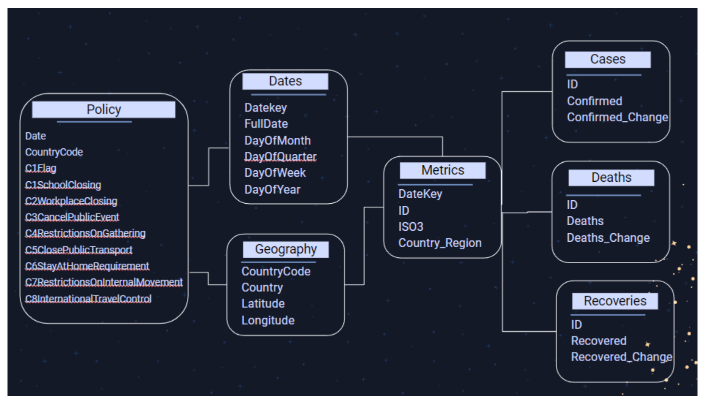
Team 3 Project Executive Summary

In our project, we've been supporting the leaders of the Commonwealth of Caladan in navigating the challenges posed by COVID-19 using data-driven insights. With Caladan concerned about a resurgence of infections, our goal was to analyze and propose health policies that strike a balance between effectiveness and minimal restrictions to ensure that the growth rate of COVID-19 deaths stays below 1% and new case growth remains under 3% over a 30-day rolling average.

To kick off our project, we gathered COVID-19 data from three sources: Azure SQL database, VM SQL Server, and Azure Cosmos DB. Each of these sources contributed information, including policy details and COVID-19 metrics like case numbers, recoveries, and fatalities. After collecting this data we imported it into a Data Lake for processing and analysis. The data transformation phase involved organizing and moving the data into an Operational Data Store (ODS). In this ODS environment, we combined datasets to synchronize information with SQL database records to gain a comprehensive understanding of the situation.



Within our ODS setup, we meticulously organized the data using a Galaxy Schema approach to facts from dimensions. Our schema featured fact tables, for metrics and policy details while dimensional tables were used to categorize dates and geography information related to cases, deaths, and recoveries. This structure facilitated a comprehensive analytical approach, allowing for detailed, multi-dimensional queries that are crucial for robust data analysis. See the diagram for the Schema in detail.

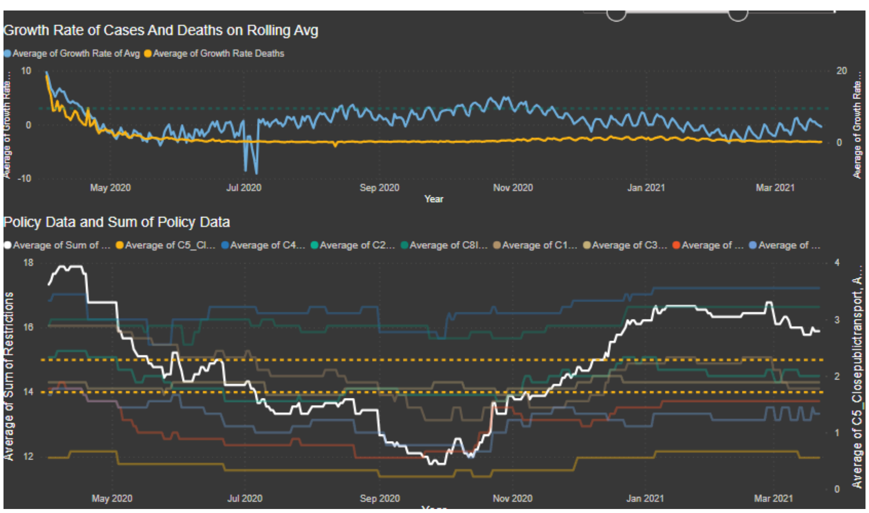
By employing the Apache Parquet format, we loaded seven refined data files to create SQL tables in Azure Synapse following our Galaxy schema rigorously. This setup empowered our team to execute queries and produce reports effectively and involved the creation of the Metrics table, which took information from the original metrics file, as well as Dates.

The analysis concluded with visualizing the data through Power BI, where we designed reports that not only emphasized statistical discoveries but also demonstrated the potential effects of different policy choices. This was done by creating new columns to view the growth rate of the 30-day rolling average of new cases as well as deaths using DAX queries in PowerBI.

During our analysis, we noted erratic data patterns from New Zealand, likely due to delayed reporting of cases and deaths. To maintain the integrity of our analysis, we excluded this outlier data, focusing on more consistent and reliable datasets.

In our research, we define the "**Policy Sum**" as a quantitative measure of the overall intensity of governmental responses to COVID-19. This indicator combines public health measures into a unified measure indicating the strictness of these measures over time. Each aspect of policy is given a value based on how restrictive it is, which influences the total policy sum.

By Utilizing the "**Policy Sum**" index, our analysis identifies what we define as the "**Golden Zone**"—ranges of policy sum values where the measures are effective yet minimally restrictive (see the above graph). Specifically, when the policy sum index is between 14 and 15, we observe optimal outcomes: the growth rates of COVID-19 cases and deaths are maintained below the critical thresholds of 3% for new cases and 1% for deaths, respectively, over a 30-day rolling average, as well as slightly beyond these timeframes. This finding highlights the effectiveness of strategic policy implementation, balancing public health imperatives with economic and social considerations. This was done by examining data, and seeing how deaths and cases were related, (notably that death growth rates correlated with the growth rates of cases, delayed). Using this examination, it was clear that on average, the policies described relate to a growth rate below the thresholds. See the diagram above, where the dotted blue line is the threshold for the cases' growth rate. This shows a clear correlation with restrictions, death, and case rates.



We used the average policy from all countries used in our analysis to form our solution. This was done by averaging different nations, so we could best account for potential variances, due to specific details of Caladan not being provided to us. Our results from our statistical analysis are as follows:

- **School Closings (C1):** Level 2, which requires shutting down specific educational levels or categories, such as high schools or public schools only.
- **Workplace Closings (C2):** Level 2, this measure mandates the closure or shift to remote work for certain sectors or categories of workers, targeting areas with higher risks of transmission while maintaining economic activities elsewhere.
- **Public Event Cancellations (C3):** Level 2, which requires the cancellation of public events to reduce large gatherings, is a crucial step in mitigating virus spread.
- **Restrictions on Gatherings (C4):** Level 3, which limits gatherings to between 11 and 100 people.
- **Public Transport Closure (C5):** Generally set at Level 0, indicating no closures. This policy may be adjusted in regions where public transport is denser and poses a higher risk. Specifics of this information were not provided, therefore, this policy has flexibility to be implemented at Level 0 - Level 1.
- **Stay-at-Home Requirements (C6):** Level 1, this policy recommends, rather than mandates, staying at home, providing a less restrictive measure that still encourages reducing public exposure.
- **Restrictions on Internal Movement (C7):** Level 1, which advises against travel between regions or cities.
- **International Travel Controls (C8):** Level 3, which bans arrivals from specific regions to control the entry of new infection sources into the population.

Further research would be required to fully implement these solutions confidently, as the specific culture, government, economy, and population of Caladan would factor into these policies.