

Road Usage and Covid Cases

Team 7

Abstract

We used the Data set provided by the data.gov.ie about cases per county in Ireland. We decided to analyse the road users in some of the major counties to see if there was any link between the number of road users and the number of Covid Cases.

We analyse and adapt the data to be usable in python. We visualise the data to see if there is any correlation.

Preparing Data

In this project we are combining two Datasets. The first is Dataset 10 from the competitor package - HSPC data containing total Covid cases, per county, each day. The second is from the National Road Authority (NRA) website, and contains traffic counts on roads across Ireland. Each needed some cleaning and transformation before they could be used together.

Covid Cases Data

This data came from the competitor packet (Database 10, County Statistics). At the begin We made a pivot table with the county's and dates with the daily sum of Covid cases as the values. We then made another table calculating the daily recorded case.

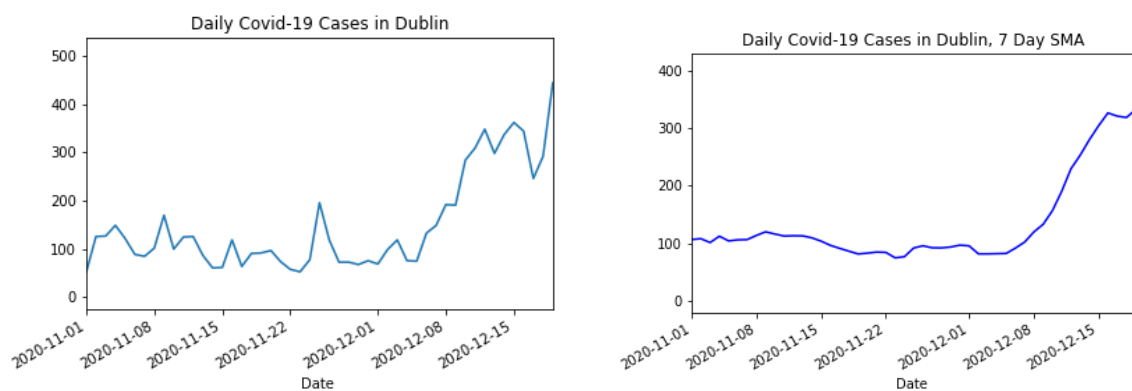


Figure 1: Daily case counts in Dublin, before and after applying a rolling average.

Traffic Data

The first tricky part about getting this data was the web interface provided by the NRA - we could not find a way to request data across the country in one go, and attempting to request data for even one location for a large enough time period caused internal server errors. We were a little worried that we might have overloaded the server at one point, though it continued to respond to smaller requests.

In the end, we restricted our time-frame to the last three months, and looked at 6 points around Ireland in particular. Nice as it would be to look at things on a larger scale, time was an issue.

The second issue was the file format. According to excel, and pandas, the incoming .xls files from the NRA were not valid .xls files, which made reading them difficult. Excel was able to open them, however, and let us daily counts and date rows into csv format.\par

There was some manual work in this, and while most of it could be automated, the invalid .xls format blocked our attempts to write the conversion to csv format in python. This is the main reason we restricted our attention to 6 roads.

Finally, once the data was imported into python, the rolling weekly average was taken of the daily average. This smoothed out the differences between the weekdays and weekend days.

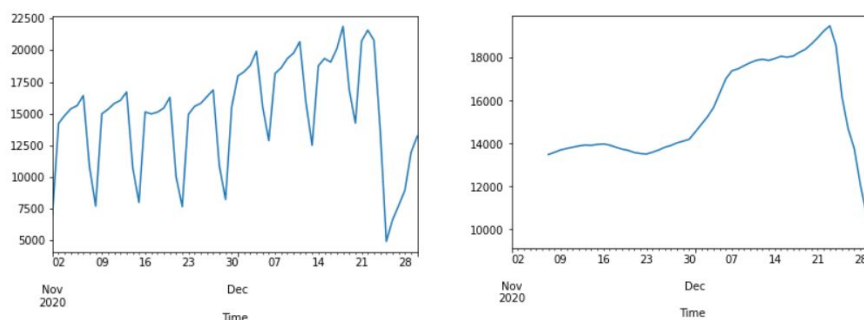


Figure 2: Traffic counts in Letterkenny, before and after applying a rolling average.

