SI 206 Final Project Plan

a. What is your group's name?

AltEdu

b. Who are the people in the group (first name, last name, umich email)?
Leonardo Carneiro Lindemberg - leolind@umich.edu

Wilson Chen - wilchen@umich.edu

c. What APIs/websites will you be gathering data from? The base URLs for the APIs/websites must be different for them to count as different APIs.

We will be using the Department of Education's Scorecard API to gather data about educational institutions and the GeoDB Cities API to gather data about cities around the world.

d. What data will you collect from each API/website and store in a database? Be specific.

We will be collecting data from educational institutions from the Scorecard API and data about different cities from the GeoDB API. For the Scorecard API, we will be fetching data on each institution, like number of students, acceptance rate, average SAT scores, graduation rate, etc. For the GeoDB API, we will fetch data on the city's population, state, and elevation. Both sets of data will be stored in a SQLite database on different tables. We will be focusing on data from US cities and institutions only.

e. What data will you be calculating from the data in the database? Be specific.

We will be calculating the number of university students in each city by performing a summation on institutions that share the same city within the Scorecard API. For our other visualization, we will calculate the average SAT score of students in all schools in a city using the Scoreboard API and performing SQL joins.

f. What visualization package will you be using (Matplotlib, Plotly, Seaborn, etc)?

For visualizations, we plan to use Plotly. It allows us to project our data onto a map of the United States. It also allows us to create more conventional visualizations like scatter plots and bar charts.

g. What graphs/charts will you be creating?

We will be creating a map visualization that projects the university student population to city population ratio onto a map of the United States. Another visualization that we will make will be

a scatter plot comparing the average SAT score of students who got accepted into institutions in a city to the city's elevation.

h. Who is responsible for what? Please note that all team members should do an equal amount of programming and total work.

The project consists of 3 parts:

1. API call functions
2. Storing retrieved data in database tables
3. Data calculation and visualizations

Each part of the project will be split in half for each team member. Each team member will write and test one of the API call functions and store the necessary data into database tables. This database will be shared between us, so there may be merge conflicts. To avoid this, we will either work on the database section together or we can work on it sequentially, with one of us finishing that part before the other. Lastly, each team member will create one of the two visualizations by making use of the API call functions we have and performing the proper data calculations for aggregate data. By splitting the work in this manner, each team member will be able to interact with the project on all of its aspects and the workload is even.