Aaron Noonan
COVID-19 DATA PYTHON PROJECT

GitHub: https://github.com/custerdome91/-UCDPA_AaronNoonan-

(Note: I created a fourth graph, a bar chart, comparing Cork and Dublin's highest COVID-19 case spikes over the time period. I did not have space to discuss it in this report but the .py file and graph are included in the .ZIP file).

Our understanding of the COVID-19 pandemic has been greatly informed by data. Worldwide, governments and public heath bodies have been releasing daily statistics on the number of cases, deaths and vaccinations that have been taking place in their respective countries. This information has allowed both governments and the general public to make effective decisions on how to deal with the virus and to plan for the future.

In Ireland, this is no different. Our government has endeavoured to use high-quality data on cases, deaths, testing, positivity rates, and vaccinations to inform them on when to impose lockdowns and when to lift them. This data was perhaps most important during the months from December 2020 through February 2021. After a six-week lockdown was lifted in early December, the lead up to the Christmas period saw a stark increase in social mixing among the population that led to an unprecedented spike in COVID-19 cases. In early January, Ireland had the highest infection rate in the world, recording a record high of 8,227 cases on January 8th. By late February, after a stringent lockdown, these numbers had decreased to around 600 daily cases.

For this project, I hope to compare COVID-19 case data across different geographical areas on the island of Ireland, and draw observations. I hope to answer two questions:

   1. *"Did population density impact COVID-19 cases rates during the spike in cases between December 2020 and January 2021?"*

   2. *"Is there a correlation between COVID-19 case increases in the border counties and Northern Ireland over the same period?"*

The first two graphs will tackle the first question, while the third will tackle the second question.

Graph 1: Did Population Density Impact COVID-19 Case Rates? Comparing Cork City and County

Aim

The aim of this graph is to compare the COVID-19 incidence rates in Cork City and Cork County. Cork City has a population density of 1,123/km2, while Cork County has a population density of 72/km2. This is a stark difference, and could reflect smaller cases numbers in the county area.

Dataset

The dataset I used for this graph is official government data that can be found at: https://data.gov.ie/dataset/covid19-leacases-mapped-historic-records/resource/0a4c2eca-6230-48c4-9d80-f9e66375e70c.

The CSV file contains COVID-19 data from August 2020 to May 2021, for the 166 Local Electoral Areas (LEAs) of the 26 counties of the Republic of Ireland. Each LEA has the incidence rate per 100,000 population of confirmed COVID-19 cases notified in the previous 14 days (P14_100K),

updated at weekly intervals. This is a standard method of measuring the incidence of COVID-19 in an area, used across the European Union and beyond.

For example, the screenshot below shows the incidence rate for the Cork South East LEA of Cork between 28/09/2020 and 23/11/2020.

OBJECTID	LE_ID	ENGLISH	COUNTY	GUID	P14_100k	EventDate
1159	1340412	CORK CITY SOUTH EAST LEA-6	CORK	2ae19629-3ef4-13a3-e055-999990000001	60.8	2020/11/23
1160	1340412	CORK CITY SOUTH EAST LEA-6	CORK	2ae19629-3ef4-13a3-e055-999990000001	81.8	2020/11/16
1161	1340412	CORK CITY SOUTH EAST LEA-6	CORK	2ae19629-3ef4-13a3-e055-999990000001	182.3	2020/11/09
1162	1340412	CORK CITY SOUTH EAST LEA-6	CORK	2ae19629-3ef4-13a3-e055-999990000001	315.6	2020/11/02
1163	1340412	CORK CITY SOUTH EAST LEA-6	CORK	2ae19629-3ef4-13a3-e055-999990000001	425.4	2020/10/26
1164	1340412	CORK CITY SOUTH EAST LEA-6	CORK	2ae19629-3ef4-13a3-e055-999990000001	425.4	2020/10/19
1165	1340412	CORK CITY SOUTH EAST LEA-6	CORK	2ae19629-3ef4-13a3-e055-999990000001	226.7	2020/10/12
1166	1340412	CORK CITY SOUTH EAST LEA-6	CORK	2ae19629-3ef4-13a3-e055-999990000001	93.5	2020/10/05
1167	1340412	CORK CITY SOUTH EAST LEA-6	CORK	2ae19629-3ef4-13a3-e055-999990000001	79.5	2020/09/28

By isolating the Cork LEAs within this CSV file, I was be able to compare their incidence rates over the time period.

