Data Analysis Project Proposal

- **Group Member:**
  - Yiqian Zheng & Yixin Wang
- **Project Title:**
  - Exploring the Intersection of Income Levels and Culinary Diversity: A Zip Code Level Analysis in the Boston Metropolitan Area
- **Problem:**
  - The culinary landscape in metropolitan areas often mirrors the socio-economic status of residents. In many major cities in the U.S, the unequal disparities between communities of different economic levels is also reflected in the dining environments of these diverse neighborhoods. For example, in higher-income neighborhoods such as West Hollywood, there is typically a greater array of culinary choices, encompassing Italian, Spanish, and Thai cuisines, as opposed to the comparatively limited options in lower-income neighborhoods like East Los Angeles. This research aims to understand how income level may reflect the diversity of food options in various zip codes within the Boston metropolitan area.
- **Solution:**
  - Through evaluating the correlation between the income level of a region and the diversity of local restaurant types, we can further validate the importance of enhancing accessibility to diverse cuisine cultures in different regions. This, in turn, promotes the integration of multiculturalism within the community.
    - In the project, we assume that diverse types of restaurants provide more accessibility to cultural diversity for local residents.
- **Goal:**
  - By leveraging the Yelp API - Business Search Data for restaurants and American Community Survey data via the United States Census Bureau API, we will analyze the distribution of different types of restaurants and explore the strength of correlation to the income level in Boston on a zip code level.
- **Data Source:**
  - **Yelp API:** We will use the Yelp API to collect data on various restaurants, including their cuisine types, ratings, and locations.
  - **United States Census Bureau API:** We will use 2019's American Community Survey (ACS) data available via the United States Census Bureau API. More specifically, we will collect population distribution data of multiple income levels of every single zip code in the Boston Metropolitan area.

- **Data Analysis Tools:**
    - The analysis will use libraries including Pandas, NumPy, and Matplotlib.
- **Analysis approach**
    - Data Collection:
        - Retrieve restaurant data from the Yelp API, focusing on different cuisine types, ratings, and locations. Retrieve demographic data from the United Census Bureau API, focusing on income levels at the zip code level
    - Data Cleaning:
        - Clean and preprocess the collected data to ensure consistency and accuracy, handling missing values and outliers as needed.
    - Data Integration:
        - Merge the restaurant data from Yelp API with the demographic data based on the city location, creating a unified dataset for analysis.
    - Data Visualization:
        - Generate a heatmap to show the correlation matrix between income levels and the prevalence of different cuisines.
        - Create bar charts to visualize the distribution of different types of cuisines in each zip code.
        - Scatter Plots on Map (folium library):
            - Plot the locations of restaurants on a map, using scatter plots with colors representing different cuisines.
    - Correlation Analysis:
        - Use statistical methods to analyze the correlation between income levels and the diversity of cuisines.
        - Identify specific cuisines that are associated with different income brackets.