Direct Comparison

It makes sense from the get go that a rise in traffic counts would correlate with a rise in cases - traffic is an indicator of human activity, and an indicator of how much the country is in lockdown. Furthermore, we know historically that the lockdown was relaxed in the months before Christmas, and that cases rose sharply around that time.

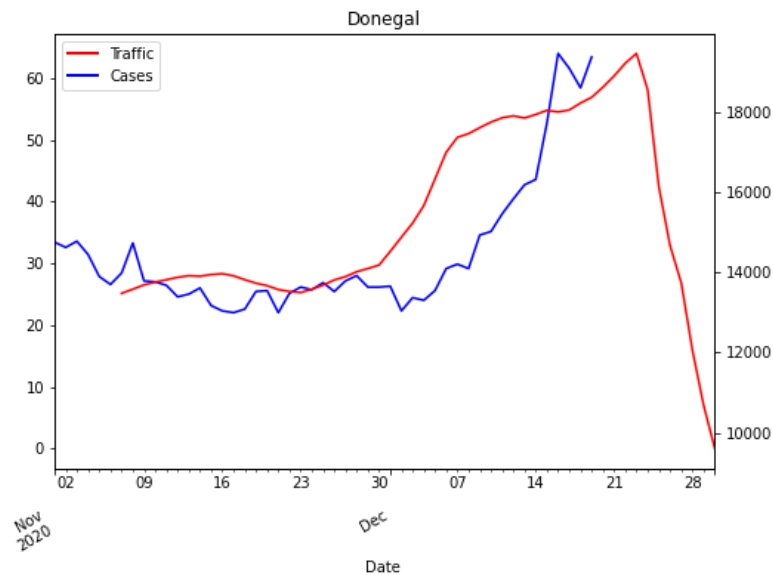


Figure 3: Traffic Counts and Daily Cases against time, with cases shifted back a week to account for the incubation period. This is a rough guess, the incubation period can apparently be as long as 2 weeks.

With that in mind, we looked at comparing the data of the two directly. One thing we noticed was that the rise in traffic data comes before the rise in COVID cases by about a week, or two. This coincides quite nicely with the incubation period of the virus, which is one or two weeks.

Still, while there is a nice visual correspondence between the two datasets, we did not get any further in trying to capture some of this intuition with statistics.

