**Project Document**

**1. Project Scope**

I gathered data from the following multiple sources:

* A csv file called covid-data from <https://ourworldindata.org/>
* A free API from <https://covid19api.com/>
* Scraped a table of European populations from <https://en.wikipedia.org/wiki/List_of_European_countries_by_population>
* A csv file called covid-tests from <https://ourworldindata.org/>

The objectives of my project were:

* Gather data from multiple sources on Covid-19
* Merge the data into one database
* Analyse the data for countries that have a population of a similar size to Ireland
* Extract the information on the countries into Dataframes
* Plot the information

**2. Description of the Data**

Sources of the data was as follows:

* A csv file called covid-19-data from <https://ourworldindata.org/> - CSV format
* A free API from <https://covid19api.com/> - JSON format
* Scraped a table of European populations from <https://en.wikipedia.org/wiki/List_of_European_countries_by_population> - HTML format
* A csv file called covid-19-tests from <https://ourworldindata.org/> - CSV format

I carried out the following to ensure the data was clean:

* Checked for null values
* Checked the datatypes of the columns and made changes as required – objects to numeric and objects to datetime
* Removed unwanted columns and rows

**3. Database**

I created a database called covid-19 with 4 tables named covid19, recovered\_persons, covid\_tests and european\_populations.

The steps I took to merge the csv files into the database are as follows:

* Saved the csv file to local drive
* Created an SQL connection to the newly created covid19.db
* Created a table in the database called covid19 / covid\_tests
* Read in the data from the csv file using Pandas and insert into a Dataframe
* Cleaned the data
* Inserted the values from the Dataframe into the database tables called covid19 / covid\_tests

The steps I took to merge the covid19 API data was as follows:

* Made a GET request to retrieve the information from the API
* Response returned in JSON format
* Converted JSON format to a Pandas Dataframe
* Cleaned the data
* Inserted the values from the Dataframe into the database table called recovered\_persons

The steps I took to merge the HTML data scraped from the Wikipedia page are as follows:

* The Python library Beautiful Soup was used to scrape the HTML code from the webpage and find the required European Populations table
* Python code was used to retrieve the content from the table and insert into a Pandas Dataframe
* Cleaned the data
* Inserted the values from the Dataframe into the database table called european\_populations

**4. End Analysis/Visualisation**

It can be seen from my analysis that Ireland have a greater number of confirmed cases of covid-19 and a higher amount of confirmed deaths in comparison to Norway and Slovakia.

However, it can also be seen that Ireland are carrying out more covid-19 tests than that of Norway and Slovakia. Is this why Ireland has more confirmed cases?

I believe it is difficult to compare countries with similar populations for several reasons:

* The number of tests being carried out by each country is different
* Are the deaths in Care homes being reported by Slovakia and Norway?
* Lockdown measures differ in each country – Slovakia for example had a very strict lockdown compared to Ireland and Norway
* How many foreign travellers were travelling back to their home countries once restrictions were put in place? Did these travellers self-isolate on their return. In Slovakia anyone returning from a foreign country was under a supervised quarantine whereas in Ireland people were just told to self-isolate.

**5. Conclusion**

I believe the data in relation to covid-19 will be better in one or two years once more is known about the virus and when more covid-19 tests can be carried out, so in that respect I am not entirely pleased with my project as I don’t believe the comparison between countries shows the full picture.

I did learn a lot from working on this project – it was a good challenge. I learnt a lot more about SQLite, Pandas, Matplotlib and Seaborn. I was also pleased that I successfully data scraped a table from Wikipedia as in my last assignment I was unable to do that task.