

PREDICTION OF CHANCE OF GETTING ADMIT INTO UNIVERSITY USING ADMISSIONS DATASET

|  |
| --- |
|  |
|  |
| Group 19  Abhinav Saluja  Prarthna Vasudevamurthy  Shivank Aggarwal  Sruthi Bandarupalli  Sindhu Nagaraj |

# INTRODUCTION

Problem Statement:

The United States has attracted millions of international students to its universities and colleges because the USA offers so many choices and some of the best faculties in the world. With more than 3,000 universities and colleges in united states, the options are almost limitless. Yet, because the choices are so varied, deciding which university to attend is not an easy choice.

There’s also uncertainty about whether the student will get admitted to a university or not. The price of the admissions process often totals hundreds of dollars, accounting for the costs associated with standardized tests, test-prep resources and application fees. This Project will help students will to guess their capacities and to decide whether to apply for a master's degree in a particular University or not.

Objective:

To provide a quality model for the project beneficiaries in order for them to predict their chances of getting university admits by selecting the most accurate model to predict the probability of admission.

Monetary Benefit:

* Students Often have to spend huge amounts of money while applying to many universities not knowing where they would get accepted, also there are a plethora of University Consulting Advisors who charge a massive amount of money from Students to help them choose where to apply
* This project will help students to narrow down their options, aiding them in applying to schools when their chance of admission is high, ultimately reducing application costs.

Project Beneficiaries:

1. Students Applying to Graduate Schools and Universities
2. University Admission Committee
3. Private University Consulting Advisors helping students with Graduate school applications and admissions process

# DATA EXPLORATION

Dataset:

This dataset is created for prediction of Graduate Admissions from an Indian perspective. 500 applicants have been surveyed as potential students for UCLA. The university weighs certain aspects of a student's education to determine their acceptance.

The dataset is clean, has 500 Observations with no null records and contains 7 parameters considered important during the application for Masters Programs

* Independent Variables

1. GRE Scores (out of 340)
2. TOEFL Scores (out of 120)
3. University Rating (rated from 1 to 5)
4. Statement of Purpose (rated from 1 to 5)
5. Letter of Recommendation Strength (rated from 1 to 5)
6. Undergraduate GPA (out of 10)
7. Research Experience (either 0 or 1, 1 meaning student has Research Experience)

* Dependent Variable

Chance of Admit (ranging from 0 to 1)

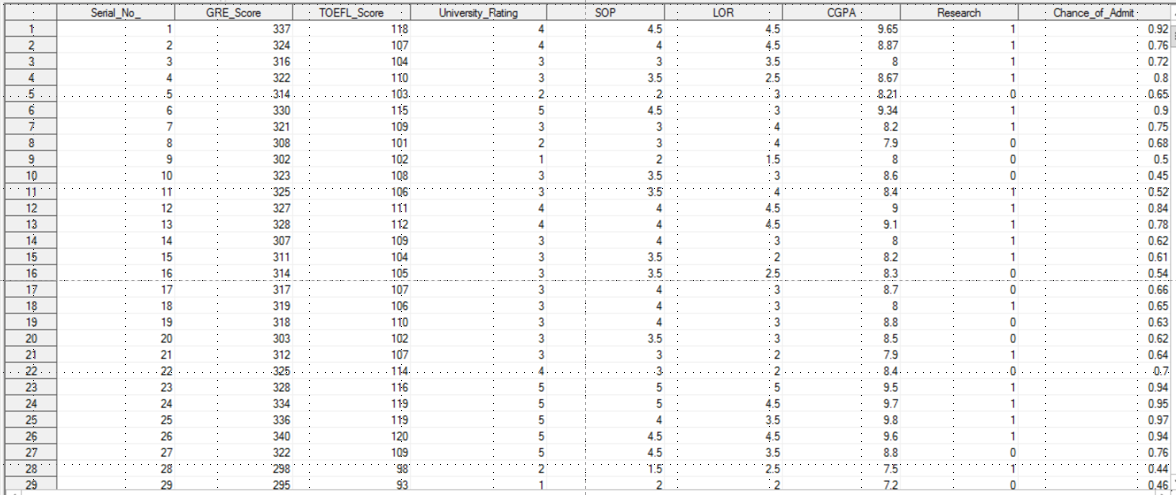


Fig1.1 Screen shot of the clean Data

Correlation Between all the variables:



Fig 1.2 Correlation matrix produced in SAS for all variables in the Dataset

* The correlation between GRE Score and TOEFL Score is very high at 0.83
* The Correlation between CGPA with GRE Score and TOEFL Score is also quite high at 0.83 and 0.81 respectively
* We observe that most of the variables in the dataset are highly corelated with each other. This is expected as the dataset is education related dataset and we can assume that all the parameters are correlated to how good a student is academically.
* We use Factorial analysis to understand the variability among these highly correlated variables.

