

ITCS 6155 - EDA and Creating a Dashboard Exercise

The University of North Carolina at Charlotte

Date: February 18, 2020

Group Members:

Aniruddha Sudhindra Shirahatti, Bharadwaj Aryasomayajula, Manoj Krishna Mohan, Sai Kumar Thallada, Sarang Padalkar.

Summary:

We have chosen the Graduate Admissions dataset to perform Exploratory Data Analysis (EDA) and create Dashboards using Google Data Studio.

Description of domain:

Getting admission for a graduate program is a lengthy process. To get admission into the desired university, it is very important for students to get insights about the university and its minimum eligibility criteria for acceptance of the application. The goal of this dataset is to use existing data and predict the chances of admission for a student.

Description of Data:

This is a Kaggle dataset owned by Mohan S Acharya which is inspired by UCLA Graduate Dataset. This dataset does not require pre-processing. The test scores and GPA are in the older format. The dataset contains 400 records and 8 features of type int and float which are considered important during the application process:

1. GRE Scores (out of 340)
2. TOEFL Scores (out of 120)
3. University Rating (out of 5)
4. SOP - Statement of Purpose Strength (out of 5)
5. LOR - Letter of Recommendation Strength (out of 5)
6. Undergraduate GPA (out of 10)
7. Research Experience (either 0 or 1)
8. Chance of Admit (ranging from 0 to 1)

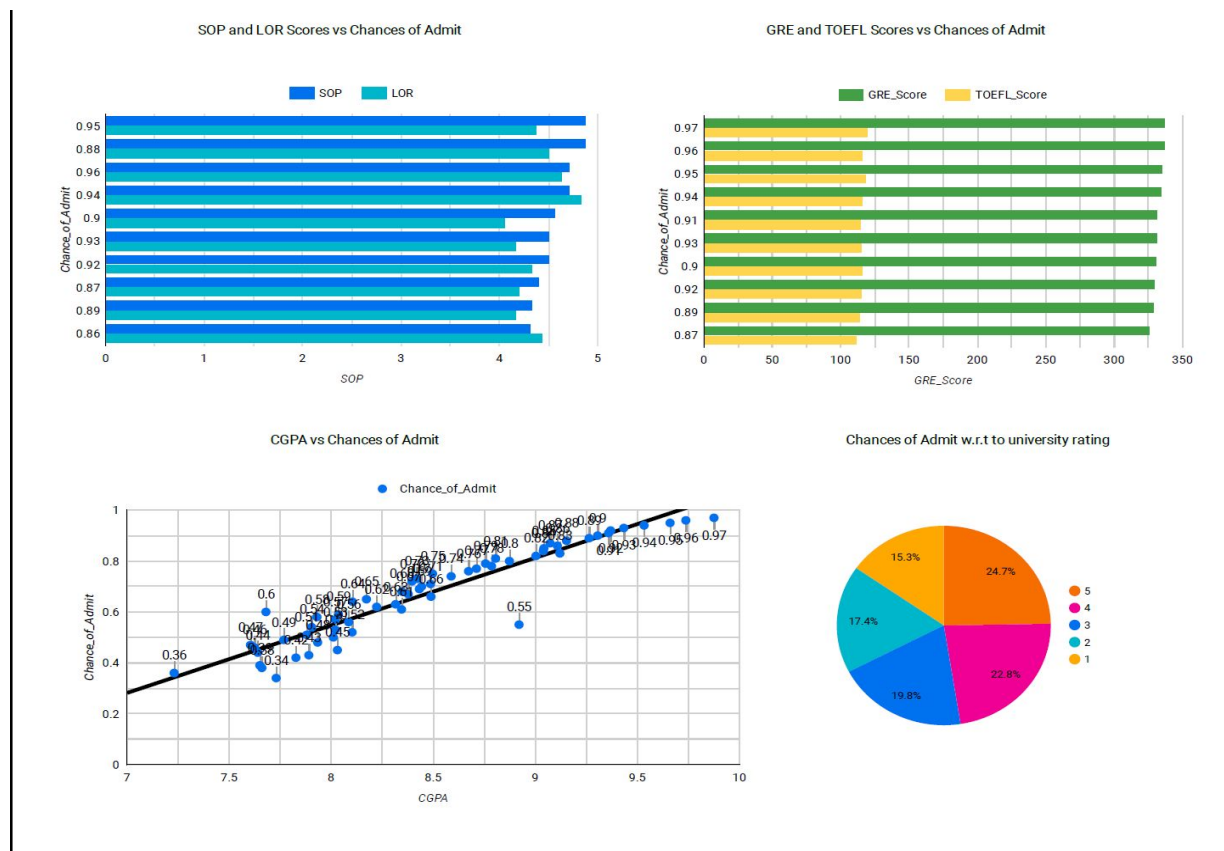
Link to Kaggle Dataset: <https://www.kaggle.com/mohansacharya/graduate-admissions>

