1. What is the name of your project and who is in the team? Please describe it as a research question and provide a short description.

The name of my project is “Does Healthy Eating Cost More?” This is an individual project by Stella Yu. Email: syu35435@usc.edu Student ID: 3420585291

The research question of the project is “Are food products with healthier nutritional profiles more expensive than less healthy alternatives?” 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.

Using publicly available food and price data, the project explores whether items classified as healthier by Nutri-Score ratings tend to have higher prices than less healthy options. The goal of the project is to examine whether a potential price gap exists between healthy food and regular food. The project focuses on commonly purchased food categories, such as breakfast cereals, breads, and sweet snacks.

2. What data did you collect? How did you collect it? How many data samples did you collect?

The data collected for this project includes both nutritional information and price data for grocery products. The nutritional data includes product names, brands, barcodes, food categories, nutrient information, and Nutri-Score grades. The price data includes the most recent available price for each product.

The data was collected using the Python requests library by accessing two public APIs. Product and nutrition information were collected from the Open Food Facts API. Price information was collected from the Open Prices API using product barcodes. The data collection process began with first getting food product data by category from the Open Food Facts API, then using each product’s barcode to look up price data in the Open Prices API. Only products that had an available price in the Open Prices API were kept in the dataset.

For the data cleaning process. First, products without available price information or name were removed, since price comparison is the focus of the project. Duplicate products were removed. Unused nutrition values included in the nutrients section were also removed. The cleaned data was then saved into a CSV file for analysis.

In total, about 300 food products were collected, with around 100 products from each food category. The categories were breakfast cereals, breads, and sweet snacks. These products were selected from common grocery categories with enough available data so that the results would reflect foods that consumers buy the most often in their everyday lives.

Compared to the original plan, my project has changed from scraping from the websites of grocery stores to using an API instead. This is because some challenges

were encountered during data collection. I was originally planning to scrape data from websites of grocery stores like Walmart, but I quickly realized that most grocery store websites do not allow web scraping. Therefore, I decided to use Open Food Facts and Open Prices, two publicly available APIs, instead of scraping from websites. The original scope of the project also needed expansion. Therefore, I expanded the analysis to include Nutrition-per-dollar metrics of sugar, fat, salt, calories, and Nutri-score analysis as a standard for whether the food is healthy.

I also encountered challenges when collecting data from the APIs. Many products listed in Open Food Facts did not have price data available in Open Prices. To solve this problem, the data collection script needed to exclude products without price information. Another challenge was that some products had missing or incomplete nutrition data, which required additional data cleaning later. Despite these challenges, the final dataset was large enough to support meaningful price comparisons between healthy and regular foods.

3. What kind of analysis and visualizations did you do?

The analysis compares food prices and nutritional value using the Nutri-Score of the products. Items were sorted into two types: healthier options rated A or B, versus standard ones marked C, D, or E. For balanced comparisons, each price was standardized to reflect the cost per 100g. Outliers of the data were left out to avoid skewing outcomes. After that, averages were worked out for every type of food and score range.

Bar graphs figures were made to display the findings. One figure shows the mean cost per 100g by food type. Another figure compares the average price per 100 grams between healthy and regular foods within each category. Additionally, each category had a figure that showed how prices change across Nutri-Score grades from A to E. Finally, nutrition-per-dollar metrics were calculated, including sugar per dollar, calories per dollar, fat per dollar, and salt per dollar. These figures compare how much nutrition consumers receive per dollar spent on healthy food compared to less healthy options.

The findings reveal that unhealthier food costs more per 100 grams compared to healthier options in nearly all groups. However, unhealthy foods also provide more sugar per dollar and fewer calories per dollar than healthy foods. Therefore, while less healthy food costs more in price by weight, you get more sugar per dollar spent. This pattern helps explain why unhealthy foods may feel more filling or satisfying per dollar, even if they are less healthy.

The initial hypothesis of this project was that healthier foods would cost more than less healthy foods. However, the findings of the project show that healthier food does not necessarily cost more than unhealthy food in terms of price per 100 grams. Even though unhealthy food might be pricier by weight, it tends to contain more sugar

per dollar with fewer calories. For healthier foods, it is less expensive by weight but more expensive in terms of calories. This suggests that food affordability depends not only on price but also on the nutrition consumers receive for their money.

4. Future Work

If given more time, I can improve the project in many directions. One possible direction would be to collect more data for analysis by including more food categories and a greater number of products. This would allow for stronger comparisons and more reliable conclusions about price differences between healthy and regular foods. The project could also be expanded to include price data from different grocery stores to examine whether the price gap varies by store. To achieve this, I would need to find other datasets that contain data from different stores, since it is impossible to scrape the websites of the grocery stores. Future work may also include comparing foods labeled organic or healthy. This can help with identifying whether these labels are associated with higher prices.