Mitigating the Next Wave:   
A Comprehensive Analysis of Global COVID-19 Policies

### **Introduction**

In the face of an impending challenge, our primary objective is to analyze and evaluate the efficacy of diverse COVID-19 policies adopted worldwide in curbing the growth rates of cases and fatalities, while focusing on two assigned metrics – "Public Trans - recommend closing" and "Testing policy - anyone showing symptoms.” This comprehensive analysis will serve as the foundation for crafting a policy plan tailored to Caladan's unique needs, ensuring the safeguarding of its population of 3.2 million from the resurgence of COVID. As we embark on this critical mission, our collective expertise will be instrumental in shaping a policy plan that not only addresses the immediate concerns of the Commonwealth but also establishes a resilient framework for future challenges. Through meticulous analysis and strategic planning, we aim to empower Caladan with a data-driven and adaptive approach to navigate the complexities of the ongoing pandemic. While COVID reporting has limitations given lack of testing for those who are deceased, attributing cause of death to COVID-19-related complications, processes for declaring deaths and causes of deaths, and lack of transparency, we hope to find some insights.

### **Data Architecture**

To forge a comprehensive dataset for analytical purposes within Azure Data Factory, we strategically unified data from three disparate sources: On-Premises SQL Server, Azure CosmosDB (SQL API), and Azure SQL Database. This intricate process, encompassing ETL, transpired in Azure Data Factory, a robust service tailored for managing large-scale data integration tasks. The datasets, originally sourced from a variety of sources including the WHO, CDC, Public Health Departments, include a diverse set of variables, including pivotal metrics like confirmed cases, deaths, and intricate policy measures. These datasets encapsulate not only the quantitative facets of the pandemic, such as economic responses and health metrics, but also qualitative aspects, such as policy implementations and regional geographic nuances.

COVID-19 Metrics from On-Premises SQL Server and Azure SQL Database:

Among the key variables from these sources are the geographical specifics, including regions within countries, latitude, and longitude. These variables enable us to discern localized patterns in the spread and impact of the virus over time. The metrics of confirmed cases, deaths, and recoveries provide a nuanced understanding of the pandemic's toll on different regions.

Azure CosmosDB (SQL API) - JSON Data:

Incorporating a multitude of policy-related variables, such as school closings, workplace restrictions, and public service/event cancellations, this dataset illuminates the diverse strategies implemented by countries to curb the virus. It also highlights the effectiveness of different levels of testing policies countries experimented with throughout the course of the pandemic. Lastly, the inclusion of economic response variables, such as income support and fiscal measures, provides insights into the broader socio-economic implications of the pandemic.

The harmonization of these datasets into the .parquet format underscores our commitment to efficient and performant analytics. This format, optimized for Hadoop Distributed File System (HDFS) underlying Azure Data Lake Storage Gen2, ensures that the data is not only easily accessible but also amenable to scalable and high-performance analyses.

Through this meticulous consolidation, we have laid the groundwork for a data-driven approach to pandemic analysis. The organized and unified dataset not only meets the immediate needs of the Commonwealth of Caladan but also positions itself as a valuable resource for other countries facing similar challenges. As we embark on the journey of automation and source control, the Azure Data Factory framework ensures that the dataset remains up-to-date and adaptable, ready to tackle the ever-evolving landscape of the ongoing global health crisis.

### **Analysis and Recommendation**

Both of our recommendations are closely related as enabling one enables the other. We highly recommend keeping testing at its highest level so that all residents can get access to open testing. Additionally, we want to ensure that public transportation remains open so that people can access their testing sites. We want to emphasize the fact that as city density increases, testing degrees should be more strict. For example, in our analysis based off of South Korea, we examined that even though the population density was quite high, due to the fact that they had high testing levels, their cases remained relatively low. While all countries mandated some sort of testing, the highest effectiveness came from the ones that implemented the greatest levels of testing which makes sense since more confirmed cases allows for further actions to be taken. We want to ensure that we can mirror these results for Caladan, so we will be following the same guidelines as them.

Looking into our data and our key influencers, we were able to quickly summarize what key factors led to changes in our assigned metrics. We found that the likelihood of not closing public transportation increased by 1.52x when the country was South Korea or when population density was more than 326 people/km2; on the latter, when pop. density was between 195 and 326 people/km2, the likelihood of not closing public transportation was magnified 1.10x. This makes a lot of sense because if we look at denser countries people rely on their public transportation more since there isn’t a lot of space for roads and cars. South Korea has a population density of about 500 km2 which aligns to why they did not close public transportation especially since there were many testing centers available in order to implement strict testing policies. By implementing these policies, South Korea had one the least accumulated confirmed cases, low fatality rate (2%), and a high recovery rate (83%) which demonstrates that South Korea was able to effectively handle the COVID-19 pandemic . If a country with a smaller density were to implement a similar strategy as South Korea, we would assume they would not view the same results as population density would greatly affect the policies’ effectiveness. Plus, each time the density of a country goes up by 175.4, the likelihood of increased testing increases by 1.75x. Most specifically, looking at South Korea, the likelihood of H2 being a 3 increases by 4.27x when the country is South Korea. Again, since we want to mimic South Korea, we would recommend keeping testing at a high level and keeping public transportation open for the civilians in the Commonwealth.

**Challenges**

Our team faced endless challenges during our project, from data engineering to data management. The most difficult portion of the project was finding relationships between different columns and producing a star schema with linkage to external data as well; even on the last day, we found that we still had to make slight changes to our connections in order to draw more detailed insights. Additionally, data management was difficult too since permissions for data lake storage and Azure synapse were unclear at first. And, collaborating with a Mac-based team on PowerBI was challenging as each and every change had to be uploaded to Github. In the future, we would love to work on a similar project but with more computational power!