Graph and Insights
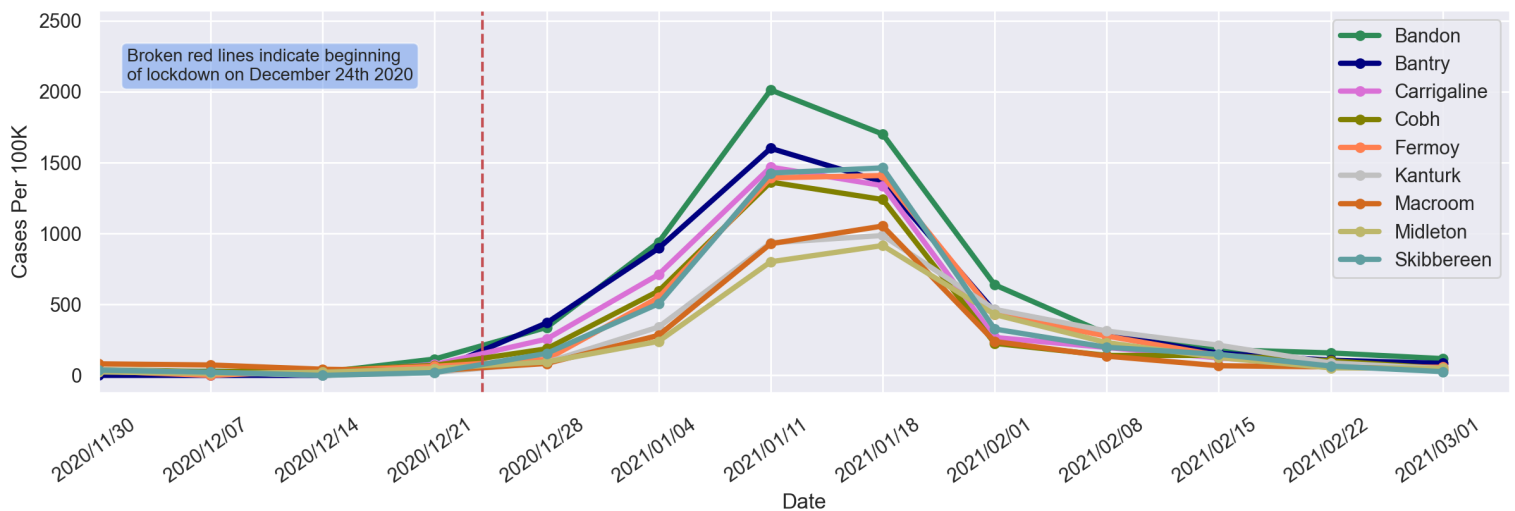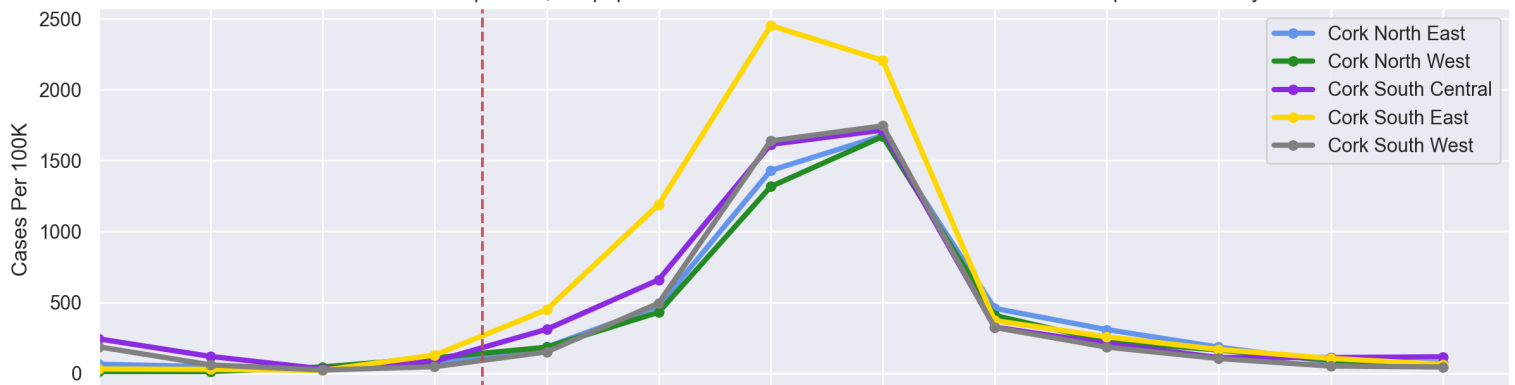
There are two graphs, the top one showing all the Cork City LEAs and the bottom one showing the Cork County LEAs. A broken red line indicates the beginning of lockdown on December 24th 2020.

