Katherine Beaty, Primah Muwanga, Gabrielle Thompson, Ruxin Wen

Professor Seferlis

Data Science 310

07 May 2024

Caladan's Prevention Analysis

Caladan's government has significant concerns about the Coronavirus influencing their commonwealth in the upcoming wave. In order to create a policy plan that can tackle this issue, our goal is to research how ten other countries have implemented their own policies, attempting to keep their death rates low and prevent new cases from appearing. Specifically, we want to find any policy that is unrestricted and can prevent the death rate from exceeding 1% with the growth rate also not exceeding 3%. By finding the most efficient policies, Caladan can avoid the harsh costs and conditions presented by the virus.

Before finding a solution, it's crucial to understand what we are working with. The dataset given offers an extensive overview of the COVID-19 pandemic through six collections: Cases, Deaths, Recoveries, Policies, Country, and Dates. Each collection serves a specific purpose: 'Cases', 'Deaths', and 'Recoveries' track daily and total counts of COVID-19 incidents, fatalities, and restoration. The 'Policies' collection records the deployment stages of various governmental strategies, 'Country' offers vital location-specific details, and the 'Dates' collection aligns these data points with exact dates, enabling detailed time-series analysis. This organized framework enhances our ability to deeply understand the progression of the pandemic and the effectiveness of the measures implemented to counter it in Caladan.

To explore our data, we generated many data pipelines and data flows to make the data usable in our final analysis. In the initial inspection, we eliminated any columns that could not be

utilized (such as the nulls) and also added unique SID's(a column comprising country code and each date) to connect our tables, defining our star schema. Now we could focus on exploring the relationships between specific variables, like the deaths and cases in regards to the policies.

We utilized several measures to decide the ranking of the policies for each country. It was determined that our first goal was to identify which policies affected each country. This meant observing each policy and accessing how impactful they were on the corresponding population. A 30-day rolling average metric was used to reduce bias and enable better comparisons by focusing on the overall trends. We incorporated a measure that counted how long the data for each country spanned. To supplement that, another measure was used to track the number of active days (where the policy was utilized regardless of restriction level) that fulfilled our requirement for the desired percentage of the growth and death rate. These measures were based on the policies. As these specific measures assess the entire duration of our data, this becomes useful for long-term analysis, which was needed for our overall goal of finding the most unrestrictive policies. In order to strengthen the helpfulness of the policies, 'Hospital Admissions' external data was added to our overall structure. This additional measure enhanced our results, revealing which policies were more successful in controlling the surge of hospitalizations.

Each country's policy success was measured by the proportion of consecutive days it remained effective, represented as a ratio like 68/430, where the numerator is the number of continuous successful days and the denominator is the total documented days of the policy. By averaging these proportions across all countries according to a specific policy, we determined the overall effectiveness of the policy.

The results became as followed:

POLICIES	School Closing	Workplace closing	Cancel public events	Restrictions on Gatherings	Close Public Transportation	Stay at Home Requirement	Restrictions on Internal Movement	International Travel Control
AVERAGE	**0.400379**	**0.378255**	**0.086096**	**0.370544**	**0.18558**	**0.33306**	**0.309719**	**0.47271**

POLICIES	Income Support	Debt Contract Relief	Public Information Campaigns	Testing Policy	Contact Tracing	Face Coverings	Vaccinations	Protection of Elderly People
AVERAGE	**0.448107**	**0.36711**	**0.479943**	**0.504124**	**0.475279**	**0.402164**	**0.122078**	**0.430479**

Based on our analysis, the most unrestrictive policies are the testing policy with a restrictive level of 1 meaning only those who had had symptoms and met a specific criteria (hospital workers, or came into contact with a known case) were tested. Our second best policy was public information campaigns with a restrictive level of 2 meaning coordinated public information across traditional or social media, as a result, everyone is able to access the same information. Finally the last policy we would recommend is contact tracing with a restrictive level of 1 which means limited tracing and not done for all cases. Each policy keeps the death growth rate below 1% and the growth rate of new cases below 3% on a 30-day rolling average. These three policies had the highest averages among the ten policies.

The external table for hospital admissions played a crucial role in formulating effective policies aimed at maintaining a death growth rate below 1% and limiting the growth rate of new cases to below 3%. This involved analyzing instances where a particular policy proved successful over consecutive days. By examining these periods, one could observe whether hospital admissions remained low or showed a decreased growth rate, providing valuable

insights into the effectiveness of the policy in question.

In regards to future research, it would be insightful to explore the conflicts between different policies and how they could potentially enhance or obstruct the success. It could also be helpful to investigate the responses from the public to see how that impacted the effectiveness of the policies.

Team Deliverables

All members contributed to weekly meetings.

Challenge 1

Katherine Beaty extracted the data from sources and landed it in a data lake using Azure Data Factory.

Challenge 2

All members contributed as a team to transform the data using data flows and data pipelines, with Katherine Beaty and Ruxin Wen doing most of the transformations.

Challenge 3

All members contributed as a team in discussing the design of the schema.

Gabrielle Thompson created the architecture diagram and schema final design in draw.io.

Challenge 4

Katherine Beaty and Primah Muwanga focused on creating the visual analysis on PowerBI.

Gabrielle Thompson, Ruxin Wen, and Primah Muwanga focused on creating the final presentation slideshow and writing the summary analysis.

Important Diagrams