Steps Completed:

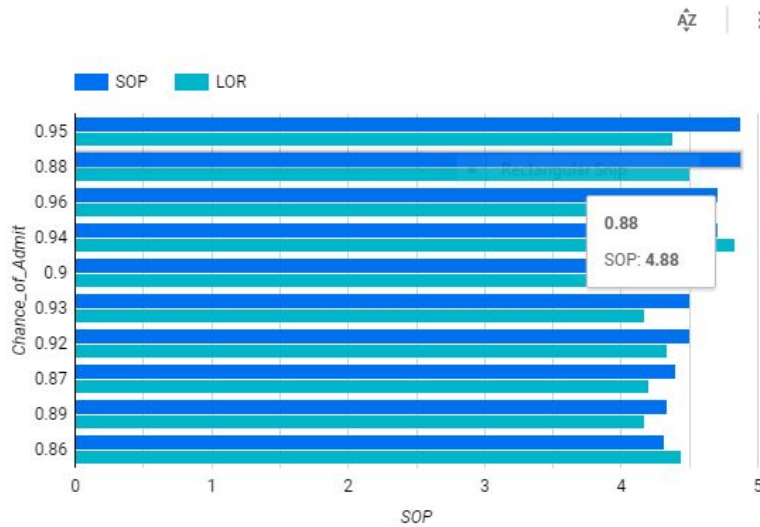
We performed the following steps for Exploratory Data Analysis (EDA) and creating Dashboards:

1. Loading Data into Google Cloud SQL
 - a. Launch the Cloud Shell terminal.
 - b. Clone files from Github repository onto the Cloud platform.
 - c. Create a SQL instance on Google Cloud Platform.
 - d. Whitelist the Cloud Shell IP address, so that we can connect to the SQL instance from the current shell.
 - e. Connect to the SQL instance and create a table in the database.
 - f. Load the data from the CSV file in the SQL table.
 - g. Confirm if the table has been loaded with data by firing a few SQL queries just to make sure that the table contains all the required data.
2. Visualizing Data with Google Data Studio
 - a. Connect the Google Data Studio with SQL database.
 - b. Select the tables or columns from the database that are useful for visualization from which we can deduce observations based on the graphs plotted.

Screenshot of Dashboard with explanation:



1. SOP and LOR Scores vs Chances of Admit



Explanation:

From the above bar plot, we can observe that with the combined scores of SOP and LOR, the chances of getting an admit increases as the combined score increases. But we can also observe that the impact of SOP on getting the chances of admission is more than the impact of the LOR.

