Mason Carlucci, Moses Chen, Sviatoslav Shevchenko, Freya Zhu
Professor Seferlis
DS 310
5/7/24

DS 310 Team 24 - Executive Summary

Determine an effective, yet unrestrictive COVID-19 policy response for Caladan
As a team of researchers, the government of the small country of Caladan has tasked us to aid its response to the COVID-19 virus. Caladan is a country with two major population centers (Duncan and Stillgard) and 3.2 million people. Given a dataset on how ten nations reacted to COVID, we must construct a modern data pipeline and use PowerBI to answer the question: What are the most unrestrictive policies they can implement to keep the growth rate of deaths below 1% and the growth rate of new cases below 3% on a 30-day rolling average? By utilizing our given data, we must determine the most effective policies concerning accomplishing our goal and recommend the policy level that Caladan should implement.

Complete an exploratory data analysis concerning our goal
Beginning with raw data tracking policies, countries, time, and the impact of COVID-19 on the population, the first step was to manipulate and collate our data. Using processes in Microsoft Azure, we transformed and moved our data to PowerBI, where our team could create a schema and analyze the data.

To build our schema, we connected our data with cases, deaths, and recoveries with data concerning dates, countries, and policies. We constructed our dashboards on PowerBI to begin analyzing the data we had worked with, focusing on 30-day averages relating to cases, deaths, and recoveries, as outlined by our instructions. Once this information was accessible, we chose the most restrictive policies and constructed our dashboards to display how the policies impacted the population in our ten given countries. The crux of our investigation is to determine which policy has the greatest proportion of days spent under the thresholds of 1% for the growth rate of deaths and under 3% for the growth rate of new cases on a 30-day average.

Definitions of the statistics used in terms of business and statistical understanding
The implementation of statistical methods supporting our Power BI report was a crucial step in validating the conclusions to be made from the Power BI dashboard visuals. To this regard, our team chose regression analysis as a suitable method for validating the effectiveness or lack thereof of a given policy, post-factum the presentation of dashboard visuals for it. Two Time-Series Regression Analysis dashboards (Cases and Deaths) were created, incorporating the regression graph, graph feature selection slicers (policy, country, date time frame), and a statistical module graph with statistical measures like R^2 and correlation coefficient.

Policy recommendations to facilitate the minimization of COVID-19 impact

To find the most effective policies to minimize new cases and deaths, we examined the growth rates of the new cases and deaths when enacting a policy at a certain level. Afterward, we gathered the number of appearances of minimized growth rates from March 24, 2020, to March 30, 2021. In doing so, we counted the remaining days below the desired thresholds and the total number of days the policy was in effect. With these values, we created a proportion of days below the threshold to quantify the effectiveness of a certain policy at a certain level of restrictiveness. Next, we compared the proportions of the percent change of deaths and percent change of new cases by the specific policy level. We concluded that the most effective policies are school closing at level 2, workplace closing at level 1, and international travel controls at level 3. Among the policies, these three had the most frequent appearances, with the highest proportions below the desired levels.

Use of auxiliary data to improve research and analysis

Through researching external articles regarding the efficacies of certain policies, we discovered that our findings were consistent with those of others' research. In an NIH article titled "Which COVID policies are most effective? A Bayesian analysis of COVID-19 by jurisdiction", they found that the three most effective policies are closing workspaces, schools, and staying home. Among these three recommended policies, our three core recommendations had two policy overlaps. The reason for our third recommendation, restrictions on international travel, is that Caladan has avoided a large influx of new cases.

Further opportunities for research and analysis

The primary limitation of this study that could be expanded upon in future research is that every country in the data pool is part of the Global North. This distinction is significant for the rhetorical ambitions of this study, as Global North countries have consistent initial conditions that skew the data of what was studied. In particular, Global North countries such as those analyzed in this study consistently have more developed healthcare systems, higher urbanization rates, and smaller shares of rural populations. These conditions significantly change the societal impacts of a phenomenon such as COVID-19. If Global South countries were involved in the sample, the data would include the effects of having conditions more reflective of the world. For instance, Global North countries took part in large vaccine and health research sharing projects during the pandemic, during which many Global South countries were left out of the loop, leading to systematically longer recovery periods, which was factored into the data pool for this study. As such, the outcomes and analyses from this study can only be reasonably asserted for the circumstances of Global North countries; further research that includes cases in the Global South would be required to arrive at conclusions that apply globally.

Secondly, outside factors that impact how a nation handles COVID-19 are variable. Policies and programs may conflate with others, potentially damaging the overall efficacy of a conglomerate

of several policies. With this, it makes it difficult to determine how specific policies impact the effects of COVID-19 and how policies interact with one another. Similarly, the culture of each nation changes how they respond to a virus and how closely they follow guidelines established by the government. Some cultures are less willing to respect policies, changing how impactful those policies are- compared to cultures that respect policies.

Team Contributions:

Sviatoslav Shevchenko - :

- Challenge 1: Created the team storage account, containers for sources and deployed data factory (Freya set up SHIR)
- Challenge 2: Created ODS and was present at query discussions.
- Challenge 3: Together with Mason, suggested the creation of a new column combining country region and date to facilitate a connection between the Union Table (Deaths and Cases) and Policies Table. Suggested connection of the Policies and Dates table through a common date column.
- Challenge 4: Created two interactive slicer regression analysis dashboards (Deaths and Cases), with statistical figures (i.e. R^2, correlation coefficient, slope equation). Suggested creation of dashboards (y1 - percent change in rolling average of deaths/cases, y2 - sum of policy implementation level, x - dates).
- Deliverables: Created data pipeline and utilized services architecture diagram.

Moses:

- Challenge 1: Landed data into CosmosDB container and SQLDatabase, created pipelines for each
- Challenge 2: Created pipelines & changed mapping. Created and queried tables from each source, landed into ODS
- Challenge 3: Connected ODS into Synapse
- Challenge 4: Existed
- Deliverables: Created several dashboards

Mason

- Challenge 1: Existed.
- Challenge 2: Provided vibes, and assisted with query creation.
- Challenge 3: Together with Sviat, suggested the creation of our schema and strategy to link the tables containing our data.
- Challenge 4: Aided the creation of dashboards and applied analysis to determine the best ways to use our data. Used Power BI dashboards (with the help of Moses) to conclude which policy options were the best, and at what level of implementation.
- Deliverables: Helped out with Power BI dashboards. Wrote and edited a majority of the executive summary. Edited our slides.

- Scheduling meetings and posting When2Meets

Freya Zhu

- Challenge 1: Set up the SHIR and landed data from the SQL Server; set up the respective pipeline
- Challenge 2: Moral support :D and assisted with transforming data by helping search up how to format queries
- Challenge 3: Assisted with the ODS
- Challenge 4: Created the schema and connected the relationships between the tables; created the columns for the rolling average for deaths and new cases; Created dashboards and began the process of using slicers and helped with the creation of the other dashboards
- Took notes at meetings and delegated work
- Created the slidedeck