Project 1 – Exploring COVID-19 rates

Link to repository:

The curation of my dataset comprises of three sources: a COVID-19 dataset, a world bank development indicator dataset, and a COVID-19 policy response dataset. To start, I decided to choose the COVID-19 dataset which only had countries and many columns of dates; mostly the start of COVID-19 up till February 16. The source had three files that displayed the cases, recoveries and the deaths. After stacking each dataset into just three columns, the country, date, and cases, recoveries, or deaths respectively, I decided to filter only the year 2020. I then joined all three into a dataset name joined\_countries that shows the country, date, cases, recoveries, and deaths.

Next, I read in the policy responses dataset. This dataset had some cleaning to be done such as renaming country names in order to join it with (countries + world indicator), renaming column names, as well as dropping unnecessary columns. The columns I decided to keep before further aggregation were Country, country code, income level, date, level 1 policy, level 2 policy, and level 3 policy. The columns I dropped were not needed for my intended analysis of looking at how the countries were doing from a big picture point of view. At the end of this curation, I chose to aggregate the level 1 policies. Level 1 policy includes measures that focus on the banking sector, liquidity and funding conditions, financial markets and non-bank financial institutions (NBFIs), and payments and financial market infrastructure (payment systems). I decided not to go into the details of the focus areas in policy level 2 and 3. For the aggregation of this datasets, I first spread level 1 policy into its respective columns. Then, I added a total measure column to give an overview of how each country did to show the number of measures taken by them.

For the World Bank Indicators column, I decided to use 2019 to showcase the numbers before COVID-19 happened. One reason for this was because the numbers for 2020 had a lot of missing information and 2019 had more. Another reason for this was because the pre-covid numbers are definitely higher for some indicators today, so a comparison could be done when more numbers have been updated. The indicators I’ve chosen are employment in agriculture (% of total employment), employment in industry (% of total employment), employment in services (% of total employment), total unemployment (% of total labor force), immunization (DPT, HepB3, and measles), international tourism (number of arrivals and number of departures), and poverty headcount ratio at national poverty lines (% of population). I chose to look at employment because of how COVID has drastically affected people’s ability to work in 2020. I am also curious to see how 2019 and the end of 2020 will look after comparing them. I also chose three different immunization to see which countries had more people who would go for a vaccination program if it was necessary and which countries had less. I also just wanted to see if it will affect COVID-19 vaccine numbers in the near future. International tourism was definitely affected, hence why I chose to look at them. Lastly, I wanted to investigate the poverty head count ratio for 2019 and see how much further or lesser it may have gone at the end of 2020.