The dataset includes 28 files scrapped from various data sources mainly the John Hopkins GitHub repository, the ministry of health affairs India, worldometer, and Our World in Data website. This dataset has daily level information on the number of affected cases, deaths, and recovery from the 2019 novel coronavirus. It is a time-series data and the number of cases given on any day is the cumulative number. Out of the 28 tables, we considered the ones which were relevant to our problem statement. These include:

A. Covid related data:

- countries-aggregated.csv: A simple and cleaned data with 5 columns with self-explanatory names.
- covid_india.csv: India specific data containing the total number of active cases, recovered and deaths statewide.
- cumulative-deaths-and-cases-covid-19.csv: A cumulative data containing death and daily confirmed cases in the world.
- daily-tests-per-thousand-people-smoothed-7-day.csv: Daily test conducted per 1000 people in a running week average.

B. Policies implemented:

- covid-stringency index.csv: The nine metrics used to calculate the Stringency Index are school closures, workplace closures, cancellation of public events, restrictions on public gatherings, closures of public transport, stay-at-home requirements, public information campaigns, restrictions on internal movements and international travel controls. The index on any given day is calculated as the mean score of the nine metrics, each taking a value between 0 and 100. A higher score indicates a stricter response (i.e., 100 = strictest response).
- covid-contact-tracing.csv: Data depicting government policies adopted in case of contact tracing. Contact tracing parameter takes values in the range [0,2]:
  - 0: No tracing policies.
  - 1: Limited tracing policies.
  - 2: Comprehensive tracing policies.

- Face-covering-policies-covid.csv:
  - 1: No policy
  - 2: Recommended
  - 3: Required in some specified shared/public spaces outside the home with other people present, or some situations when social distancing not possible

- 4: Required in all shared/public spaces outside the home with other people present or all situations when social distancing not possible
- 5: Required outside the home at all times regardless of location or presence of other people

- income-support-covid.csv: Income support captures if the government is covering the salaries or providing direct cash payments, universal basic income, or similar, of people who lose their jobs or cannot work. The values range from [0,2]:
  - 0: No income support.
  - 1: covers less than 50% of lost salary.
  - 2-: covers more than 50% of the lost salary.
- internal-movement-covid.csv: Showing government policies in restricting internal movements. Ranges from 0 to 2 where 2 represents the strictest.
- international-travel-covid.csv: Showing government policies in restricting international movements. Ranges from 0 to 2 where 2 represents the strictest.
- public-gathering-rules-covid.csv
  Restrictions are given based on the size of public gatherings as follows:
  - 0: No restrictions.
  - 1: Restrictions on very large gatherings (the limit is above 1000 people).
  - 2: gatherings between 100-1000 people.
  - 3: gatherings between 10-100 people.
  - 4: gatherings of less than 10 people.
- school-closures-covid.csv: School closure during Covid.
- stay-at-home-covid.csv: Countries are grouped into four categories:
  - 0: No measures.
  - 1: Recommended not to leave the house.
  - 2: Required to not leave the house with exceptions for daily exercise, grocery shopping, and ‘essential’ trips.
  - 3: Required to not leave the house with minimal exceptions (e.g. allowed to leave only once every few days, or only one person can leave at a time, etc).
- workplace-closures-covid.csv: Workplace closure during covid. Ranges from [0,3] where 3 represents required for all.

All of the above mentioned tables were used for visualisation to get more insights. The other files in the dataset were either not relevant to our problem statement or they had a lot of missing data in different countries and hence, we decided to drop them. When pivot tables were created for the mentioned files, missing values were generated. The data cleaning process was done the values were filled using fillna().