*Causes of BMI*

1. **Abstract (10 pts)**

Obesity is a growing public health problem in today's society. According to the National Heart, Lung, and Blood Institute, obesity is a chronic health problem that raises the risk of heart disease which is the leading cause of death in America. This report examines factors that contribute to obesity according to one's Body Mass Index (BMI). Body Mass Index determines if one has a healthy body weight in relation to their height. BMI categorizes underweight as a BMI less than 18.5, normal weight between 18.5 and 24.9, overweight between 25 and 29.9, and obese as greater than or equal to 30. Through the use of data analysts, statistical tests, visuals, and data models this project helps to determine what causes high BMI resulting in ways to combat high BMI. This will be done through RStudio taking advantage of their tools and functions to analyze a data set over possible causes for obesity and high BMI. Some findings were some correlations to BMI, for example, the amount of high caloric intake and one physical activity. These variables will be explored along with many others.

1. **Introduction (20 pts)**

According to the National Heart, Lung, and Blood Institute, 3 in 4 adults aged 20 or older are obese or overweight. 1 and 5 Children between the ages of 2 and 19 also find themselves to be overweight or obese. Obesity creates a high risk of heart disease and heart disease being the leading cause of death in the United States this is a problem the public health should know about and be given solutions to combat it. This report will use data from patients and investigate relations as to why people have a high body mass index or a low one. The goal is to give the public ideas and actions to help keep their BMI levels down and stray away from being categorized as overweight or obese.

These goals will be found through data finding, data cleaning, analyzing the data, visualizing the data, and communicating the data. The usage of Exploratory Data Analysis will be used to find the best variables that have a good correlation to one’s BMI. This is done through Correlation tests, T-tests, and ANOVA tests. Within RStudio visualization tools will be used to communicate the data and results of some of the strong relations to BMI levels. The usage of models and their prediction will created throughout the process of our EDA. The models will be used to help predict BMI using variables from the data.

1. **Dataset (30 pts)**

The data set for this was posted by Fatemeh Meharparvar on Kaggle. The data set was based on an individual's eating and health habits within Mexico, Peru, and Columbia. The data set contains clean data to help predict one’s obesity levels. It consisted of 2111 lines of clean data with 17 columns to base prediction possible predictions. The columns given were the patient's age, weight, if they consumed high-calorie meals, the amount of physical activity they did daily, how many veggies and water they consumed, what type of transportation they used daily, their obesity level, if they monitored their calories, and if there was any family history of being overweight. BMI levels were not a part of the data but I was able to include them by calculating BMI levels with the height and weight of each row. Not much went into this data set on my end, it was already clean and well-maintained by Meharparvar.

1. **Exploratory Data Analysis (EDA) (60 pts)**

*Rmd and PDF are provided for this part*

1. **Models (75 pts)**

*Rmd and PDF are provided for this part.*

1. **Summary of learning (5 pts)**

Some of the most interesting findings I found doing this project were the factors/variables that caused and did not cause high BMI. Veggies consumed, water consumed, and if a patient eats high-calorie meals all had a strong correlation with BMI. These three variables represent the importance of what someone puts in their bodies is prevalent. Another variable correlated with BMI was if a patient monitors their calorie intake. All four of these variables illustrate a healthy diet/lifestyle is key to maintaining a normal/healthy BMI. A surprising variable that was not correlated with BMI was the amount of meals one eats daily. This could give proof that what one eats is more important than how much one eats. If one eats 5 meals a day consisting of chicken and vegetables it would be better than eating 3 meals a day consisting of fatty foods. Overall, I feel more informed about what is correlated with BMI than before I started this project.

This project taught me many things and was a great experience for my knowledge of the realm of data science. I got to understand how the data process works firsthand. I made many mistakes, for example, using a correlation test for numerical and binomial data. Building models for me was the most challenging. Understanding how they worked and what they did took time to gain. I also could not figure out how to make models better at predicting BMI. It took me extensive research, time, and using trial and error to make good models. It was a process of making mistakes and learning from them. I had to learn through reading articles and textbooks I found online. I found myself using the world's new Google, ChatGPT, to learn how to build visuals and understand topics. This was the first time I have used ChatGPT for a major project and it was life-changing. I did not ask ChatGPT to do this project for me but asked it to teach me how to do it. With ChatGPT I was able to learn on the fly and ask specific questions Google could not answer or would take me an extended amount of time reading through articles to answer.