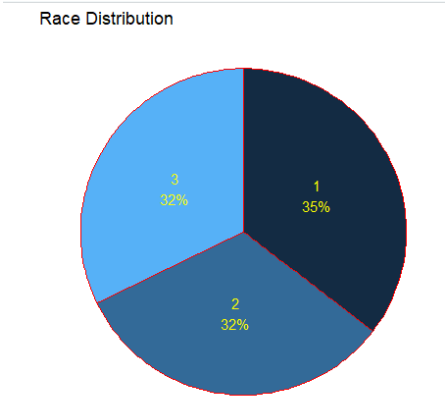
Data cleaning

In this step the dataset was checked for missing values. The numerical variables namely GPA and GRE were found to have outliers and a bit of abnormal distribution. The outliers were removed and then the dataset became fit for analysis.

Descriptive statistics

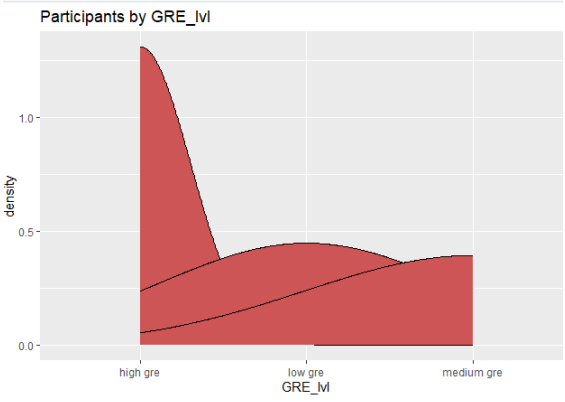
We found the following insights from the dataset:

- The dataset is almost uniformly distributed in case of Racial groups but Hispanics are slightly in majority.

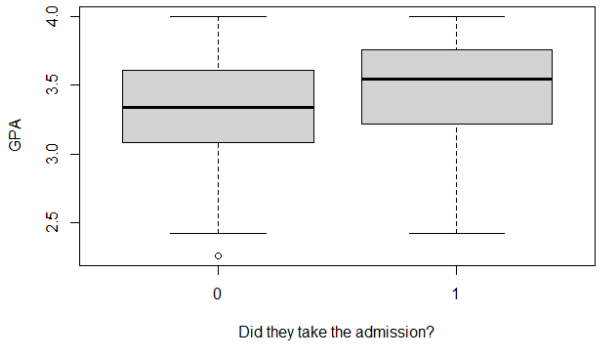


- There is moderate positive correlation between GRE and GPA.

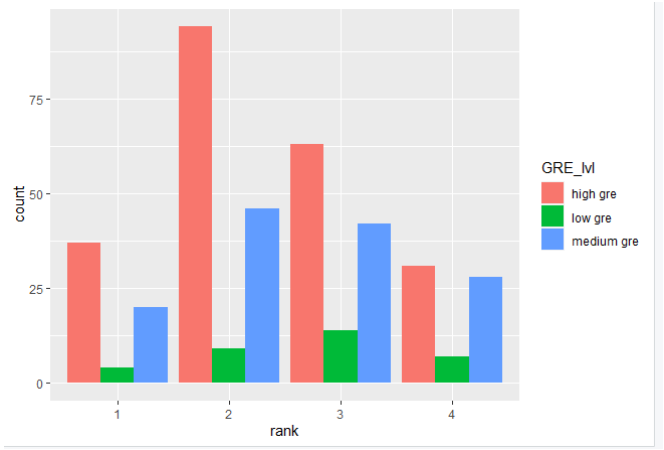
- Majority of the respondents got a high GRE.



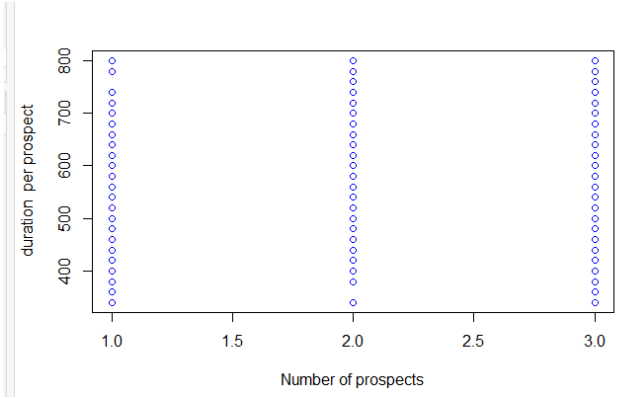
- It has been found that those with higher GPA are more likely to get an admission.



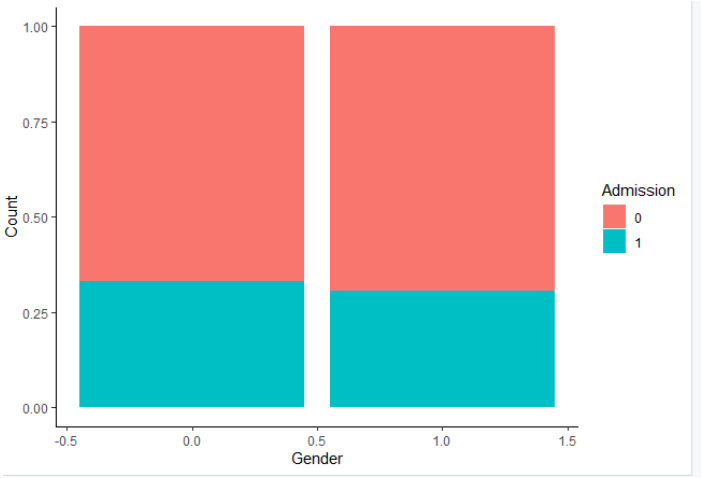
- Those with Rank 2 have highest GRE.



- The GRE marks doesn't depend on social economic status.



- Females are more likely to get admission than men



Model Building

In this step we are going to build models that will classify whether the student will get admission or not. They will use Machine learning algorithms for that purpose of classification.

We split the dataset into 70:30, 70% data is to be used for training the model and 30% data for testing the model.

Model1: Logistic Regression

The model is significant and is able to explain the phenomenon.

The accuracy of the model is 72%.

Model2: SVM

The model is significant and is able to explain the phenomenon.

The accuracy of the model is 67%.

Model3: Decision Tree classifier

The model is significant and is able to explain the phenomenon.

The accuracy of the model is 68%.