The Team (we should probably include potential roles here. Like research, debugging, coding, visualization, and that sort of thing)

Jatin Patel

Nimisha Patel

Margaret Ulrich

Problem Statement a.k.a Decompose The Ask

(Is there a correlation between the school ranking vs. acceptance rate, tuition fees, SAT/ACT Score requirements? (This question is not answered with our hypothesis.))

Does ranking necessarily correlate to acceptance rate, tuition fees, SAT/ACT score requirements? (This is a question we actually approximately answer, so is more suited to be our ask.)

Identify Data Sources

<https://www.kaggle.com/theriley106/university-statistics>

Data source will be from Kaggle.com. Kaggle.com has a data set with university ranking data of about 300 universities taken from U.S. News.

Data is available in CSV and JSON format from Kaggle.com.

(Do we need additional sources, as a supplement? I’m sure what we have is good enough, but why not add additional research into the top five, say. Not sure where to obtain additional data. It’s only a thought.)

Define Strategy and Metrics (I think strategy and metrics should be here, along with our hypotheses. I’ve added what I think would work.)

Hypothesis 1: There is a higher concentration of engineering schools in SW vs. NE and vice-versa, there is a higher concentration of Business schools in NE vs. SW. To investigate this, we intend to visualize the relationships of Business and Engineering with their locations.

Hypothesis 2: There is no correlation between the cost of attending highly ranked college vs. lower ranked college. To investigate this, we hope that a scatter plot of cost of tuition to ranking of school will demonstrate that this is not a relevant correlation.

Hypothesis 3: Higher ranking colleges do not always require higher SAT/ACT scores. To do this, a simple scatter plot of college rank by SAT and college rank by ACT will suffice to demonstrate that the data either does point to a correlation or proves it does not.

Hypothesis 4: NE colleges are more expensive than any other region. To show this, a heatmap of tuition-cost by location of college will be sufficient to prove the hypothesis.

Description of Data Analysis Tools You Plan to Use

We plan to use Python’s PANDAS, MATPLOTLIB, NUMPY, gmap and other modules to extract and analyze the data.

We will use PowerPoint to present the data findings.

Describe the Data Products Your Project Will Produce

We will be using various statistical and analytical methods to illustrate our findings. Each of our hypotheses will require at least one visualization, as outlined above.