2. GRE and TOEFL Scores vs Chances of Admit

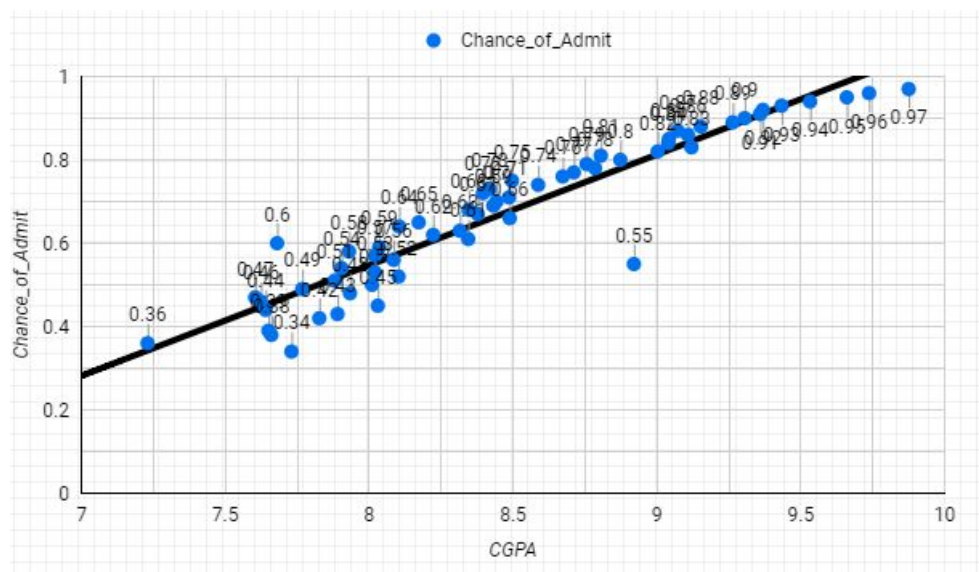


Explanation:

From the above bar plot of combined GRE and TOEFL vs the Chances of admit, we can observe that most of the average Data of TOEFL scores are very nearby to each other.

But the GRE Scores see a significant increase along with the chances of admission. From this plot, we can infer that the GRE Scores play a major role as a variable in determining the chances of admission than the TOEFL scores.

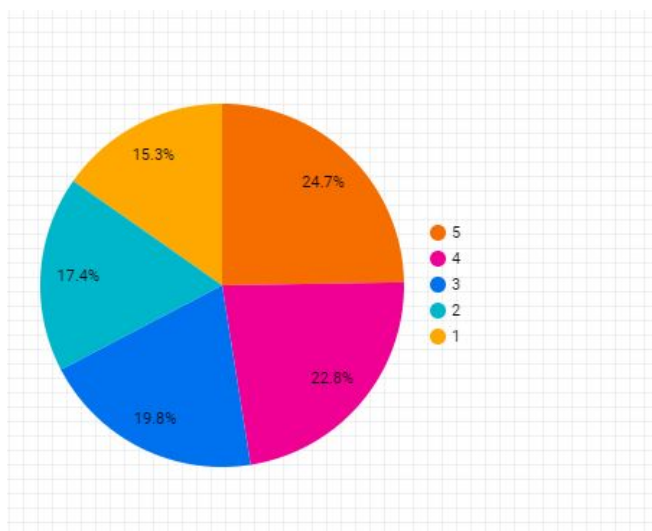
3. CGPA vs Chances of Admit



Explanation:

The above scatter plot is a simple CGPA vs Chances_of_admit plot. It depicts the linear relationship between CGPA and admittance. Higher the CGPA, the better the chances of getting an admittance. Although there appear to be outliers around the trend line which defines the relationship. Also, we need to note that the CGPA taken into consideration for plotting, ranges from 7.0 to 10. A linear trendline shows how the data is being fitted to the equation of a linear line.

4. Chances of Admit w.r.t to university rating



Explanation:

This PIE Chart depicts how difficult the chances of admission vary from each University based on the ranking. The ranking is based on a scale of 1-5. Better the University, the better the ranking. The PIE chart shows the distribution of how much the chances of admission are when compared to the University rating.

Conclusion:

Google Data Studio offers advanced reporting visualization that includes customizable dashboards, charts and graphs. With Data Studio, users can connect to various data sources to visualize, and share data from a variety of web-based platforms. It allows you to transform your data into appealing and informative reports for your audience. By using the Google Data Studio, we had visually analyzed the various factors that influence and depend on a candidate's chances of getting admission into a particular university. As per our analysis, the chances of an Indian student getting admittance in highly rated university is high if they have very high GRE, TOEFL, CGPA scores and a highly rated supporting documents like SOP and LOR's. Thus, the predicted output gives them a fair idea about their chances of getting an admit to a particular university

Citation:

1. <https://www.qwiklabs.com/focuses/1157?parent=catalog>
2. <https://www.qwiklabs.com/focuses/1156?parent=catalog>
3. <https://www.qwiklabs.com/focuses/1158?parent=catalog>
4. <https://www.kaggle.com/mohansacharya/graduate-admissions>
5. https://github.com/GoogleCloudPlatform/data-science-on-gcp/tree/master/03_sqlstudio