**Data Analysis for COVID-19 and Economic Statistics for Countries**

Data is critical to support countries in managing the global coronavirus (COVID-19) pandemic. This ETL project provides an array of real-time data, statistical indicators, and other types of data that are relevant to the coronavirus pandemic. Data is drawn from different authoritative sources which is then cleaned and updated.

* Step1—data is extracted from different sources and downloaded in pyton in its origin format.
* Step2—data is cleaned by dropping irrelevant columns and formatted to meet the requirements
* Step3—data is then merged to form the final tables
* Step4-- - A Python script containing several functions to create plots and get insights from the global dataset.
* Step5-- - queries for .

# ETL Process

# EXTRACT

Information for this report was extracted from a variety of reliable online sources on March 11th 2021. For the economic data provided, it was sourced from ourworldindata.org. The data was freely available in CSV format and was subsequently downloaded for transformation and analysis purposes. This file contained information relating to the economic statuses of each country as of the current date as well as some information regarding COVID statistics of each country listed.

Information regarding COVID specifically was taken from the World Health Organisation’s specific site dedicated to coverage of the virus. The CSV file extracted not only displayed the cumulative cases and deaths so far but also displays the most prevalent transmission classification for each country, giving some reasoning to the results displayed in the table. The site is regularly updated to ensure that correct information is available for display.

Population data was extracted to provide reasoning for trends listed in the data above. It would help answer questions such as why the data displayed for the country had a transmission classification of “sporadic” or why some countries had higher cases than others. Kaggle user Tanu N Prabhu provided a public CSV file that displayed the population and density statistics for several countries. This file was extracted for analytical purposes.

# TRANSFORM

Once the files were downloaded from their respective sources, they were imported into Jupyter Notebooks for transformation purposes. Cleaning of a majority of the data included deleting unnecessary columns and renaming others to enhance readability. As well as that, all of the tables extracted were checked for obsolete values within the rows. If any were found (such as statistics for global and continent locations), they were removed as the data is aimed to focus more on the results for individual countries rather than clusters of countries.

# LOAD

# ANALYSIS

## Effects on GDP per capita

Our first chart plots nations’ deaths per million from COVID-19 against the percentage change in per capita GDP during the second quarter of 2020.

# TECHNICAL ISSUES