**Executive Summary**

**This project aims to analyze the trends and patterns of COVID-19 cases in a specific region, using available data from public health sources. The motivation for this project is to understand how the pandemic has impacted the local population and to identify any potential risk factors that could contribute to the spread of the virus.**

**The main data question driving this project is: “What are the key factors that have influenced the spread of COVID-19 cases in the selected region?” To answer this question, we will be using various data sources such as daily case counts, hospitalization rates, economic and demographic information.**

**Assumptions and challenges that may arise during the project include the accuracy and completeness of the data, as well as potential changes in testing protocols or reporting practices that could affect the interpretation of the results. Additionally, the impact of public health interventions such as vaccination campaigns or lockdowns on the spread of the virus will need to be carefully considered and analyzed.**

**Overall, this project aims to provide a data-driven understanding of the COVID-19 situation in the selected region, which could inform public health policies and interventions to help mitigate the spread of the virus.**

**Motivation**

1. **Demographic factors: COVID-19 has been shown to affect different demographic groups in different ways, with older adults and people with underlying health conditions at higher risk of severe illness. I could investigate whether certain demographic groups in Tennessee have been disproportionately affected by the pandemic and explore potential reasons for these disparities.**
2. **Economic factors: The COVID-19 pandemic has had a significant impact on the economy, with many businesses and industries facing closures or decreased demand. I could analyze the relationship between economic indicators (such as unemployment rates or business closures) and COVID-19 case counts in Tennessee and explore potential reasons for any observed trends.**

**Data Question**

*Present your question. Feel free to include any research/articles that are relevant or show where others have attempted to answer this question.*

**Minimum Viable Product (MVP)**

1. **Data Collection and Cleaning: The first step in this project will be to collect and clean data on COVID-19 cases, hospitalizations, deaths, and demographic information in Tennessee. This will involve sourcing data from various public health agencies and cleaning it to ensure consistency and accuracy.**
2. **Data Exploration: Once the data has been cleaned, I will explore it using descriptive statistics and visualizations to identify trends and patterns in the spread of COVID-19 in Tennessee. This will involve creating maps, charts, and graphs to help us understand how the virus is spreading and which regions may be most affected.**
3. **Hypothesis Testing: Based on our data exploration, I will develop several hypotheses to test potential factors that may be contributing to the spread of COVID-19 in Tennessee. This could include factors such as demographic characteristics, economic indicators, public health interventions, and geographic location.**
4. **Data Analysis: Using statistical analysis techniques, I will test my hypotheses and explore the relationships between different variables in the dataset. This will involve conducting regression analysis, correlation tests, and other statistical methods to identify significant relationships between variables.**
5. **Presentation of Findings: The results of the data analysis will be presented in a clear and concise report, which will include visualizations and tables to help convey my findings. I will also provide recommendations for future research and potential policy interventions that could help slow the spread of COVID-19 in Tennessee.**

**Schedule (through)**

1. **Get the Data (6/1/2023)**
2. **Clean & Explore the Data (6/15/2023)**
3. **Create Presentation of your Analysis (6/20/2023)**

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

1. **Internal demos (6/22/2022)**
2. **Demo Day!! (6/29/2022)**

**Data Sources**

1. **<https://www.tn.gov/health/cedep/ncov/data/downloadable-datasets.html>**

* **This data set contains all covid-19 cases listed by counties in TN. It has data dates from 2020-2023. In this data I will be able to focus on county specific and year specific which is 2021 in Davidson county, TN and categorize them by age group. The link to this dataset is** [Age by County](https://www.tn.gov/content/dam/tn/health/documents/cedep/novel-coronavirus/datasets/Public_Dataset_Daily_County_Age_Group.CSV)

1. [**https://censusreporter.org/profiles/05000US47037-davidson-county-tn/**](https://censusreporter.org/profiles/05000US47037-davidson-county-tn/)

* **Sex by Age (year 2021) Data** [**https://censusreporter.org/data/table/?table=B01001&primary\_geo\_id=05000US47037&geo\_ids=05000US47037,04000US47,01000US**](https://censusreporter.org/data/table/?table=B01001&primary_geo_id=05000US47037&geo_ids=05000US47037,04000US47,01000US)
* **Race and Ethnicity Data** [**https://censusreporter.org/data/table/?table=B03002&primary\_geo\_id=05000US47037&geo\_ids=05000US47037,04000US47,01000US**](https://censusreporter.org/data/table/?table=B03002&primary_geo_id=05000US47037&geo_ids=05000US47037,04000US47,01000US)
* **Economics Data (Income)** [**https://censusreporter.org/data/table/?table=B19001&geo\_ids=05000US47037,04000US47,01000US&primary\_geo\_id=05000US47037#valueType|estimate**](https://censusreporter.org/data/table/?table=B19001&geo_ids=05000US47037,04000US47,01000US&primary_geo_id=05000US47037#valueType|estimate)
* **Poverty Data** [**https://censusreporter.org/data/table/?table=B17001&primary\_geo\_id=05000US47037&geo\_ids=05000US47037,04000US47,01000US**](https://censusreporter.org/data/table/?table=B17001&primary_geo_id=05000US47037&geo_ids=05000US47037,04000US47,01000US)
* **Household Data** [**https://censusreporter.org/data/table/?table=B11002&primary\_geo\_id=05000US47037&geo\_ids=05000US47037,04000US47,01000US**](https://censusreporter.org/data/table/?table=B11002&primary_geo_id=05000US47037&geo_ids=05000US47037,04000US47,01000US)
* **House Marketing Data** [**https://censusreporter.org/data/table/?table=B25075&primary\_geo\_id=05000US47037&geo\_ids=05000US47037,04000US47,01000US**](https://censusreporter.org/data/table/?table=B25075&primary_geo_id=05000US47037&geo_ids=05000US47037,04000US47,01000US)
* **Educational Attainment Data** [**https://censusreporter.org/data/table/?table=B15002&primary\_geo\_id=05000US47037&geo\_ids=05000US47037,04000US47,01000US**](https://censusreporter.org/data/table/?table=B15002&primary_geo_id=05000US47037&geo_ids=05000US47037,04000US47,01000US)

**Known Issues and Challenges**

* *Known Issues that I came across was being able to find data that would correlate with my hypothesis. I am trying to analyze data during the pandemic in the year of 2021. Analyzing how the county was affected by the pandemic specifically looking closely at demographics and economic standpoint during covid-19.*
* *Issues that I could come across during my project can be seeing how I can correlate completely different datasets from two websites to accomplish my goal.*