Project Proposal

1. Name of the Project and Team Members

The project title is "Does Healthy Eating Cost More?". This is an individual project by Stella Yu.

2. What Problem Are You Trying to Solve?

The project is trying to investigate the problem of the price gap between healthy foods and regular foods sold in grocery stores. The project's findings will help consumers determine whether eating healthier will be more expensive and illustrate the price differences between foods labeled as healthy and regular foods.

3. How Will You Collect Data and From Where?

Data will be collected by web scraping food products from online grocery stores such as Walmart. I will use Python libraries requests and BeautifulSoup to perform the web scraping. I will first extract product names, prices, and serving sizes for several common food products. Products with names that contain keywords like "healthy" or "low sugar" will count as healthy foods, while other products that do not contain those keywords will be seen as regular foods. The prices will be standardized by weight or serving size for fair comparison.

4. What Analysis Will You Do and What Visualizations Will You Create?

For the data analysis of this project, I will use Pandas and NumPy to perform the data analysis process and create data visualizations. I will analyze the data by calculating the price difference between healthy and regular foods. I will also calculate the average price of healthy foods and regular foods to compare them with each other. Additionally, I will analyze the type of food that tends to exhibit the most significant price difference between the two. For the Visualizations, I will create bar charts to illustrate the price differences between the two categories. In addition, I will also use box plots to show the price distribution in each group.