Danielle Louie
DSCI 510
Final Project Progress Report

Project Scope

Originally, I wanted to analyze all user wearable devices, with a focus on specifically Apple Watches and Fitbits. However, I decided to focus specifically on Apple Watches and Fitbits only due to availability of data. In addition, I decided to narrow down to five chronic diseases to analyze: alcohol, arthritis, cancer, cardiovascular disease, and sleep. These conditions were chosen mainly due to data availability.

Data Sources

In the script **load.py**, I have two functions: **get_csv** and **get_chronic_data()**. The first function **get_csv** is used to extract data obtained in the csv form, which includes data collected from Apple Watches and Fitbit, and returns the data as a dataframe. It is also used to extract the data from the Nutrition, Physical Activity, and Obesity Behavioral Risk Factors file, which also was downloaded as a csv. The second function **get_chronic_data()** is used specifically to acquire data about chronic diseases in the United States, which requires a RESTful (Representational State Transfer) web API. The file is downloaded as an Excel file, and then converted into a csv file. The data acquired are described in the following paragraphs.

Apple Watch and Fitbit Data

The **aw_fb_data.csv** file contains 20 features of data collected from either Apple Watches or Fitbits. Some examples of features that I will be using in my project include age, gender, height, weight, steps, heart_rate, calories, and distance. This file is used to set a baseline for an average person's normal conditions in order to help identify anomalies captured that could indicate chronic disease.

US Chronic Disease Indicators

The **U.S._Chronic_Disease_Indicators.csv** file contains public surveillance indicators related to chronic diseases and their risk factors. Some of the important indicators for this project include health status, sex, age, and location. This data, along with other scientific literature, can help to establish baseline standards for physical attributes or behaviors that might signal early warning signs of chronic health issues.

Nutrition, Physical Activity, and Obesity Behavioral Risk Factors

The data from **Nutrition_Physical_Activity_and_Obesity_-_Behavioral_Risk_Factor_Surveillance_System.csv** contains demographic information mainly related to obesity and weight, but it provides fairly good information as chronic diseases have some positive correlation with obesity. Paired with scientific literature, this file can help provide insight into possible physical indicators for different demographic groups.

Issues / Difficulties

One difficulty that I faced so far was access to the Fitrockr website, which contained numerous data from various user wearable devices types. I tried requesting access to an API key but was told that it would require a 3-month paid subscription in order to receive authentication. Although I tried to reach out several times for other options, I eventually decided that I would not be able to obtain the API key, and I slightly adjusted my focus to only observe Apple Watches and Fitbits, which I was able to acquire a large csv file for.

I also had to find a new data source to replace Health Inequality Data because only data from the Eastern Mediterranean was available, which does not align with my project as my data is more focused on the United States. Fortunately, I was able to find data for Nutrition, Physical Activity, and Obesity Behavioral Risk Factors that contain features related to my project.

I also noticed as I researched more information about chronic illnesses and physical indicators that there were assumptions that needed to be made for this project, such as what is considered a "normal" heart rate and what is an "abnormal" one. In order to understand the data I have better, I have included a list of scholarly articles and literature (from my current progress) that are helping to create the framework for my analysis.

- Kfir Ben-David, et al. "Tracking Cancer: Exploring Heart Rate Variability Patterns by Cancer Location and Progression." Cancers, vol. 16, no. 5, 27 Feb. 2024, pp. 962–962, www.mdpi.com/2072-6694/16/5/962, https://doi.org/10.3390/cancers16050962.

- Malcolm Arnold, J., et al. "Resting Heart Rate: A Modifiable Prognostic Indicator of Cardiovascular Risk and Outcomes?" Canadian Journal of Cardiology, vol. 24, May 2008, pp. 3A15A, https://doi.org/10.1016/s0828-282x(08)71019-5.

- Waldeck, Miriam R, and Michael I Lambert. "Heart Rate during Sleep: Implications for Monitoring Training Status." Journal of Sports Science & Medicine, vol. 2, no. 4, Dec. 2003, p. 133, pmc.ncbi.nlm.nih.gov/articles/PMC3963245/.