- As expected, Cork City did have a higher COVID-19 incidence rate over the period December through February. This was particularly the case for the Cork South East LEA, which spiked at 2,452 cases on January 11th.

- By comparison, the highest COVID-19 incidence rate in Cork County was in Bandon, which spiked at 2,012 cases on January 11th.

- From the graphs, we can see in the week between January 4th and January 11th, there was a steady increase in cases for both Cork City and Cork County.

- The implementation of lockdown saw both regions return to relatively low levels of COVID-19 by the start of March. The downward trend between mid-January and March was largely the same between both regions.

- While there was not a huge difference in COVID-19 incidence between both geographical areas, it is clear that Cork City suffered a higher incidence along with a steeper increase in cases in early January. It is therefore reasonable to believe that Cork City's higher population density was a contributing factor to this.

Coding Notes

In order to draw an effective comparison between the two geographical areas - city and county - I needed to import the CSV file, clean the data up, and then isolate each Cork LEA from the rest of the CSV file.

From the "*CorkLEA.csv*" file I created, I worked to create a DataFrame for each of the Cork City LEAs and for the Cork County LEAs, and then slicing that by the dates I required. Below is an example of how I did this for Cork North East. I repeated this for each LEA.

I isolated Cork North East from the rest of the CSV

```
cork_NE = cork.loc[cork['ENGLISH'] == "CORK CITY NORTH EAST LEA-6"]
```

I then created a new DataFrame for Cork North East

```
cork_NE_df = pd.DataFrame(cork_NE)
```

However, this created a DataFrame for Cork North East between August 2020 and May 2021. In order to use only data between December 2020 and February 2021, I used the following code, which used the DataFrame's index numbers to slice between the dates required.

```
dec_mar_NE = cork_NE_df.loc[258:438, :]
```

In order to plot this on a graph, I needed to create two variables, one for each axis. I used the following code to isolate the dates (X axis) and the incidence rate per 100,000 over 14 days (Y axis).

```
cork_NE_df2 = dec_mar_NE['P14_100k']
cork_NE_df2_date = dec_mar_NE['EventDate']
```

Graphs 2: Did Population Density Impact COVID-19 Case Rates? Comparing Dublin and Cork

Aim and Dataset

The aim of this graph is to compare COVID-19 incidence rates between Cork City and Dublin City. There is a significant difference in population density between the two geographical areas. Cork City has a population density has 1,123/km2, while Dublin City has 4,811/km2. From this, it would be reasonable to expect a higher level of COVID-19 transmission among the more densely populated Dublin City.

This graph will use the same dataset as above. For the sake of clarity, I have defined Dublin City as the five most central LEAs as shown on the following map: http://www.boundarycommittee.ie/ Maps Dublin%20City%20LEA%20Report%20Map%202018.pdf

These are *Cabra-Glasnevin*, *Clontarf*, *North Inner City*, *South East Inner City*, *South West Inner City*. Cork City's five LEAs remain the same as in the previous graph.

