Milestone 1 - Project Proposal

Topic:

- Name: Assessing the Impact of Chronic Diseases on COVID-19 Mortality in the U.S.

- Description: This project investigates the correlation between chronic diseases and COVID-19 mortality rates across various states, age groups, and racial/ethnic demographics from 2020 to 2023, aiming to identify significant risk factors and inform public health measures.

Business Problem:

The COVID-19 pandemic has disproportionately affected different population groups, particularly those with pre-existing chronic conditions. Understanding how these conditions influence COVID-19 outcomes can help optimize resource allocation, improve public health responses, and guide preventative measures. This project seeks to answer critical questions such as:

- Which chronic conditions have the strongest association with COVID-19 mortality?

- How do COVID-19 outcomes vary across different age, racial, and geographic groups when chronic conditions are considered?

- What patterns can inform future interventions for high-risk populations?

Datasets:

- COVID-19 Deaths by State and Age (2020-2023): This dataset captures the underlying conditions contributing to COVID-19 deaths, segmented by state and age groups, to assess regional impact and age-based vulnerability.

- U.S. Chronic Disease Indicators: A dataset providing detailed insights into chronic disease prevalence, behavioral risk factors, and healthcare access across states, enabling a nuanced analysis of risk factors.

- COVID-19 Cases and Deaths by Race/Ethnicity: This dataset offers a breakdown of cases and deaths by racial/ethnic groups, adding a crucial dimension to the analysis and allowing exploration

of disparities and social determinants of health.

Methods:

- Descriptive Analysis: Basic statistical measures will outline trends in COVID-19 outcomes and chronic disease prevalence across states, age groups, and racial/ethnic categories.
- Correlation Analysis: To understand the relationships between chronic conditions and COVID-19 mortality rates.
- Regression Models: Multivariate regression models will be used to identify significant predictors of COVID-19 mortality, controlling for confounding variables such as age, sex, and socioeconomic status.
- Data Visualization: Comprehensive charts, maps, and infographics will be created to clearly communicate trends and findings.

Ethical Considerations:

- Privacy: Although data is likely aggregated and anonymized, there is a need to ensure compliance with privacy standards to prevent any potential identification of individuals.
- Bias & Fairness: There is a risk of reinforcing stereotypes or biases when interpreting data related to race, ethnicity, or socioeconomic status. Results will be framed responsibly, highlighting systemic factors rather than attributing outcomes to individual characteristics.

Challenges/Issues:

- Data Quality & Completeness: Datasets might have missing values or reporting inconsistencies that could affect analysis accuracy. Techniques like imputation and sensitivity analysis will be considered to manage these gaps.
- Confounding Variables: Factors like healthcare access, socioeconomic status, and regional healthcare infrastructure might confound results, making it challenging to isolate the impact of chronic conditions on COVID-19 outcomes.
- Generalization of Findings: Results from specific states or demographic groups may not be

generalizable to the entire U.S. population, requiring careful interpretation.

References:

The project will rely on data from the Centers for Disease Control and Prevention (CDC), along with academic literature on the intersection of chronic diseases and COVID-19 outcomes. Key references will include peer-reviewed studies on COVID-19, chronic disease epidemiology, and public health interventions aimed at mitigating pandemic impacts.