Sidney Julian
Teo Chagil-Mason
William Burke

1. Intro to the Data

- The Nutrition, Physical Activity, and Obesity dataset is available on Kaggle and contains information on several factors related to nutrition, physical activity, and obesity for various states in the United States. The data is collected from various sources, including the Centers for Disease Control and Prevention (CDC), the National Center for Health Statistics (NCHS), the U.S. Census Bureau, and is harvested from the Behavioral Risk Factor Surveillance System.
- The dataset includes variables such as the percentage of adults who are obese, the percentage of adults who engage in physical activity, and the percentage of adults who consume fruits and vegetables. It also includes information of income, age, race, and location.
- The dataset can be used for various analyses related to nutrition, physical activity, and obesity in the United States. It can also be used to identify trends and patterns in these factors across different states and over time. Additionally, the dataset can be used to develop models and predictive algorithms to identify potential risk factors for obesity and related health conditions.
- The population is adults in the United States
- Some sources of sampling error could be small sample size, convenience bias, unknown additional medical factors, or lifestyle influences.

2. Analysis Process and Selected KPIs

- Cleaned data (created columns containing averages)
- In excel we made a pivot table to create a new table with the calculated averages for each state with different variables
- Transformed question column
    - Horizontally split question column to separate different questions into distinct columns.
- Created new tables and connections
- Data cleaning in Excel, imported to Power BI to plan out visualizations
- Made scatter plots and applied trend lines to the plots
- Created a new measure using the Corr(x,y) function to measure the correlation between the three explanatory variables and the dependent variable (Obesity Rate)
- Imported Power BI data to Tableau for final dashboard
- KPI's include:
    - % of adults with obesity classification
    - % of adults who achieve at least 300 mins/week of moderately intense activity or 150 mins/week of vigorous-intensity training

- % of adults who have no leisure time physical activity
- % of adults who consume fruit less than once a day

3. **Selected Visualizations**
    - (insert dashboard here)
    - Scatter plot of 300 min/week and obesity
    - Scatter plot of spend leisure time being active and obesity
    - Scatter plot of consume fruit more than 1 time a day and obesity
    - Bar chart of Gender / Question
    - Heat map of fruit consumption by state
    - Gender bar chart
    - Race/ ethnicity bar chart

4. **Narrative Summary**

Presentation Outline:

- Basic overview of data
- Biases and population of dataset
- Questions from dataset used in analysis
- Explanation of correlation plots & respective coefficients
- Main recommendation of eating more than one fruit per day
- Explanation and demonstration of remaining visuals and how they interact:
    - Race/Ethnicity bar chart
    - Fruit consumption heat map
    - Gender comparison chart
    - Obesity and fruit consumption by gender chart

Dataset Context:

- Obesity rates vary widely across different states in the US. The dataset includes information on the percentage of adults who are obese in each state, and there is a significant difference between the states with the highest and lowest rates.
- Physical activity levels also vary by state. The percentage of adults who engage in regular physical activity varies from state to state, and there is also a difference in activity levels between men and women.
- Fruit and vegetable consumption is lower than recommended in many states. The dataset includes information on the percentage of adults who consume fruits and vegetables on a regular basis, and the numbers suggest that many adults are not meeting the recommended daily intake which is highly correlated to obesity.