Executive Summary

The dataset we considered in the term project paper is collected from Kaggle and it is regarding the prediction of Graduate Admissions from an Indian perspective. The dataset was built with the purpose of helping students in shortlisting the universities with their profiles. The dataset contains 500 rows which are reduced to 486 rows (after pre-processing) and 8 columns and with each of these columns representing an attribute which is considered important during the application for the Graduate Program. They are:- GRE Score, TOEFL Score, University Rating, SOP, LOR, CGPA, Research, and Chance of Admit.
In this paper we performed 5 visualizations on the dataset, and we have gathered the following insights:

- In the first visualization, we considered TOEFL, GRE scores along with the CGPA to predict the average Chance of admit. In this visualization we have observed that the dark blue color bubbles represent higher chance of admit and as the blue shade of the bubble gets lighter the chance of admit also gets decreased.
- In the Scatter plot we have considered the ratings of LOR and SOP to depict the chance of the admit. Based on the analysis of the visualization the plot shows that if the students have good LOR and SOP ratings, then he/she is more likely to be accepted into the university. From the visualization, it is clear that darker the color of the circle the higher chance of admit.
- In the tree-map visualization all the predictors are taken into consideration to analyze the chance of admit like CGPA, University Rating, GRE score, LOR, Research, SOP and TOEFL Score. The Visualization divided the data into 5 categories associated with all the attributes and represented each category with a color.
- Moving on to the next visualization, we have observed that there is a linear rise in the CGPA scores as the university rating increases.
- In the next line chart visualization, we considered average SOP and average LOR ratings as the attributes to predict the chance of admit and we observed that average SOP score shows a steep rise as compared to the average LOR score to secure an admission into better universities.

The three hypothesis we used for our analysis are:

1. How important are GRE, TOEFL scores and CGPA important in getting an admit?
2. Does higher rating of SOP and LOR lead to greater chance of admit?
3. What are the chances of getting an admit from the university considering all variables (i.e. GRE Score, TOEFL Score, University rating, CGPA, Research)?

The models built for the hypothesis were KNN, Regression Tree and Multiple Linear Regression. The best model was chosen based on higher R2-score and lower SSE, MSE and RMSE values. Based on these model evaluation metrics we found the best model to be multiple linear regression for all three models.

Upon completion of our analysis, we have stated some concerns about the dataset such as its size, which is not sufficient to give a real-time analysis to be used as an accurate prediction and also addition of an attribute which improves the effectiveness of our responses. We have concluded that most of the students are taking admissions into tier 2 and tier 3 universities.