A Critical Analysis of COVID-19 Policy Effectiveness
CDS DS 310 Data Mechanics

During the global COVID-19 pandemic, Caladan officials wanted to develop strategies to mitigate the spread of the virus and minimize its impact on public health and the economy. Our team started a comprehensive data engineering project to collect, process, analyze, and visualize data from various sources to achieve this. The ultimate goal was to recommend effective policies that could keep the growth rate of deaths below 1% and the growth rate of new cases below 3% on a 30-day rolling average, keeping the level of restriction to a minimum.

We extracted the raw data from 3 sources: a Cosmos database containing policy data, an on-premise Virtual Machine containing half of the case, country, recovery, and death data, and an Azure SQL database containing the other half. Through Azure Data Factory manipulation, we created dataflows that converted the raw data into parquet files that we loaded into a data lake. Each source was assigned a designated container in the data lake for efficient management. To transform the raw data, we created an Operational Data System (ODS) via a separate container located in the data lake.

Having made the ODS, we loaded the processed data via Azure Synapse Analytics, creating External Tables in the Synapse Serverless Pools to connect Fact and Dimension files stored in the lake. We subsequently loaded the data into Power BI for exploratory data analysis to determine the most effective COVID-19 policies. To create our schema, we used the policy data as our fact table and then connected it via the country code (ISO3) and data columns. To avoid many-to-many relationships, we created a new column that combined ISO3 and date. To begin the analysis, we created measures that quantified countries' growth and death rates, in addition to their respective 30-day rolling averages. Having visualized this, we determined the most significant period during the pandemic to be between mid-March and mid-August when cases rose drastically and then started to drop as policies began to roll out. We chose this volatile period to analyze which policies were effective in the reduction of the growth rate for both cases and deaths.

Next, we tried to narrow down which specific countries' policies we were interested in analyzing. We decided to analyze Korea, Italy, and New Zealand since they were the only countries throughout our selected time period that had an average case growth rate and deaths growth rate below 3% and 1%, respectively.

In order to analyze the different policies to make our recommendations, we first normalized all the policy data on a scale of 0 to 1, ranked by level of restrictiveness, to eliminate any potential biases during the analysis. Next, we graphed the growth rate of cases and deaths with a legend for each policy's restrictiveness level to see if the value for each given policy correlated to a change in growth rate. In particular, we focused on policies that had fluctuating results over different periods of time to highlight the impact the level of restriction had on the growth rate among the nations. In light of this, policies such as closing public transportation and contact tracing were not picked for our recommendations, since they had zero or very little fluctuation.

The exploratory data process was a catalyst in narrowing down which policies we wanted to recommend and further analyze.

Recommendations

Moving into our recommendations, we focused on the three main policies:

1. Moderate/low stay-at-home restrictions
2. Moderate Enforcement of Testing tracing
3. Moderate Restriction international travel controls

Each of these policies saw fluctuating levels of restrictiveness among the three selected countries, reflected varying degrees of enforcement level during the chosen time period, or was a combination of the two. While the stay-at-home restrictions set in place by Italy, Korea, and New Zealand lowered its restrictiveness to 0.33/1 or 0/1, the growth rates continued to remain below the 3% and 1% thresholds for the cases and deaths. To further investigate, we used the built-in Power BI key influencers visualization that revealed lowering stay-at-home restrictiveness to a level of 0.33 or less was 11.58 times more likely to cause the death growth rate to be below 1%. The function also suggested that maintaining the level of restrictiveness for the stay-at-home policy to between 0 and 0.66 was 1.34 times more likely to keep the case growth rate to be below 3%. Having yielded these statistically significant results, we concluded that while maintaining this enforcement level did not cause lower rates, there is no evidence to suggest that it caused the growth rates to surpass their ideal thresholds.

The other two policies we focused on were testing policy and international travel restrictions. While these two policies were more restrictive than the stay-at-home, we believed the variance and the drop in restrictiveness of the policies over time and among the countries would uncover interesting trends within the data. Although Korea was very strict with their implementation of the testing policy at a 1/1 level, both New Zealand and Italy were less strict with a 0.66 level. Based on the key influencer visualization we found statistical evidence that a less restrictive testing policy with levels of 0.33 and 0.66 was 1.25 times more likely to have a death rate below 1%, compared to more restrictive enforcement. Because of this, we concluded that lowering the restrictiveness of the testing policy to a moderate level would be a crucial step in keeping the death rate below 1% for Caladan.

While New Zealand was stricter with their international travel restrictions, Italy and Korea contrastingly enforced lower levels of restrictiveness with levels of 0.75/1 and quickly converged to 0.5/1 and even 0/1 for some time. This variance allowed us to analyze how modifying the enforcement levels in a rapidly changing environment impacted the growth rates of cases and deaths. The Key Influencers Analysis found that international travel restrictions kept the case's growth rate below 3% 1.6 times more with 0/1 enforcement, which we concluded was an outlier as Italy had a 0.1 restrictiveness level for around one week. However, enforcement of 0-0.5/1 also proved effective in keeping the growth rate below 3%, as it did not cause the growth rate to increase significantly 83.70% of the time.

Although our Power BI mainly focused on these three policies we also analyzed all the other policies with significant variance and volatility. For example, we saw that the workplace-closing policy and restrictions-on-internal-movement policy each were able to have a lower restrictiveness level while keeping our growth rates below the required level.

In conclusion, we found that analyzing policies with variance over our selected time period (3/11/2020 - 8/11/2020) allowed us to determine how reducing the restrictiveness level of each policy impacted if growth rates for cases and deaths were below 3% and 1% respectively. We were able to accomplish this by exploring each policy and relying on two points of criteria: sufficient variance over the time periods for the selected countries (Korea, Italy, and New Zealand), and whether the growth rates were still below the required levels for the majority of the time. We found this method successful in analyzing the different policies, and are confident that our recommendations are the strongest choice for Caladan to implement.

Appendix:

Policy Appendix: Throughout this project, we did not factor in the economic/financial-related policies as most of these policies only occurred a few times throughout the entire time period. In addition, we decided to not consider any of the health policies past h5 since the Consolidated project details document did not have a description for those policies.

Contributions:
Joshua Leeds (Project Manager): Joshua started the ELT process by extracting the data from the SQL database and loading it into the data lake with the data factory. He also contributed to and completed the data flow to load the data into the ODS. He completed the visualizations, DAX Commands, and data analysis in Power BI. Finally, Joshua wrote parts of the final report, specifically the data analysis sections.

Parin Shaik (BI Analyst): Parin orchestrated the ELT process by extracting data from the Cosmos database and loading it into the data lake. She initiated the dataflow that transformed and merged the data, loading it into the ODS. Parin also created the slideshow, architecture, and schema diagrams, and contributed to the creation and analysis of Power BI visualizations. Finally, Parin analyzed the data and prepared the write-up.

Anmar Abdi (Data Engineer): Anmar deployed the data factory and configured the SHIR. He contributed to the ELT process by extracting the data from the VM. He crafted the initial data flows during the transformation and mapping stages. He also performed the data cleaning and created the schema within Power BI. Finally, he created the initial Power BI measures and helped with visualizations.

Jimin Park (Data Architect): Jimin transformed the raw data into data lake storage and loaded the data, by guaranteeing the accuracy and efficiency of our data transfer processes. Leveraging insights from data analysis, she provided recommendations that guide our policy effectiveness. She also initiated the paper by crafting the introduction and outlining our objectives. Finally, she assisted the Power BI visualization, and presentation slides.