Predicting Graduate Admission in New York state colleges

1. **Introduction**

For those students looking to further their education the state of New York, has 240 universities and colleges to choose from. With New York City having many schools in such a small area, you are guaranteed to find a program of your liking in this great city. There is no question that social aspects are a main factor to come to college in New York, but some of the highest quality schools are also available to you in order to provide you with an exceptional education that you can bring with you for the rest of your life. If you are looking to settle down and find a job in New York, you will be more likely to find a job as employers will be able to recognize the name of your college or university. Additionally, there are many other great college towns in the state of New York to attend college, such as Buffalo, Albany, Ithaca, and Syracuse, where one can begin the first step to achieving their dreams.

1. **Problem Description:**

To apply for a master's degree is a very expensive and intensive work. With this project, students can predict their capacities and guess in which college they will get admission for a master's degree admission from an Indian perspective.

For this project, I am going to put on my entrepreneur hat and create a simple guide to help master’s degree aspirants to predict in which college they will get graduate admission and what will be the expected cost for study.

1. **Data acquisition and cleaning**

**Data sources:**

For this project, I downloaded or scraped from multiple sources were combined into one table. The links for the data sets are given below:

1. <https://www.free-4u.com/Colleges/New-York-Colleges>
2. <https://www.kaggle.com/mohansacharya/graduate-admissions>

Features in the dataset: There are 13 columns and 239 rows in the data set

Columns are listed below:

* School (Name of the college)
* City (In which city the college belongs to)
* Enroll (The number of students enrolled in the previous year)
* In-State Cost (Total fees for in state students)
* Out-State Cost (Total fees for out state students)
* GRE Scores (290 to 340)
* TOEFL Scores (92 to 120)
* University Rating (1 to 5)
* Statement of Purpose (1 to 5)
* Letter of Recommendation Strength (1 to 5)
* Undergraduate CGPA (6.8 to 9.92)
* Research Experience (0 or 1)
* Chance of Admit (0.34 to 0.97)

1. **Exploratory Data Analysis**

There were missing values in the In\_sate cost and Out\_stae cost fields it has been handled by using idmax function. The 3 most important features for admission to the Master: CGPA, GRE SCORE, and TOEFL SCORE and the 3 least important features for admission to the Master: Research, LOR, and SOP. Most of the candidates in the dataset have research experience. Therefore, the Research will be a unimportant feature for the Chance of Admit. The correlation between Chance of Admit and Research was already lower than other correlation values.

Understanding the relation between different factors responsible for graduate admissions.

The Instatecost and Outsatecost fields are divided into three categories (High, Avg, and Low) cost, more than 70% of colleges have high fees structure. From the graph it is visible that the cost of study is more in New York state colleges.

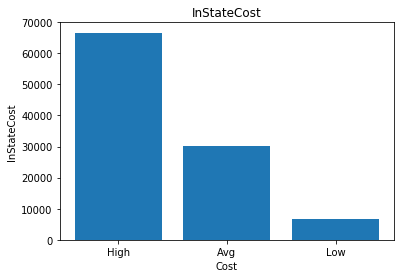


Fig1: Insatecost fees distribution

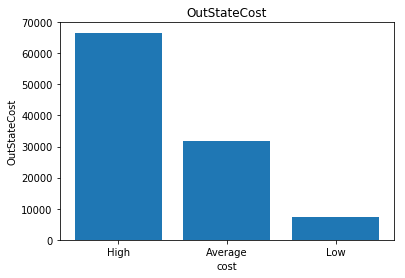
****

Fig 2: Outsatecost fees distribution

Relationship between GRE and TOFEL score:

People with higher GRE Scores also have higher TOEFL Scores which is justified because both TOEFL and GRE have a verbal section which although not similar are relatable.

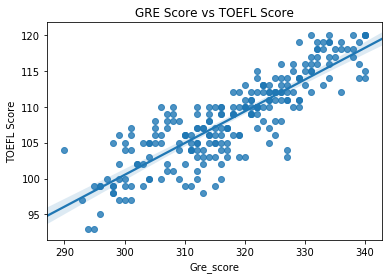


Fig3: GRE Score vs TOEFL Score

There is a linear positive relationship between the GRE Score and TOEFL score.

Relationship between GRE score and CGPA:

It is an assumption that people with higher CGPA usually have higher GRE scores but in the case of this dataset it is true.

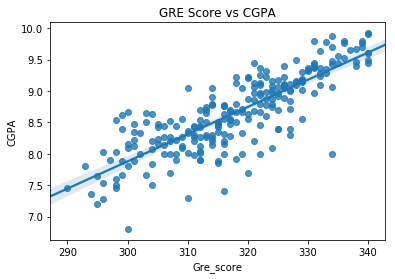


Fig4: GRE Score vs CGPA

From the above graph there is a linear positive relationship between the GRE score and CGPA.

LOP and SOR are the two features which are having less impact on graduate admission.

LORs are not that related with CGPA, so a person’s LOR is not dependent on that person’s academic excellence.

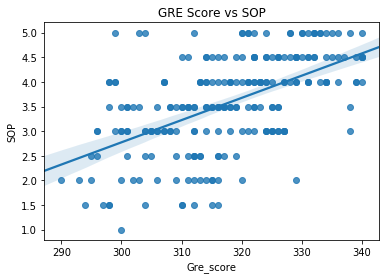


Fig5: GRE Score vs SOP

GRE scores and LORs are also not that related. People with different kinds of LORs have all kinds of GRE scores.

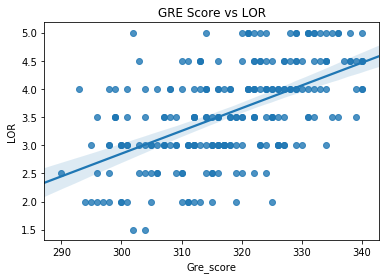


Fig6: GRE Score vs LOR

1. **Predictive Modeling:**

Regression models are used for predicting the graduate Admission. Regression analysis is a form of predictive modelling technique which investigates the relationship between a dependent and independent variable.

For predicting chance of admission here I have used three different types of regression models, Linear regression, Random forest regression and Decision tree regression.

I have used R square as the tuning and evaluation metric. R-squared is a statistical measure of how close the data are to the fitted regression line. It is also known as the coefficient of determination, or the coefficient of multiple determination for multiple regression. The definition of R-squared is straight-forward; it is the percentage of the response variable variation that is explained by a linear model. R-squared = Explained variation / Total variation.

**Comparison of Regression Algorithms**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Linear Reg | Random forest | Decision tree |
| R square value | 0.6820 | 0.6234 | 0.3786 |

Linear regression and random forest regression algorithms were better than decision tree regression algorithm.

1. **Conclusions:**

In this study I have predicted the chance of admission in New York state colleges. From the study it has been observed that GRE score and CGPA have more impact on chance of Admission.

1. **Future directions:**

We can improve the prediction using classification algorithms.