SI 206 Final Project Plan

1. **What is your group's name?**
    a. Analyzing the Impact of Temperature on COVID-19 and Flu Trends in Michigan and Nationwide
        i. Research question: How does temperature affect flu or COVID-19 cases in Michigan?
2. **Who are the people in the group (first name, last name, UMich email)?**
    a. Ashley Xu, ashleyxu@umich.edu
    b. Tharron Combs, thacombs@umich.edu
3. **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.**
    a. FluView | Delphi Epidata API
        i. Base URL: https://api.delphi.cmu.edu/epidata/
        ii. Endpoint: /fluview
    b. COVID-19 Statistics API
        i. Base URL: https://covid-api.com/api
        ii. Endpoint: /reports
    c. Web Services API (version 2) Documentation | Climate Data Online (CDO)
        i. Base URL: https://www.ncei.noaa.gov/cdo-web/api/v2/
        ii. Endpoint: /data
4. **What data will you collect from each API/website and store in a database? Be specific.**
    a. Delphi Epidata API: We will collect the number of daily flu cases in Michigan and nationally.
        i. Number of daily influenza-like illness (ILI) cases in Michigan. Number of daily ILI cases nationally.
        ii. Weighted ILI percentage (`wili`) for both Michigan and national levels. Metadata such as `epiweek` (epidemiological week), `region`, and `num_patients` (total patients).
    b. COVID-19 Statistics API: We will collect the number of daily COVID-19 cases in Michigan and nationally.
        i. Number of daily confirmed COVID-19 cases in Michigan.
        ii. Number of daily confirmed COVID-19 cases nationally.
    c. Climate Data Online (CDO) API:

i. Daily maximum temperature (`TMAX`) and minimum temperature (`TMIN`) in Michigan.
ii. Daily average temperature (`TAVG`, if available).
iii. Metadata such as `station_id` and `location`.

d. Data Storage: We will store in tables
i. `flu_data` table with the columns: `date, region, wili, num_ili, num_patients, epiweek`.
ii. `covid_data` table with the columns: `date, region, confirmed`,
iii. `climate_data` table with the columns: `date, location, tmax, tmin, tavg, station_id`.
iv. Shared Keys:
1. `date`: Used to join data across all tables.
2. `region`: Used to join region-specific data (e.g., Michigan) between flu and COVID-19 datasets.

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

a. We will be investigating possible correlations between the average air temperature in a given region (Michigan or nationally), flu cases in the given region, and COVID-19 cases in the given region.
b. Data to investigate:
i. Average air temperature (TAVG) in Michigan and daily flu cases (num_ili).
ii. Average air temperature (TAVG) in Michigan and daily COVID-19 cases (confirmed).
iii. Daily flu cases and daily COVID-19 cases in Michigan.
c. Calculate daily, weekly, and monthly averages (as necessary) for:
i. Flu cases (`num_ili` and `wili`).
ii. COVID-19 cases (`confirmed`).
iii. Air temperature (`TMAX`, `TMIN`, and `TAVG`, or manually calculate `TAVG` if missing).
d. Trend analysis:
i. Analyze seasonal trends in flu and COVID-19 cases relative to temperature changes.

ii. Identify peak periods for flu and COVID-19 cases and compare them to temperature anomalies.

e. **Statistical Analysis**. Use statistical tests to analyze data:
    i. Linear Regression:
        1. Linear regression lets us quantify the relationship between a continuous variable (e.g., temperature) and another continuous variable (e.g., flu or COVID-19 cases)
        2. This test allows us to model how changes in temperature directly influence the number of cases and test for statistical significance.
        3. E.g. finding how a 1°F increase in average temperature may affect the daily number of flu cases in Michigan
    ii. ANOVA:
        1. ANOVA is useful for comparing flu or COVID-19 case counts across distinct groups, such as temperature ranges (e.g., cold, moderate, hot) or seasons.
        2. This test can determine whether differences in disease cases are statistically significant between these temperature categories.
        3. E.g. Investigating if flu cases are significantly higher during colder temperature periods compared to warmer periods

6. **What visualization package will you be using (Matplotlib, Plotly, Seaborn, etc)?**
    a. We will use Matplotlib to create scatter plots and line charts such as:
        i. Scatterplot for: "Correlation between daily flu cases and air temperature in Michigan."
        ii. Line chart for: "Trend of weekly COVID-19 cases over time."
7. **What graphs/charts will you be creating?**
    a. We will create visualizations of:
        i. Number of Michigan COVID-19 Cases vs. Air Temperature
        ii. Number of Michigan Flu cases vs. Air Temperature
        iii. (Extra Credit) Number of Michigan influenza cases vs. the number of Michigan COVID-19 cases
        iv. (Extra Credit) National trends of COVID-19 and flu trends vs. weather
8. **Who is responsible for what? Please note that all team members should do an equal amount of programming and total work.**

a. Ashley Xu: Delphi Epidata API implementation and 1 corresponding visualization representing flu trends.
b. Tharron Combs: COVID-19 API and 1 corresponding visualization representing COVID-19 trends.
c. Shared tasks: Database setup, Climate data API, Statistical calculations for correlation, and final report