This project and dataset are interesting to me because they can reveal how different health outcomes cluster within populations. Understanding these clusters is highly relevant in a public health context and it can be useful in health interventions. This project is a perfect combination of my major in data science and my minor in public health.

This project analyzes health metrics to identify communities with similar health risk profiles through graph-based clustering techniques. The application processes health-related data by normalizing certain metrics, such as BMI, blood pressure, and cholesterol levels to ensure consistency. Each of the nodes in the graph represents an individual's health data. I use a clustering algorithm to detect communities within the graph, grouping individuals with similar health characteristics. I then use silhouette scores to see the effectiveness of these communities, which measures how similar an individual is to their cluster compared to others. This analysis helps in understanding the clustering quality and in potentially tailoring public health interventions to specific groups.

Running the project:
To run this project, I first found the path to my project by using cd path/to/project in the terminal. Once in the correct path, use cargo build to compile everything and finally use cargo run to execute and print the project outputs.

Simplified description of what my output contains:
First, it normalizes the data to standardize health measures. It then constructs a graph with each node representing a person's health indicators, simplifying network-based research. After that, the program analyzes the graph created. First, it ensures each node is unique and linked correctly to produce the correct graph. Then it finds graph communities using graph-based clustering. People with similar health qualities form each community. Each community has role model representatives depending on various criteria. An individual's silhouette score indicates how well they fit into their chosen community. The overall community structure is shown by the average silhouette score. This project does vertex degree analysis and cluster analysis to assess the health data network's connectivity.