DS310 Project Executive Summary: Team 18

Overview:

This project was designed so that we could construct a fully functional Modern Data Pipeline utilizing COVID-19 data, then create calculations on growth vs policies enacted by different governments based on this data. This project gave us a sense of some of the problems we need to solve in the real world. We worked on answering the following question: What are the most restrictive 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?

As a brief background, we moved source data from different sources (Cosmos DB, Azure SQL DB, and a Virtual Machine) into multiple blob storages using ADF data copy. We then transformed the data using data ADF flows and loaded it into an operational data storage. We then connected and loaded the data to Azure Synapse Analytics for connection with PowerBI for analysis.

Something we implemented into our project was a UN population dataset, which had country populations, country population densities, median age, and other statistics for all countries. We Imported this data from their downloadable Excel file and joined the UN data with our existing country data. Essentially, it functions as a dynamic link to their locally downloaded Excel file. Our data flow streamlines the information, filtering it to encompass only the data from the year 2020 and genuine countries, verified through their ISO3 codes (excluding regions, etc.), while removing any extraneous columns that aren't part of our intended analysis. We were able to incorporate statistics and visualizations of policies pertaining to countries with varying population densities, forming the basis for our conclusions.

Metrics and Analysis:

For our base metrics of cases and deaths growth rates we first created a 30-day rolling average of new cases and deaths of each country respectively to smooth out the data, and remove potential extremities created from countries batch reporting on certain days. From there we calculated the growth rate of that rolling average to normalize our data between countries of different population sizes, as well as to achieve our goal format of a growth rate to keep under our desired values of 3% for cases and 1% for deaths

To decide which policies would be most effective for us to vary, we first split our metrics data based on each possible value of each policy, and averaged the growth rates over all days and countries where the policy was implemented at that specific value. From there, we took the standard deviation of values for each of the policies, to get a measure of how much changing that policy changes the average growth rates. From this we concluded that C8 (international travel restrictions) was the most impactful policy to vary, with a standard deviation of 0.055 between its 5 possible values for cases growth rates. Beyond this the next largest values were C1, C2, and C3, so we looked to vary those policies as well in cases where just implementing a value of C8 was not sufficient to keep our growth rates at our desired values. After this we did our analysis by finding the value of C8 that minimizes average case and death growth rates for each quarter, then looked at growth rates filtered on that C8 value to find if implementing other policies was necessary.

Policy Recommendations:

We found policies that solved our original problem and broke them up by quarter and country population density. We considered countries with 200 persons per square kilometer as dense populations, and those with less than 200 persons per as sparse populations. Here were the results we found: in Quarter 1 of the year 2020, for countries with a sparse population density, we concluded that International travel restrictions with screening arrivals (C8 = 1) and to not cancel public events (C3 = 0) were the most effective policies. For dense countries in the first quarter of 2020, International travel restrictions was also the best policy, but this time, with quarantine arrivals from some or all regions (C8 = 2). Moving on to Q2 of 2020, international travel restrictions with ban arrivals from some regions (C8 = 3) and recommended school closings (C1 = 1) were two of the most important policies for sparsely populated countries. For dense countries in Q2 for 2020, required cancellations of public events (C3 = 2) was deemed the most significant. In Q3 of 2020, for countries with sparse populations, three policies were found to be significant: international travel restrictions with banned arrivals from some regions (C8 = 3), required cancellations of public events (C3 = 2), and required school closings on some levels (C1 = 2). For the densely populated countries, screening arrivals for international travel (C8 = 1) was an important policy. In the last quarter of 2020 for sparsely populated countries, the two policies that proved the most useful were the banning of all arrivals and closing borders for international travel (C8 = 4), and recommended public event cancellations (C3 = 1). For the denser countries, screening arrivals for international travel (C8 = 1) proved the most significant policy.

Moving onto the 2021 results, in Q1 for countries with a sparse population, the banning of all arrivals for international travel (C8 = 4) and required workplace closing (C2 = 3) were the two most significant policies. Among the dense countries, the three policies that proved the most effective were: banning arrivals from some regions during international travel (C8 = 3), recommended closing in the workplace (C2 = 1), and recommended school closing (C1 = 1).

As we can see, the most recurring policy to implement is international travel restrictions (C8), closely followed with the cancellation of public events (C2) and school closings (C1).

Future Studies:

We can highlight several areas focusing on the long-term impacts of travel restrictions on countries dependent on tourism. Exploring alternatives like virtual tourism and strengthening local industries could help these economies recover and diversify, reducing the impact of global crises.

Studying the resilience of these economies to sudden shocks is also vital. Understanding what helps these areas bounce back could guide future economic policies and support systems. Additionally, the pandemic has significantly influenced travel behaviors and safety perceptions, revealing shifts in preferences between domestic and international travel. These changes could reshape the tourism landscape significantly.

The role of international cooperation during the pandemic, through strategies like travel bubbles and open corridors, also deserves attention. Examining these policies' effectiveness could provide a blueprint for managing future global health emergencies, ensuring better preparedness and resilience.

In conclusion, addressing these topics is crucial for preparing tourism-dependent economies for future challenges and ensuring their long-term sustainability.

Appendix I: Team responsibilities

Challenge 1: Extract and Load
Matthew: CosmosDB Data Source, UN Population dataset
Hal: On premises SQL Data
Keenan: Azure SQL Data

Challenge 2: Transforming Data
Matthew: Metrics Dataflow, integrating UN Data into Countries dataflow
Hal: Countries Dataflow
Keenan: Dates Dataflow
Brenda: Policies Dataflow

Challenge 3:
Matthew: Moving data into external tables in Azure Synapse

Challenge 4:
Matthew: PowerBI Schema design/implementation, visualizations, metric analysis, recommendation implementation of analysis

Challenge 5:
Matthew: Downloading data factory deliverables, visualizations from PowerBI, metrics and analysis section of Executive Summary
Hal: Data Architecture and Data Model diagrams, assisting on Executive Summary, editing recorded presentation
Keenan: Writing Executive Summary
Brenda: Creating Slideshow, future studies section of Executive Summary