

College Search Data

Authors:

Masaki Nawa, Ryuichirou Sonoda, Austin Yu

Project Goal:

We are writing a program that provides recommended lists of various colleges based on the user's input such as their SAT/ACT score, the location of the college, percent admitted, Tuition and fees, and the total enrollment. We will be giving out names of different colleges that could be the safety schools for the user.

Description of data:

- We would be using a dataset, "College Admissions" from Kaggle.com (<https://www.kaggle.com/samsonqian/college-admissions>)
- The data will be a csv file.
- Rows of the data:
 - o Name (of the college)
 - o SAT Critical Reading 75th percentile score
 - o SAT Math 75th percentile score
 - o ACT Composite 75th percentile score
 - o Estimated undergraduate enrollment, total
 - o Percent admitted - total
 - o Total price for in-state students living on campus
 - o Total price for out-of-state students living on campus
 - o State abbreviation
- Note: this dataset contains 110 rows of different variables, but we will only be using the variables listed above.
- We would not be using the 25th percentile scores for the SAT and ACT because our goal is to generate safety schools for the user. 75th percentile scores provide accurate results of whether the user has a chance of getting in or not.

Description of algorithms:

We first categorize the size of the colleges into three groups: small, medium, and large. These will be determined based on the number of undergraduate enrollments. We then filter out the colleges based on the given size that the user wants to attend. Next, we will sort the SAT scores from lowest to highest after we add the math score and the critical reading score. We will also sort the ACT scores from lowest to highest. We then filter out colleges that have less than or equal to the user's score(s). After that, we filter out the remaining list of colleges by taking out colleges that are more than or equal to the given amount of money their family can afford. Once we get a new list, we sort the colleges based on the preferred location and percent admitted. If there are more than ten colleges in the final list, we return the first ten colleges.

Outcomes:

- Preferred outcome:
 - We get a list of colleges that meets the size of the college they want to enroll, standard of the user's SAT score and amount of money that can be afford. In addition, the final list is sorted based on the preferred location.
 - If the list has more than ten colleges, display the first ten.

- Satisfying outcome:
 - A procedure that returns a list of colleges.
 - Display ten list of colleges.

- Reach outcome:
 - Since we have the tuition and fees data for in-state residents and out-state residents, we can make a procedure that takes an input where the user lives and display the in-state price for the in-state colleges.
 - We could make a procedure where the user can search for a specific college and add some colleges with similar SAT/ACT score and percent admitted rate.