**Executive Summary**

*This project explores how social factors like income, education, and race relate to health outcomes. The aim is to identify the communities most at risk,so health-related efforts can be focused where they’ll have the biggest impact. The project uses a mix of public datasets to highlight which groups are most affected and where support is most urgently needed. There may be challenges in getting clean, complete data and ensuring the results aren’t skewed by cofounding variables.*

**Motivation**

*Health isn’t distributed equally, and where people live (along with their background and resources) can heavily influence how long and how well they live. By looking at socioeconomic patterns tied to race, education, and income, I hope to uncover where the system is falling short and where change could save the most lives. This topic feels personal and timely, especially given the gaps in care many communities face today.*

**Data Question**

*Where should we prioritize health interventions to save the most lives- based on income, education, and race disparities?*

**Minimum Viable Product (MVP)**

*The goal is to produce a visual summary (possibly an interactive dashboard) that shows where disparities are most severe. It will highlight the geographic areas where targeted intervention would make the most difference and break down the contributing factors. The result will make it easier for decision-makers to see where to act and why.*

**Schedule (through <date of demo day>)**

1. Get the Data (7/25/25)
2. Clean & Explore the Data (8/10/25)
3. Create Presentation of your Analysis (8/15/25)

* Should be a presentation, but could include a Jupyter Notebook or dashboard in Excel, Tableau, or PowerBI

1. Internal demos (8/25/25)
2. Demo Day!! (8/29/25)

**Data Sources**

* *Publicly available health and demographics data (CDC, Census Bureau, County Health Rankings, etc.)*
* *Regional statistics on health by ZIP code or county*
* *Social vulnerability metrics*

**Known Issues and Challenges**

* *Some areas may have incomplete or outdated records*
* *Health outcomes are complex and tied to many overlapping factors*
* *May need to align datasets with different boundaries (e.g., ZIP vs. county)*
* *Accessing certain databases might require credentials*
* *It’s possible the available data won’t reflect more recent events or trends such as COVID19*