Project Report

Extract

### The data was sourced from the following locations:

[EU unemployment:](https://www.kaggle.com/gpreda/unemployment-in-european-union) as a TSV file.

[EU covid-19 cases:](https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide) as a CSV file.

Transformation

### EU unemployment:

* The data was in a Tab Separated Format, but the first part of each row was in CSV format, so this had to be removed and sorted.
* Colons, “:”, were used as null values in the data, which needed to be removed.
* There were also “ e” and “ u” that needed to be removed to be able to import these values.
* The TOTAL, Seasonally Adjust figures for both genders, T, were pulled out of the datset and imported into PostGres.

### EU Covid-19 Cases:

* The data was for Covid-19 cases worldwide, this had to be filter by the continent for Europe only
* Then grouped the data by countries and months
* Got the total value of cases and deaths for each of the countries and months
* Renamed the months from numbers to names
* Also Exported the Dataframe out to a csv file

Load

* Loaded both of the final DataFrames into Postgres via Pandas
* Used PostGres in favour of MongoDB as it uses a simple table structure