Erin Quense

Springboard Data Science Career Track
Capstone 2 Project Proposal
Obesity in America

1. **Problem Statement:** Between 2011 and 2016, what factors impacted the prevalence of obesity in America?

2. **Context:** Healthy people 2020 is a health initiative, restated every 10 years, with the goal of helping Americans improve their health and well-being. One of the objectives is to decrease the prevalence of overweight and obesity, known risk factors for chronic disease, in the population. To do this, public health officials need to know the driving forces of overweight and obesity and where to target their prevention and maintenance efforts and would therefore like to know which factors have the largest impact on overweight and obesity outcomes.

3. **Criteria for Success-** Build models that will predict the incidence of obesity based on risk factors.

4. **Scope of Solution Space**: Obesity trends and impacting factors in America from 2011 to 2016.

5. **Constraints:** none observed

6. **Stakeholders**: public health officials

7. **Data Sources:** This data set includes national and state specific data on adult's diets, physical activity, and weight status, with socioeconomic factors including income, education, race, age, and gender. https://chronicdata.cdc.gov/Nutrition-Physical-Activity-and-Obesity/Nutrition-Physical-Activity-and-Obesity-Behavioral/hn4x-zwk7

8. **Approach:** Using unsupervised learning, determine the national overall prevalence of obesity in 2011 versus 2016. Then, look for patterns in the dataset between diet, physical activity, socioeconomic factors and overweight and obesity. Supervised learning will be used to train a model to predict the incidence of obesity based on risk factors.

9. **Deliverables:** Github repository containing notebook for each step, a slide deck, and a project report.