Course: DSC 680 – Applied Data Science
Project Name: Obesity and Poverty in the United States

Proposal and Data Selection – Milestone I

Data sources

This project is very special. I have been working on it for about six months now, and I am very excited that I will professionally complete it. I have used government statistics because national governments are often the only institutions with the resources to collect comprehensive social statistics, and thus publish many social statistics available. There are two datasets that I will examine. The primary data has not been generated by surveys, interviews, and experiments. It is a normal dataset that was generated from Data.gov, and it is designed for understanding and solving the research problem at hand. The secondary data truly follows the definition of a real secondary data. It is generated by census.gov. It will serve as supporting data for the project.

- Data source: https://catalog.data.gov/dataset/national-obesity-by-state
- Data source: https://www.census.gov/topics/income-poverty/poverty/data/tables.html

Approach

I will use R and Power BI for this project. I am not very comfortable using R, but I am sure that its applications are very well aligned with this type of projects. I will use ggplot which gives me a coherent way to produce visualizations by expressing relationships between the attributes of data and their graphical representation. I will use R Markdown to allow me to create documents that serve as a neat record of your analysis. And finally, I may use Dplyr for data manipulation if I need it.
I will also use Power BI for the following reasons.

- Power BI offers a wide range of custom visualizations. That means visualizations made by developers for a specific use. Custom visuals are also available on Microsoft marketplace.

- In Power BI, I also have the option to upload and view my data in Excel. I will be able to select, filter, and slice data in a Power BI report or dashboard and put it on Excel. I will be able to open Excel and view the same data in tabular form in an Excel spreadsheet.
- Another reason why I will use Power BI is because the data visualization tool is very attractive, intuitive, and interactive. It is very easy to create and understand data through visualizations in Power BI.

Explain the Datasets

As I mention above, there are two datasets that will be used and analyzed in this project.

1. The first data is called "National Obesity by State". This dataset will be analyzed to determine the states that have the greater obesity rate in the United States.
2. The second data is called "Poverty by State". This dataset will be used almost the same way as National Obesity by State. The goal of this dataset is to use it to prove that poverty plays a big role in obesity in the United States.

Data Analysis and Research Questions

- Are poverty and obesity associated?

  Poverty rates and obesity were reviewed across all 51 states including Puerto Rico in the U.S. In contrast to international trends, people in America who live in the most poverty-dense counties are those most prone to obesity. Counties with poverty rates of greater than 35% have obesity rates 145% greater than wealthy counties.
- How is poverty linked to obesity?

I will prove that individuals who live in impoverished regions have poor access to fresh food. Poverty-dense areas are oftentimes called "food deserts," implying diminished access to fresh food.

I will also prove the evidence of the association between sedentariness, poor health, obesity, diabetes, other metabolic diseases, and premature death. Sedentary individuals move 2 hours per day less than active individuals and expend less energy, and they are thereby prone to obesity, chronic metabolic disease, and cardiovascular death.

- What can be done to solve the obesity problem?

  I will create focus on predictive analysis to pinpoint a diverse set of physical activities and nutritional programs in neighborhoods and educational institutions may be the best way to reduce the risk of obesity among poor youngsters. I will also some such that need reliable funding from governments and charities to boost and leverage valuable programs in neighborhood centers and schools serving disadvantaged Americans.

Project Risk

Because the datasets that I am using are very vague, it is very easy to be off task. I mention off task because I will have to extend my datasets to analyze the years. I know that analyzing the years will be a little difficult because based on my initial research of the years they do not match accordingly to each state.

1. I will try not to miss very important features during the process.
2. I will make sure that my predictions are based on the comparisons between the years and between the states.

3. I will make sure to concentrate three major areas of the dataset - High risk states, Low risk states, and Average risk states.

References:

1. https://www.niddk.nih.gov/health-information/health-statistics/overweight-obesity#estimation
2. https://catalog.data.gov/dataset/national-obesity-by-state
3. http://cs.carleton.edu/cs_comps/0910/netflixprize/final_results/knn/index.html
4. https://www.health.harvard.edu/staying-healthy/obesity-in-america-whats-driving-the-epidemic
5. https://www.sciencedirect.com/topics/computer-science/decision-trees
6. Food Service Guidelines Federal Workgroup. Food Service Guidelines for Federal Facilities. Washington, DC: U.S. Department of Health and Human Services, 2017. https://www.cdc.gov/obesity/downloads/ guidelines_for_federal_concessions_and_ vending_operations.pdf (accessed July 24, 2019).
7. "Operation Live Well." Health.mil, Military Health System. https://health.mil/Military-Health-Topics/Operation-Live-Well (accessed July 24, 2019).
8. American Academy of Pediatrics. "Promoting Food Security for All Children." Pediatrics, 135(5): e1431–e1438, 2015. https://pediatrics.aappublications.org/ content/136/5/e1431 (accessed July 24, 2019).
9. Reference: "The relationship between childhood obesity, low socioeconomic status, and race/ethnicity: Lessons from Massachusetts," Childhood Obesity, December 2015, Vol. 11, No. 6.
10. Additional authors: Robert Rogers; Taylor F. Eagle; Anne Sheetz, M.P.H.; Alan Woodward, M.D.; Robert Leibowitz, Ph.D.; MinKyoung Song, Ph.D.; Rachel Sylvester; Nicole Corriveau; Eva Kline-Rogers; Qingmei Jiang and Elizabeth Jackson, M.D., M.P.H.