Project Report Gavin Horan

# GitHub URL

https://github.com/gwavin/UCDPA\_gavinhoran.git

# Abstract

I mainly decided to focus on ECDC COVID Data, because it is of vast interest personally, and professionally. We may find ourselves in a position where we may have to make a decision as to whether or not we close or limit hours or access to the pharmacy in the Rotunda, and this type of analysis could be invaluable.

I have looked at the links between the countries in the EU, and decided to focus on those I considered to have health systems and populations which I considered to be close match to us.

# Introduction

COVID-19 and the resulting SARS-COV-2 is the biggest health emergency faced in modern memory. The pathogen reflects many pandemics of the past, but is occurring in a society where we can record a lot more data that we could in previous pandemics.

This is a blessing and a curse, in that it can be difficult to pull out the useful information from the flood that has come to us.

# Dataset

# The ECDC is an EU agency which engages in disease surveillance, epidemic intelligence, and response, among a wide range of other roles.

# As a result of this, they have an excellent selection of datasets from which I drew to build reports which I hoped would be informative. These were located at <https://www.ecdc.europa.eu/en/covid-19/data>

# In order to harness the power of API’s, I also sampled some datasets from Kaggle.com.

# The use of Kaggle required the downloading of the Kaggle commandline interface tool .

# According to Kaggle, this is the easiest way to interact with Kaggle’s public API. This is implemented via python, and was ran from the terminal in PyCharm.

# Simply by running “pip install Kaggle” from the terminal I took advantage of the python package manager, PIP, and installed the CLI tool with ease. Resources could then be selected by searching with were selected for download were then easily accessible from within PyCharm.

# kaggle datasets list -s health > KaggleHealthList3.csv downloaded the list of Health data files which were available for examination.

# Implementation Process

# In order to harness the power of API’s, I also sampled some datasets from Kaggle.com.

# The use of Kaggle required the downloading of the Kaggle commandline interface tool .

# According to Kaggle, this is the easiest way to interact with Kaggle’s public API. This is implemented via python, and was ran from the terminal in PyCharm.

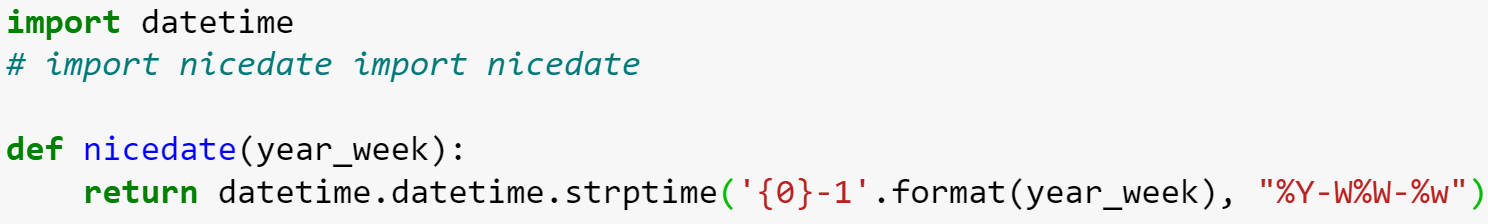
# Simply by running “pip install Kaggle” from the terminal I took advantage of the python package manager, PIP, and installed the CLI tool with ease. Resources could then be selected by searching with were selected for download were then easily accessible from within PyCharm.

# kaggle datasets list -s health > KaggleHealthList3.csv downloaded the list of Health data files which were available for examination.

I also examined the NHANES

I created a function for re-use with the ECDC dataset; they have an annoying habit of using 2020-W30 etc to refer to the 30th week of the year.

I rewrote a code snippet to deal with this;



This takes in the nasty format, interprets it by slicing out the year, and the week number, and then incorporating those into a datetime which includes day 1 by default. This meant that the labels at the bottom of my graph could look much nicer.

In order to become familiar with the various utilities that one might take advantage of I also created and used a Jupyter notebook for the project.

# Results

I got some nice charts;

Put these below; everything essentially that is a .py in my UCDPA folder on github.

On the evening of the 23rd

# Insights

I'm going with "This regression analysis shows that the impact of the independent variable is not constant. Early in the pandemic, tests done were linked strongly to new cases diagnosed. This implies that the increase in testing facilities early in the pandemic meant that we were catching a lot but not all of the cases that we might. However, since we have increased testing capacity, peaks are more easily detectable, since when a "wave" of infection begins, or ends, this is visible in the test results returned. As the number of available testing facilities has grown to an appropriate level, tests.

Holland seem to be going slightly ahead of us.

Lots of peaks indicating that there was a wave; these speak of

# References

(Include any references if required)