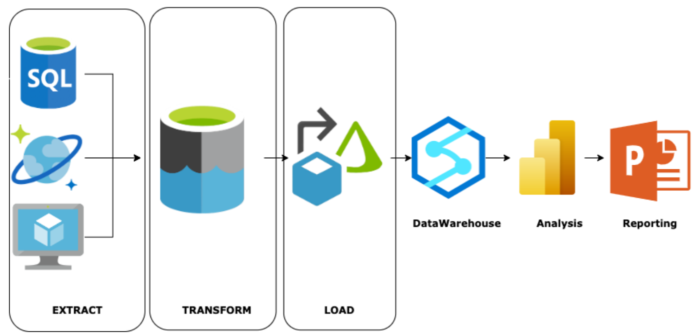
Anna V., Anshi M., Jack M., Kiril Y., Naeco L.

Data Mechanics

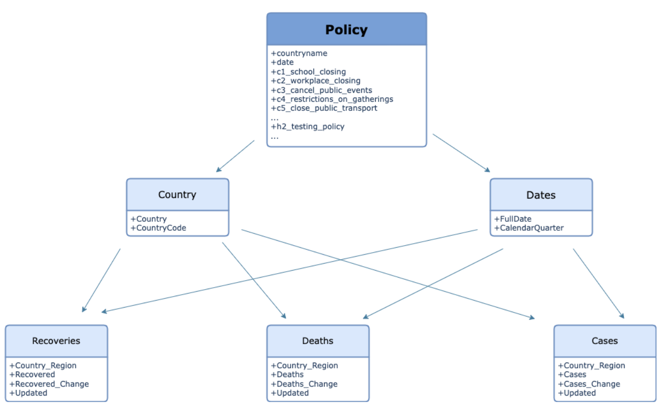
Professor Seferlis

16 December 2023

Team 12 Report



We analyzed Covid-19 data to determine which policies were most effective. We used an ETL structure to pull and aggregate data from multiple sources and organize the dataset based on metrics (cases, deaths, recoveries, etc.) and policies and, using a snowflake schema in PowerBI, we were able to connect and discover relationships between the many variables. We were able to analyze our two assigned policies and which policies were most successful in slowing the spread of Covid-19.



Our two assigned policies were immediately perceived as effective: the required closing of public transportation (c5_2) and open public testing (h2_3). For these two policies, we analyzed Covid data from ten countries: New Zealand, Russia, South Korea, UK, Canada, France, Germany, Italy, and Japan. To analyze these policies we created a line and stacked column chart depicting when a policy was put in place and the new cases on a month by month change.

Through our analysis of policy c5_2 (required closing of public transportation), we discovered Italy and New Zealand were the only countries to elevate the policy to level two–complete closure of public transport. As shown by the line chart detailing the sum of change in deaths and closed public transport over time, after elevating to level two, Italy experienced an immediate drop in deaths, and New Zealand experienced an immediate drop in confirmed cases. This is shown to a lesser extent in countries that only elevated to level one (c5_1). For example, after implementing the recommended closing of public transport, Canada experienced an eventual decline in both change in deaths per day (from 136 on January 8th to 19 on March 24th) and change in confirmed cases per day (8,147 on January 8th to 4,050 on March 24th). This is evidence that requiring rather than recommending the closing of public transport was more beneficial in slowing the spread of the coronavirus.

We found open public testing (h2_3) to be less effective than the required closing of public transportation (c2_3) but still helpful in slowing the spread of Covid-19. An argument for its effectiveness can be made by Germany's case and death trends from November 2020 to March 2021. In November, Germany downgraded the policy from open public testing (h2_3) to only symptomatic testing (h2_1) and witnessed an immediate rise in both cases and deaths, ultimately peaking in January. However, looking at, for example, Russia or South Korea, two countries that had open public testing throughout the Covid-19 pandemic, even with an elevated testing policy of h2_3, cases and deaths still peaked at a similar time to Germany.

To effectively analyze which two policies were most impactful, we first calculated correlation coefficients for 12 policies. From those 12, we narrowed it down to four to further analyze: restrictions on gatherings (c4), school closings (c1), stay at home requirements (c6), and testing policy (h2). We analyzed these four similarly to our two assigned policies, looking at the change in cases over time with a secondary axis depicting the degree of the policy. Ultimately, we discovered c4 and h2 to be superior for slowing the spread of Covid-19. This is depicted through both percentage changes in case numbers (in Canada from January to February of 2021, there was a 55.84% decrease in new cases. At that same time, Canada had a testing policy of h2_2) and correlation coefficients. It's important to note that this disregards policy degree specific analysis. For example, when analyzing our policies, we found that c2_3 was more impactful than h2_2, however, when looking at the broader impact, h2 was more effective in slowing the spread of Covid-19, shown by both the correlation coefficient (0.54 for h2 versus 0.24 for c2) and the percentage change in case numbers.

Reflecting on the project, we're confident in the conclusions drawn and happy with the methods used. Overall, the project had a much broader scope than we originally anticipated. We also failed to anticipate the learning curve involved in big data analysis, especially while facing machine limitations (only two of our members could access PowerBI). The sheer number of policies and the intricacies of their effects, specifically in terms of degree changes, made analysis difficult and time consuming. This is most evident in our intuition based choosing of 12 policies to analyze. We disregarded policies such as public information campaigns based on our own personal ideas of their effectiveness rather than something shown by data.