Project Proposal

Team Information

- **Team Name**:
- **Team Members**:
  - Yanzhen Chen, chen.yanzh@northeastern.edu
  - Siyuan Chen, chen.siyuan5@northeastern.edu
  - Guo Tang, tang.guo1@northeastern.edu
  - Yifan Wang, wang.yifan17@northeastern.edu

Problem Statement

Overweight and obesity have become pressing public health issues in the United States, linked to a range of chronic conditions like diabetes, heart disease, and high blood pressure. This project will explore weight status among adults aged 20 and older, focusing on normal weight, overweight, and obesity categories. We plan to analyze how different factors such as age, gender, socioeconomic status, and ethnicity influence these weight categories. By identifying key trends and relationships, we aim to provide insights that could support public health initiatives aimed at reducing obesity and its associated health risks.

We plan to use publicly available data from the U.S. government's data portal, Data.gov, which provides extensive datasets related to health and nutrition. Specifically, we will be working with the National Health and Nutrition Examination Survey (NHANES) data to study weight categories (normal weight, overweight, and obesity) among adults aged 20 and over. This dataset contains comprehensive information about weight status along with key population characteristics such as age, gender, ethnicity, and socioeconomic indicators, making it highly relevant for our analysis.

Data Sources and Links:

- **Primary Data Source**: Data from NHANES accessed through the Data.gov platform: Data.gov - National HealthData.
- **Supplemental Information**: NCHS, National Health and Nutrition Examination Survey. For more information on the National Health and Nutrition Examination Survey, see the corresponding Appendix entry at https://www.cdc.gov/nchs/data/hus/hus19-appendix-508.pdf.

Ethical Considerations:

- **Privacy Concerns**: Since NHANES data is publicly accessible and anonymized, it does not contain any personally identifiable information. Nonetheless, we will be cautious and adhere strictly to data usage guidelines to ensure privacy protection.

- **Bias in Data**: Although NHANES aims to provide a nationally representative sample, we acknowledge potential biases due to nonresponse rates or underrepresentation of specific subpopulations. We will take these limitations into account during our analysis.

Goals of Analysis

- **Hypothesis 1:** There is a significant correlation between age and the likelihood of being overweight or obese.
- **Hypothesis 2:** Gender and ethnicity are associated with variations in the prevalence of obesity among adults aged 20 and over.

Success Metrics

- We will measure the **correlation coefficients** between demographic factors and weight status to determine the associations.
- **Statistical significance** of coefficients in regression models will help validate the impact of individual factors.

Data Science Algorithms

We plan to use the following data science algorithms and techniques for our analysis:

1. **Descriptive Statistics**: To understand the basic distribution of weight categories across different demographics (e.g., age, gender, income).
2. **Correlation Analysis**: To identify the relationship between weight status and individual demographic characteristics.
3. **Regression Models**:
    - **Logistic Regression**: To predict the likelihood of an individual falling into an overweight or obese category based on demographic factors.
    - **Multiple Linear Regression**: To assess the impact of continuous variables (e.g., age, income) on weight status indicators like BMI (Body Mass Index).
4. **Data Visualization Techniques**: To create clear visual representations of trends and relationships using histograms, bar charts, and scatter plots.

Division of Labor

- **Yifan Wang:** Mainly responsible for the production and content writing of result reports as well as various documents in the project. Also responsible for code creation and presentation of results at the end of the project.
- **Guo Tang:** Primarily responsible for writing, running, checking, and correcting errors in code. Also responsible for checking the completeness of reports, objectivity of information and correctness of content.

- **Yanzhen Chen:** Primarily responsible for writing code, integrating resources and applying data. Also responsible for the preparation of the presentation of the results and the checking of the final report.
- **Siyuan Chen:** Primarily responsible for the preparation of the final presentation content and assigning work for the final presentation. Also responsible for checking the correctness of the code as well as the completeness and objectivity of the reports.