Factor Analysis:

Factor analysis is a statistical method used to describe variability among observed, correlated variables in terms of a potentially lower number of unobserved variables called factors. For example, it is possible that variations in six observed variables mainly reflect the variations in two unobserved (underlying) variables. Factor analysis searches for such joint variations in response to unobserved latent variables. The observed variables are modelled as linear combinations of the potential factors, plus "error" terms. Factor analysis aims to find independent latent variable. The theory behind factor analytic methods is that the information gained about the interdependencies between observed variables can be used later to reduce the set of variables in a dataset.

An important feature of factor analysis is that the axes of the factors can be rotated within the multidimensional variable space. In simple terms the factor analysis program looks first for the strongest correlations between variables and the latent factor, and makes this Factor 1. Visually, one can think of it as axis (Axis 1) then the factor analysis program looks for the second set of correlations and calls it Factor 2, and so on.

Another option/ Criterion for determining the number of factors is the Scree Plot. The Cattell scree test plots the components as the X-axis and the corresponding eigenvalues as the Y-axis. As one moves to the right, toward later components, the eigenvalues drop. When the drop ceases and the curve make an elbow toward less steep decline, Cattell's scree test says to drop all further components after the one starting at the elbow.

The factoring in our analysis is as follows:

Factor 1:

* CGPA
* University Rating
* LOR
* SOP

Factor 2:

* GRE
* TOEFL
* Research

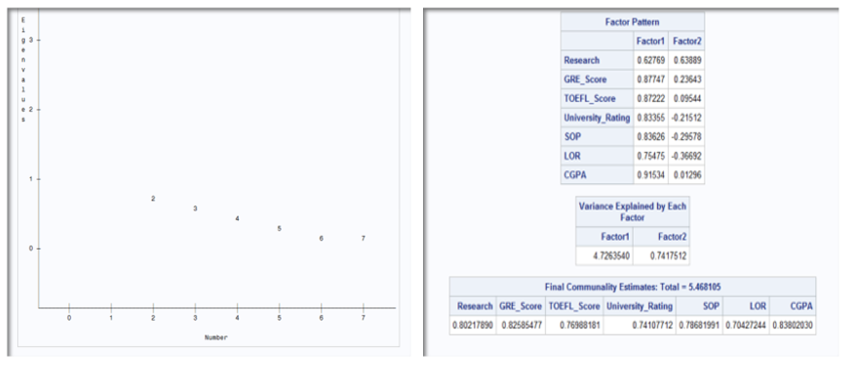


Fig 1.3: 2 factors will be retained by the NFACTOR criterion

Rotation Methods:

The unrotated output maximizes variance accounted for by the first and subsequent factors, and forces the factors to be orthogonal. This data-compression comes at the cost of having most items load on the early factors, and usually, of having many items load substantially on more than one factor. Rotation serves to make the output more understandable, by seeking so-called "Simple Structure": A pattern of loadings where each item loads strongly on only one of the factors, and much weakly on the other factors. Rotations can be orthogonal or oblique (allowing the factors to correlate).

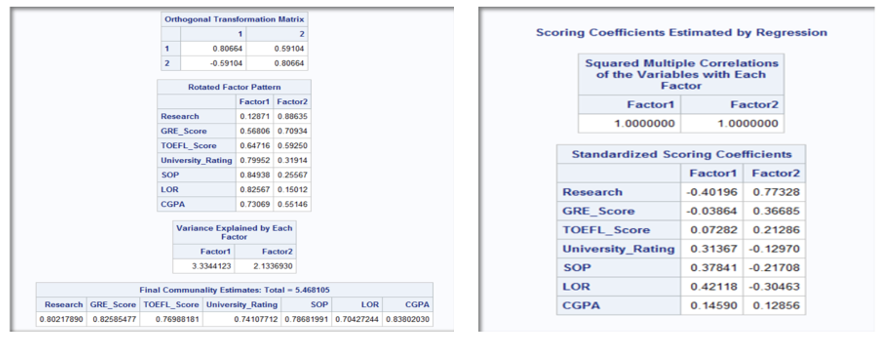
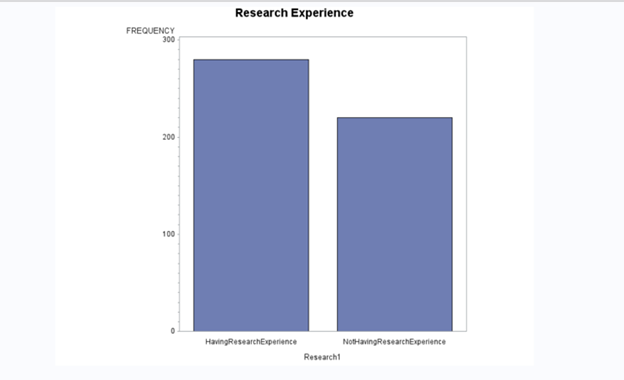