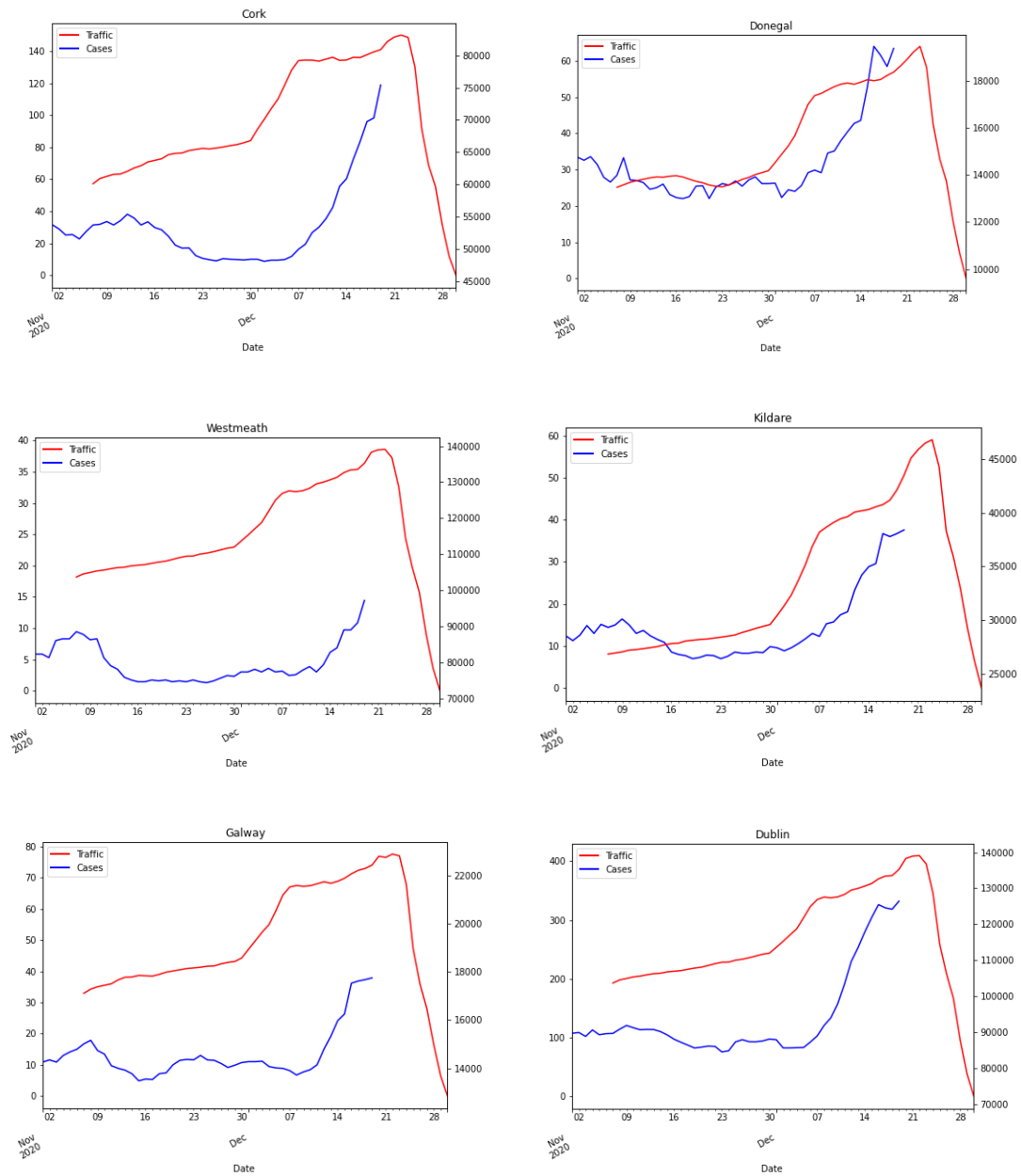


Figure 4: Comparisons of traffic vs cases in Donegal and 5 other locations.

Regression Analysis

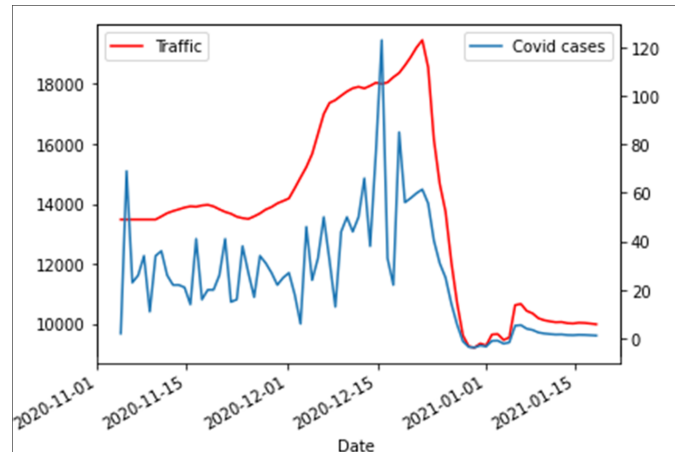


Figure 5: Regression prediction vs the traffic data that it is based on. The prediction does not match the reality - it is completely dependent on the one feature it is based on.

We made an attempt at creating a regression model to predict, based only on traffic. From the plot we can see that a simple regression based on our data collection that it failed entirely to predict the events after the New Year, which is natural - when the traffic dropped dramatically due to lockdown, the cases took quite some time to follow suit.

If we had been able to collect more data then the model will have had more dependencies and the model would have performed much better than the current model. Because of the simplicity of our data a linear model may not have been the best model to predict data on.

The traffic data still could be very useful as a predictor in conjunction with other features, and perhaps over a longer time frame that included previous entries into lockdown.

Conclusion

Ireland has at its disposal a comprehensive data of traffic data, which can be related to activity within counties, movement between counties and movement across the border to northern Ireland. This project took a step towards looking at how this data might be used - at least directly, in prediction.