Graph and Insights

Incidence rate per 100,000 population of confirmed COVID-19 cases notified in the previous 14 days
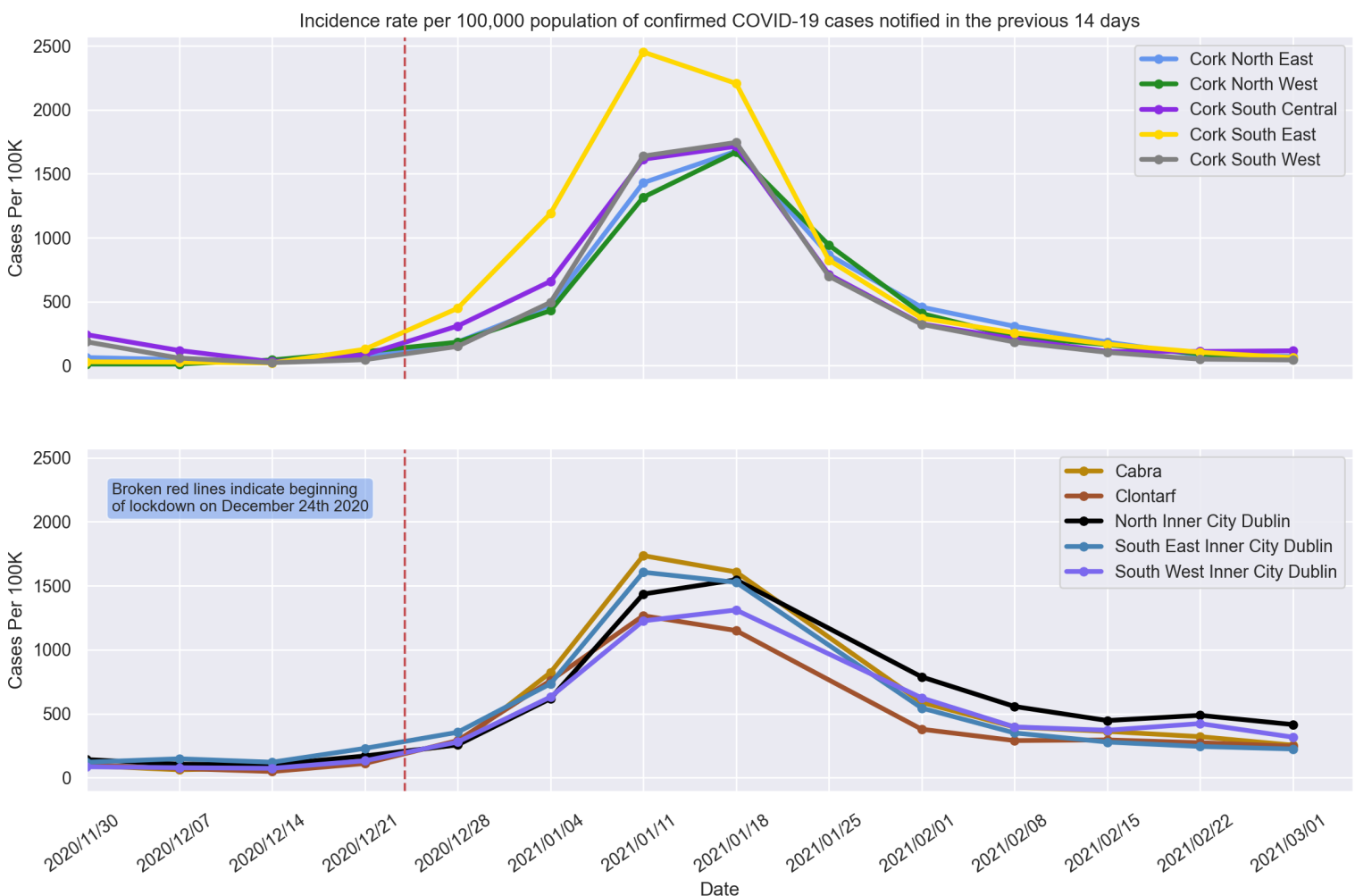
Although there was an expectation that Dublin City would have a higher COVID-19 incidence rate over the time period, this turned out not to be the case, despite the fact its LEAs are more densely populated.

- Dublin City region registered a far lower incidence rate than Cork City over the time period. While, as noted previously, Cork South East spiked at 2,452 cases on January 11th 2021, Dublin City's Cabra spiked at only 1,735 cases on the same date.

- Outside of these two LEAs, most other LEAs in both regions spiked at similar numbers, somewhere between 1,300 and 1,700 cases per 100,000 people.

- As with Cork City and County, Cork City and Dublin City cases spiked at around the same time, roughly 2.5 weeks after lockdown was implemented on December 24th.

- It is clear from the graph that although cases in Cork City were trending toward 0 by the start of March, cases in all Dublin City LEAs analysed remained stubbornly high, with North Inner City Dublin having an incidence rate of 415 per 100,000 by March 1st.

- Although the population density of an area may not have directly resulted in higher COVID-19 incidence rates over the time period, it is possible that population may have kept Dublin City's case numbers higher while Cork City's continued to drop off as lockdown progressed.

Graph 3: Is there a correlation between COVID-19 case increases in the Border Counties and Northern Ireland over this period?

Aim

The aim of this graph was analyse if there was any correlation between the increase in COVID-19 cases in border counties between December 2020 through February 2021, and the increase in COVID-19 cases in Northern Ireland during the same period.

Unfortunately, this analysis was hampered by a lack of county-by-county data for Northern Ireland. This information is not released by the UK government at this time.