Fig 1.4 Rotation Method using SAS

* The initial solution results in strong correlations of a variable with several factors or in a variable that has no strong correlations with any of the factors.
* In order to make the location of the axes fit the actual data points better, the program rotates the axes. Ideally, the rotation will make the factors more easily interpretable.

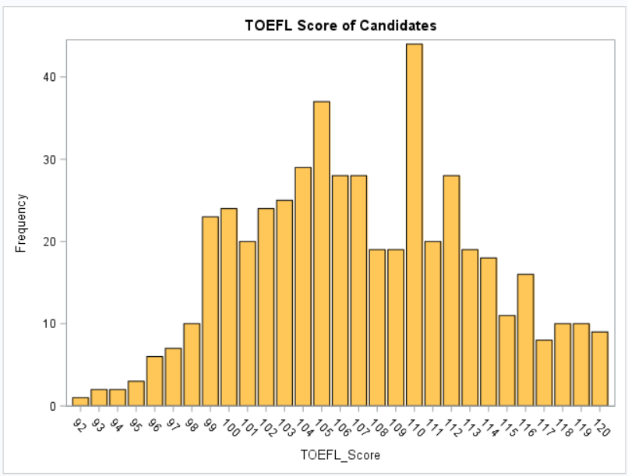
# Data Visualization

Number of Students involved in research



The graph clearly indicates that a greater number of students have been involved or have research experience as compared to students with no prior research experience.

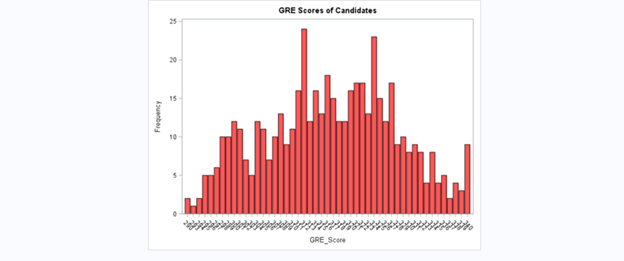
Plot for TOEFL Scores



This histogram shows the frequency for TOEFL Scores.

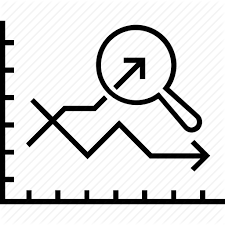
There is a density between 92 and 120. The minimum and maximum scores among university students are reported to be 92 and 120.

Histogram – GRE Scores

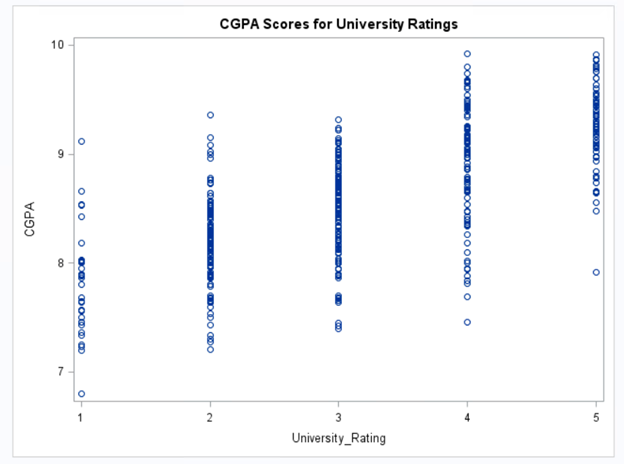


* This histogram shows the frequency for GRE scores.
* There is a density between 290 and 340. Being within this range would be a good feature for a candidate to stand out.
* How many students have good chances for getting an admit?

