Data Sources provided in CSV, locally stored format to have it available ASAP. Continuing to work on connectivity to API to load data dynamically for final product, however, for the purpose of cleansing, data attributes are consistent regardless of how they are loaded. Web-Scraping of data is also a work in progress (using BeautifulSoup in Python to scrape Australia vaccination data). Classification of Covid-19 Measures is also WIP.

Possible problems with obtaining data:

NL:

* API connectivity to data sources
* Manual pulling of temperature data

Israel:

* Data in Hebrew; not configured in Excel.
* Check screenshot for English equivalent of temperature data.
* Temperature data in 2 separate columns, one for minimum, one for maximum... must get the average.

Australia:

* Covid data by age group is only available from 06-2021. Overall data is available from the start of pandemic and included in DS6.1.
  + From start of pandemic to 06-2021, there were ~5000 cases that we would be missing.
  + From 06-2021 to now, there were 64.000
* Vaccination Data not readily available.
  + Web scraping
* Hospitalization data not readily available in data set
  + Web scraping or using alternative source with data quality check against governmental media reports.
* Temperature data in 2 separate datasets, one for minimum, one for maximum... must join and get the average.

Data cleansing Idea List:

1. Aggregation to required level... most data is by towns/municipalities, need to aggregate at nation level