Rama Pamudji
11170220

Dataset: https://www.kaggle.com/gsutters/economic-freedom

Description search process

I found the dataset by searching on Kaggle.com with the search terms “world” and “economy”. I searched for the specific terms because the COVID-19 numbers have a lot to with the economy of countries.

I chose this specific one about economic freedom in the world because it had a lot of variables and also made use of ISO codes to make the merging of the datasets easier.

Chosen option

The chosen dataset contains information about the economies of multiple countries in the world. The dataset also shows the ranking of countries based on: size of government, legal system and property rights, sound money, freedom to trade internationally and regulation.

I chose for option 1 because the economic freedom dataset does not contain any information regarding time. I aggregated the COVID-19 dataset by country.

The datasets were merged together by means of universal ISO codes so any misspelling of country won’t cause a problem.

Table 1. Correlation Overview

Variables from the COVID-19 dataset	Variables from the Economic Freedom Dataset	Pearson’s R
total_cases_per_million	rank	-0.32
hospital_beds_per_thousand	1_size_government	0.61
gdp_per_capita	ECONOMIC FREEDOM	-0.24

Analysis description

In my main vision of finding correlations between the two datasets the focus was set on the total cases of COVID-19 and the total deaths of COVID-19 per country. Unfortunately the correlations between the two datasets were not as high as expected. To find good correlations I’ve tested many different combinations of variables. I expected total deaths and 1_size_government to have a high correlation, this was unfortunately not the case.

Furthermore, I started to focus on the hospital_beds_per_thousand and gdp_per_capita as these are very relevant topics to world economy. Comparing them to 1_size_government and economic freedom showed that there was a slight correlation, even a high correlation between hospital_beds_per_thousand and 1_size_government. In short, the variables were chosen by means of trial and error, finding the right and high correlations between variables.