

## Covid-19 Policy Analysis

Team 21: Jaejoon Kim, Urvi Sharma, Henry Willis, Jamie Lin, Idan Gutkind

7 May 2024

### **Scope of Problem**

The COVID-19 pandemic has hit many of the world's countries hard, demanding governmental responses to reduce the damage done. Caladan, a midsize commonwealth, was not hit hard by the coronavirus, but commonwealth leaders are looking to implement new policies based on the effectiveness of other policies in the past. Our task was to determine the least restrictive policies that keep the growth rate of deaths below 1% and the growth rate of new cases below 3% on a 30-day rolling average.

### **Data Exploration Steps**

We used ELT and determined our solution by extracting, loading, transforming, and then analyzing the given data of global policies implemented by 10 representative countries. The data we need is located in three places: Cosmos DB, Virtual Machine, and Azure SQL DB. We loaded our data to Azure Data Factory, where we cleaned up our data to fit our needs such as deleting unnecessary columns and uniforming the format. We used exploratory analysis to formulate our hypotheses and confirmatory data analysis to validate whether or not the initial hypotheses were accurate.

### **Exploratory Data Analysis**

For exploratory analysis, our team created slicers for countries, policies, and policy levels in order to see how all three impact the growth rate of cases and deaths over a 30-day rolling average. We used a line chart with a threshold at 0.01 (1%) for deaths and 0.03 (3%) for cases to

depict what changes in policy/policy level will give us a growth rate for both under the threshold. While selecting different policies and levels for our countries, we looked at where the cases and death growth rates are falling/minimized below the thresholds. We noticed strong negative trends between policy implementation and growth rates after the spike of COVID-19 to the beginning of July, and using this time frame, we continued our exploration. We also imported external data for the country's population. Based on our knowledge of Caladan (population of 3.2 million and two urban population centers in Duncan and Stillgard), we concluded that it is similar in population density to that of South Korea, Japan, and France. Therefore, we continued our analysis by focusing on the effects of policies in South Korea, Japan, and France.

### **Recommended Policies and Reasoning**

We analyzed and compared the death and cases growth rate of several different policies in the time frame of April 1, 2020 to June 1, 2020. We chose this time frame because it was after the spike of COVID-19, therefore policies will have time to be implemented. We also saw low deaths and cases growth rates within this time frame consistent amongst policies. External research was also conducted to help form our hypothesis. For example, during our observed time frame, Japan reopened schools which could explain the initial fluctuation in death rates for certain policies. Through our EDA, we've hypothesized that school closing, facial coverings, and restrictions on gatherings are the best recommended policies that kept cases growth rate under 3% and death growth rate under 1%.

### **Confirmatory Data Analysis**

We sought to confirm our findings by using statistical evidence. The key-influencers visualization in Power-BI selects the top policies that had the highest impact on deaths and cases. So, it showcased the relationship between policy level and the change in cases/deaths. We

identified the negative relationships that were the strongest by standardizing (dividing the growth rate of deaths and cases by policy growth rate) and compared them with our initial hypothesis. The graph confirmed that school closing is effective, because an increase in policy C1 (school closing) led to a decrease in case growth rates for Japan and Korea, and it also had the highest level of decrease within that period. The same also confirmed that facial coverings were a strong key influencer in reducing deaths in France and Korea.

### **Conclusion**

After analyzing our statistical evidence, we decided to reject our initial hypothesis that restriction on gathering is an effective policy but going forth with school closings and facial coverings. We found that the school closing and facial coverings policy kept deaths and cases growth rates below the desired threshold, while also functioning at a low restrictive level

## **Appendix: Team Contributions**

**Henry:** Created storage account and containers to start the project. Helped deploy data factory and configure the Self Hosted Integration Runtime to extract data from the virtual machine.

Fixed schema issues when importing data. Created Data Warehouse and loaded data into Synapse Analytics. Performed EDA to highlight policies relevant to Caladan. Researched reasons for outliers during our time frame.

**Urvi:** Sourced data from Cosmos DB and converted it into Parquet in the data lake. Assisted with data flow/pipeline. Created visualizations to help analyze data.

**Jamie:** Assisted in determining the 30-rolling average. Helped analyze data through visualization.

**Jaejoong:** Configured Azure Data Factory settings to extract data from external sources into a Data Lake, ensuring efficient data availability. Developed and normalized an Operational Data Store (ODS) in Azure. Designed schema in Power BI, analyzing and visualizing data and preparing it for big data processing in Azure Synapse.

**Idan:** Created draw.io diagrams for architecture and schema. Setup GitHub and Virtual Machine. Worked with the rest of the team to extract and load all of the data into Azure. Helped analyze and choose best policies given our data visualizations.