Data 375 Project
By: Adian Duarte & Ryan Wixon

By the end of the semester, we hope to have a joined dataset which could allow us to identify if there is a connection between the prices of fruits and vegetables with the rates of obesity (and perhaps even other ailments) of Americans. Seeing as both of our data sources contain information on a state-wide level, we believe the best way to do this would be to generate and analyze statistics on a state by state basis. With just a cursory glance over the initial data, one can find numerous null entries, as well as extraneous data that would only serve to impede analysis. Thus, the project will require us to utilize numerous data cleaning methods in order to get a working result.

Ryan: I am hoping to break into the software engineering field after I graduate. While this class, and by extension the project, is not heavy on advanced programming, I think there are nonetheless ways in which I can showcase my skills. My primary focus will initially be on simply getting a working dataset, but if time permits, I can already think of a few ways to improve/streamline the result. Perhaps I could write a script to automate the process, or even create a (simple) website to display the results in a more interactive manner.

Adian: I want to potentially get into analytics but I am not sure what exactly I want to focus on (I just like to look at data and identify trends). So with this project I want to refine and build on my skills especially when it comes to working with large datasets and being able to manipulate them in order to find an answer. Primarily I want to focus on the data visualization piece and being able to express the data in a form that is easy to read for the end user. In my opinion I feel that just being able to work with and manipulate datasets in general is a huge piece of the job when it comes to the analytics field. And I think practicing and having this skill under my belt will be super beneficial once I join the workforce

Potential Datasets:
https://catalog.data.gov/dataset/alzheimers-disease-and-healthy-aging-data
https://www.kaggle.com/datasets/everydaycodings/produce-prices-dataset