Team 7 Project Proposal

Our Project is an Exploratory Data Analysis project focusing on a major area of concern in community Healthcare: Heart Disease. We are exploring various factors that potentially have an impact on whether a patient suffers from a heart attack and a stroke. Our team found a large dataset, “Indicators of Heart Disease (2022 Update)”, from the website Kaggle. This dataset consists of 2022 annual CDC survey data of over 400,000 adults from all 50 states and has categorical and numerical data on a wide array of factors of a patient’s health status. This dataset was published by Kamil Pytlak and has a CC0:Public Domain license.

Our main question in this EDA project is: *Does the region of the United States in which a patient lives play a part in their risk of suffering from a heart attack and a stroke?* To explore this question, we broke down the dataset into the 4 US census regions: South, West, Midwest, Northeast. For each region we will explore what correlations gender, ethnicity, general health, smoking status, alcohol use, sleep, weight, and BMI impact the likelihood of suffering from a heart attack and/or a stroke. We will explore which factors are likely more significant with respect to heart attacks and strokes.

References:

[Indicators of Heart Disease (2022 UPDATE) (kaggle.com)](https://www.kaggle.com/datasets/kamilpytlak/personal-key-indicators-of-heart-disease/data)

<https://github.com/kamilpytlak/data-science-projects/blob/main/heart-disease-prediction/2022/notebooks/data_processing.ipynb>