Dataset

I used two datasets for this graph. One for the Irish border counties and another for the Northern Ireland data.

The Irish border counties dataset I used can be found here: https://data.gov.ie/dataset/covid19countystatisticshpscireland1/resource/9ec1cba9-a9a9-4890-a4cc-980f4676197b

Unlike the previous dataset, this one contains COVID-19 cases numbers for each county on a cumulative basis, with one row per date, dating from February 27th 2020 to May 11th 2021. The screenshot below shows a sample of the CSV file, with some columns narrowed to make it viewable.

OBJECTID	ORIGID	CountyName	PopulationCensus16	TimeStamp	IGEasting	IGNorthing	Lat	Long	L	ConfirmedCovidCases	PopulationProportio	C	C	S
1	1	Carlow	56932	2020/02/27	278661	163444	52.7168	-6.8367	h	0				
27	1	Carlow	56932	2020/03/01	278661	163444	52.7168	-6.8367	h	0				
53	1	Carlow	56932	2020/03/02	278661	163444	52.7168	-6.8367	h	0	0			
79	1	Carlow	56932	2020/03/03	278661	163444	52.7168	-6.8367	h	0	0			
105	1	Carlow	56932	2020/03/04	278661	163444	52.7168	-6.8367	h	0	0			
131	1	Carlow	56932	2020/03/05	278661	163444	52.7168	-6.8367	h	0	0			
157	1	Carlow	56932	2020/03/06	278661	163444	52.7168	-6.8367	h	0	0			
183	1	Carlow	56932	2020/03/07	278661	163444	52.7168	-6.8367	h	0	0			
209	1	Carlow	56932	2020/03/08	278661	163444	52.7168	-6.8367	h	0	0			
235	1	Carlow	56932	2020/03/09	278661	163444	52.7168	-6.8367	h	0	0			

The Northern Ireland dataset can be found here: https://coronavirus.data.gov.uk/details/cases

Unlike in the Republic of Ireland, the UK government does not break down COVID-19 case data by county in Northern Ireland, so I was limited to using case data for the entire country. This will be reflected in the graph. See sample screenshot below.

UKCOVIDDATA

areaType	areaName	areaCode	date	newCasesByPublishDate	cumCasesByPublishDate
nation	England	E92000001	2021-05-17	1645	3889132
nation	England	E92000001	2021-05-16	1471	3887487
nation	England	E92000001	2021-05-15	1614	3886016
nation	England	E92000001	2021-05-14	1828	3884402
nation	England	E92000001	2021-05-13	2229	3882574
nation	England	E92000001	2021-05-12	1804	3880345
nation	England	E92000001	2021-05-11	2117	3878541
nation	England	E92000001	2021-05-10	2009	3876424
nation	England	E92000001	2021-05-09	1418	3874415
nation	England	E92000001	2021-05-08	1742	3872997

