Extract

We wanted to consult current data regarding coronavirus pandemic statistics, this because we didn’t have the chance to do it for the previous project and we think is really relevant and interesting data. We came across a site called Rapid API where we saw there was current covid information available free of charge, we began consulting the information and saw that the last update had been 9 days ago which was really good for a free API. Se we established the connection to the API using the request/get method and converting the response into json to get it ready for further steps.

Transform

After we requested the data, we decided to gather all the countries info into groups by continent, G7 and G20.

For this we performed two methods.

1. Got the list of the country members by continent from another table hosted on internet, then matched that list with the countries retrieved by the request.

2. For the smallest groups (G7, G20), we looked directly into the indexes of those countries in the request results.

By the time we built the groups, we did some “country\_name” changes so both names in the list and the request matched. Eg. From “United Kingdom” to “UK”.

Once the all the groups were done, we had to transform all the numeric variables into integer values, since all the data was string datatype. For this we created a function “replace(region\_list)” to do the task.

Load

Two methods were proposed to load the data to Mongo after the transformation was made:

1. The first one consisted of creating a single dictionary that contains each of the regions and groups. Then, the connection to mongo was established and the database was created. It should be mentioned, that inside of this database, only a single collection was created, and by using the update() method from pymongo, the dictionary with all the regions and groups was loaded. It must be said that with this method the way to extract the data from the Mongo database is by using the find() method from pymongo, and then iterating on this element.

2. The second method proposed was to create a collection for each region and group inside the database. Then, by using the method insert\_one() from pymongo, an iteration was made on each list to load the data. Unlike from the first method, the way to query documents, in this case, is by using directly pymongo query and projection operators.