Let’s have a look at relations between the variables

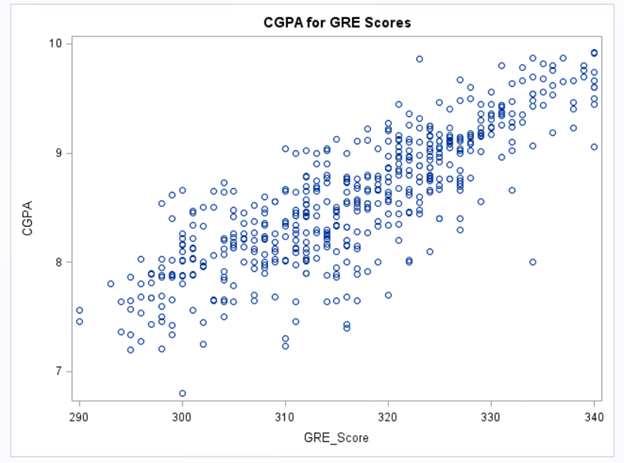


CGPA v/s University Ratings



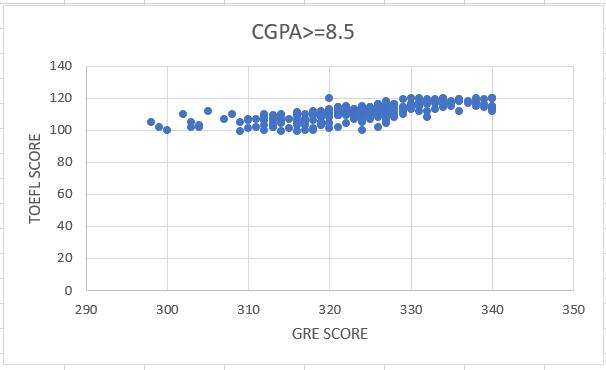
Clearly, high CGPA is associated with higher University Ratings. This is a clear indication that a higher CGPA in your undergraduate college can land you to a good university with a high ranking.

CGPA v/s GRE Scores



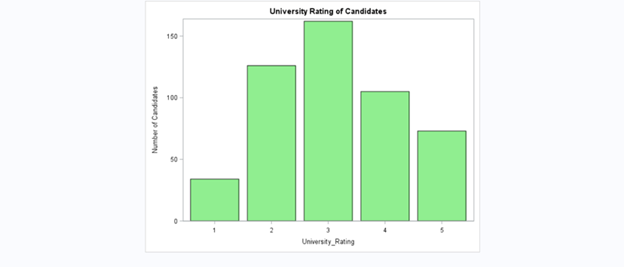
Students who have high CGPA have a tendency to score high score in GRE. We can associate it with high Intelligence Quotient (IQ). Students with high IQ have a capability to score well in exams, be it the university exams or other competitive exams like GRE.

TOEFL v/s GRE (CGPA>=8.5)



University students having high score have a tendency to score high in TOEFL given that their CGPA is higher than 8.5.

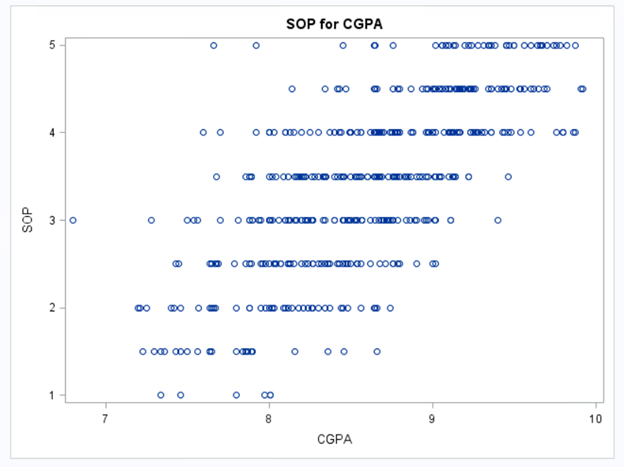
University Rating v/s Number of Students



Most of the students from our data set are from the Average Universities with Rating 3.

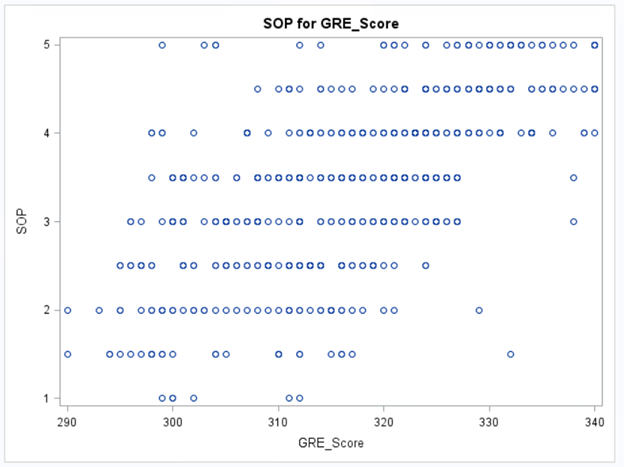
The data set has less number of students from low ranking universities.

SOP v/s CGPA



Candidates with high CGPA usually have high SOP scores.

GRE Score v/s SOP Score



Candidates with high GRE scores are expected of getting high SOP scores too.