Graph and Insights

COVID-19 Cumulative Cases From December 2020 through February 2021
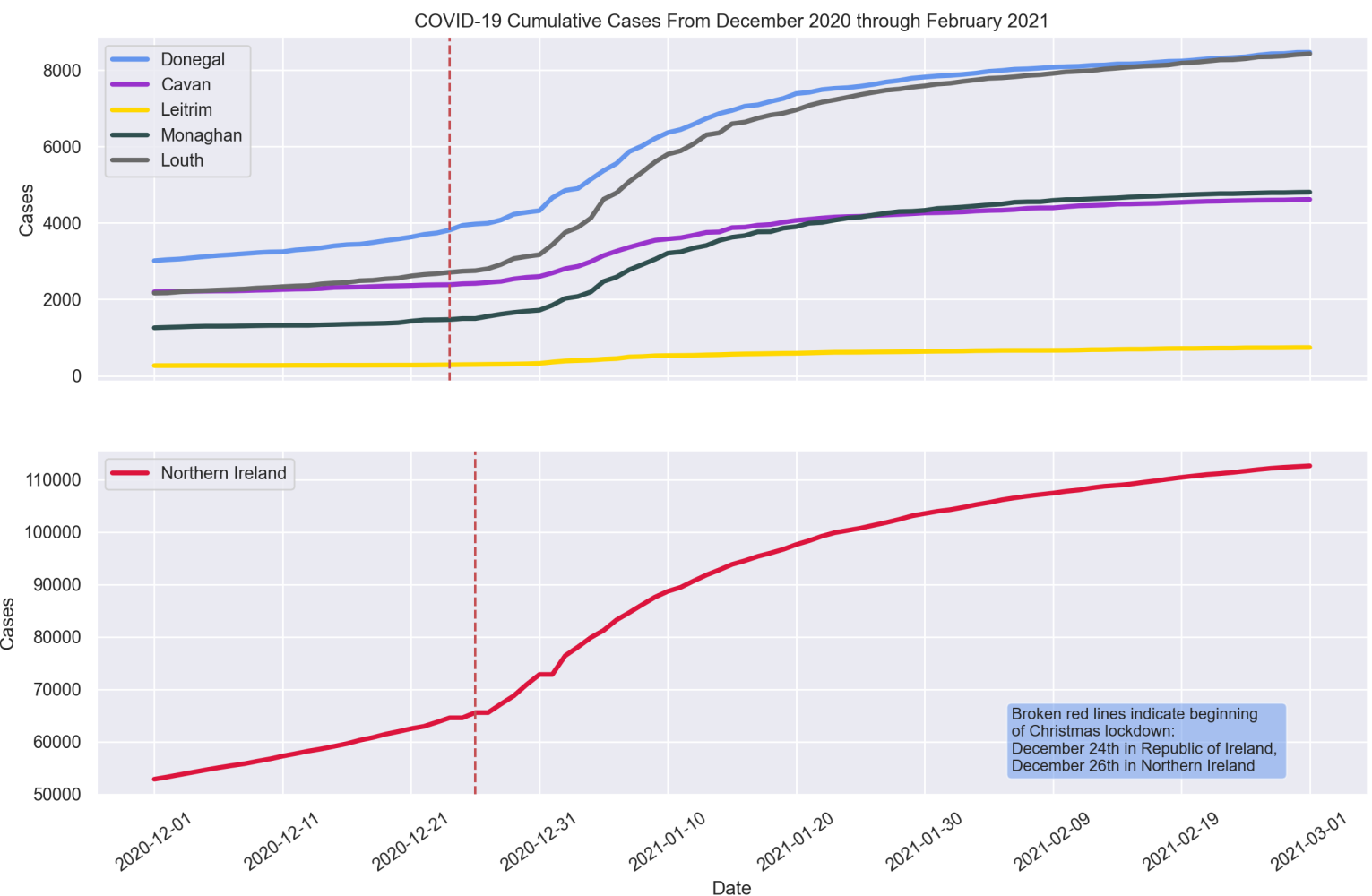
Unlike previous graphs, there was a limitation here in that I was unable to break down the county-by-county data for Northern Ireland. Although the lack of available data made it difficult to answer if there was a correlation, we can nonetheless take some insights from the graphs:

- It is clear when comparing the two graphs that cases numbers in the Border Counties and Northern Ireland increased along a similar trajectory over the time period.

- Most of the Border Counties saw significant spikes in cases from the end of 2020, with the exception of Leitrim, whose cases numbers stayed remarkably low. Leitrim's border with Northern Ireland is much smaller than the other counties so it's possible this was a reason for its lower case numbers.

- Northern Ireland also saw a significant spike in cases, beginning just after entering lockdown.

- Northern Ireland's cumulative case numbers increased by 71% between the implementation of lockdown and March 1st 2021. However for Donegal and Louth, these increases were 121% and 211% respectively. This is represented in the graph by the much steeper incline for both counties.

Coding Notes

As the Border County dataset was arranged differently to the previous one, I was unable to use the *.loc[]* method to isolate the dates for each county to December 2020 through February 2021. In order to overcome this, I created a boolean mask that I used to create a new DataFrame with only the dates I required.

```python
import datetime as dt

donegal_df['TimeStamp'] = pd.to_datetime(donegal_df['TimeStamp'])

donegal_mask = (donegal_df['TimeStamp'] > '2020/11/30') &
(donegal_df['TimeStamp'] <= '2021/03/01')

donegal_df2 = donegal_df.loc[donegal_mask]
```

In the example above, the result was a DataFrame (donegal_df2) containing all the case data for Donegal, between November 30th 2020 and Match 1st 2021.

Finally, I created two new variables for each axis. I used the following code to isolate the date (X axis) and the incidence rate per 100,000 over 14 days (Y axis).

```python
donegal_df3 = donegal_df2['ConfirmedCovidCases']

donegal_df3_date = donegal_df2['TimeStamp']
```