Analyzing Universities in the United State

Team 7

Dataset

The dataset for this project is obtained from us-news.

The dataset contains aggregated metrics such as average SAT scores, for universities in the United state. The dataset contains other information such as the tuition and acceptance-rate and the data has been verified not to contain significant outliers.

Research Questions

1. Number of public universities vs private universities per state
   This allows us to find out states which have higher public universities to help students choose from them as they would cost less than private universities.

2. Distribution of Tuition and enrollment in universities across the states.
   The aim is to have an overview of how expensive the tuition of universities is in the United States. Also, we aim to determine the relationship between the tuition fee of a university and the number of enrolled students in the university.

3. Discover which state has universities with the lowest tuition.
   It would be useful to identify the state in the USA that has universities with the lowest tuition, as it would help economical students focus their applications within a particular state.

4. Identify top ranking universities where a good percentage of their students receive aid.
   For economical students with the aspiration to study in top ranking universities, identifying universities with high ranking where a good percentage of their students receive aid would be helpful.

5. Cluster universities into groups based on the average SAT scores of the students.
   The aim here is to create groups where the average SAT scores of the students in the university is what determines the cluster. This would help determine the influence of the SAT scores when compared to another metric like tuition.

Plan

In accomplishing this project, we plan to create tables and graphs to explain the dataset. We will also create a geographic map to visualize the location of the universities. To create the map, we would be using geodata from us-cities. Graphs and tables would be created using python and I would be creating the map using the python folium package.

Additional Dataset used is United States cities databases which has city name and ID, state name, latitude, longitude etc.

Reference

Additional Dataset: https://simplemaps.com/data/us-cities

EDA- https://towardsdatascience.com/exploratory-data-analysis-8fc1cb20fd15

EDA- https://www.stat.cmu.edu/~hseltman/309/Book/chapter4.pdf

EDA- https://plotly.com/python/2D-Histogram/

https://riptutorial.com/matplotlib/topic/3266/basic-plots

https://matplotlib.org/

https://seaborn.pydata.org/

https://python-visualization.github.io/folium/