**1.Introduction:**

Ireland had spent more than €17.1 Billion in response to Covid 19 by the end of February,2021 according to the state’s report, out of which more than €2.1 Billion was spent solely by the Department of health including more than €290 Million spent on tracing itself but still the number of cases is high.

This brings us to the question of how we can understand this trend and locate the areas/counties where the chances of breakout of COVID-19 are the highest. We plan to investigate it and see if we can predict the areas where the next breakout will happen so that the allocation of funds to those counties can be better.

**2.Dataset Features and Cleaning:**

To find the data that we could use, we started by looking at the various Irish government provided resources that could give us the number of cases and deaths weekly/daily by every county along with the vaccination numbers in each county. But neither the HSE nor the COVID-19geohive had weekly data for the number of cases and deaths.

Finally, the CSO website had a weekly sheet for the number cases/deaths by county, age, sex. We then scrapped the data from every weekly sheet with the aim of creating a final dataset of the number of cases/deaths in every county as well as by age, sex etc.

We then added another ordinal field to the dataset which considered the status of lockdown during that week, ranging from 0-5.

The dataset though was really dirty, we applied various techniques in order to clean the dataset ranging from stripping white spaces replacing random values (such as , .. etc) with blank spaces and filling in the none values with 0.

Also, we were not able to find any dataset which gave out weekly figures for the number of people vaccinated in each county. The only dataset available at CSO was divided based on occupation.

The following charts show the trends for the number of cases and deaths in Ireland for 2020 and 2021. The bar chart shows the number of cases on a scale of 0-10k whereas the red line shows the number of deaths on the scale of 0-800.

Chart, histogram

Description automatically generated

Chart, histogram

Description automatically generated

As we can see the number of cases in Ireland were not that high until mid-October 2020 after which they reached an all time high during the end of 2020 and beginning of 2021 before plummeting. The cases remained stable until August 2021 before rising again during the months of October and November in 2021. As for deaths, the number were high during the beginning of the pandemic after which they plateaued because of the policies of the government before rising again in the beginning of 2021 after which the have remained stable and being 0 as well for a few weeks in between.

**3.Methods:**

**ARIMA** (Auto Regressive Integrated Moving Average) - After studying various papers, we found that most of the time series gave the best results with a ARIMA model. Hence, we decided to give it a go. ARIMA is essentially a univariate model i.e., it uses a single predictor to forecast the value of interest. Integrated represents the differentiating of raw observations to make the time series stationary i.e., it does not have any seasonality. The term Autoregression means that the variable is predicted using its own values. ARIMA uses moving average model to predict the value of the variable based on the current value along with the error terms of the past values.

**4. Experiments and Results:**

After cleaning the dataset, we decided to run the ARIMA model for just one county (DUBLIN) and to select the values of the hyperparameters p, q and d. The values p, q, and d describe the various terms in ARIMA, the p describe the AR i.e., which order of Auto Regressive model we want, I define the Integrative part, which is order of differentiating we require, and d represents the order of Moving Average.

After referring various literature, we decided to run the model for values of p in range of 0 to 10 and for q and d in range of 0 to 3, we filtered off the values for which minima was not reached and after that the results were as follows:

ARIMA(0, 0, 0) RMSE=1762.856

ARIMA(0, 0, 1) RMSE=1116.679

ARIMA(0, 0, 2) RMSE=897.241

ARIMA(0, 1, 0) RMSE=531.029

ARIMA(0, 1, 1) RMSE=521.330

ARIMA(0, 2, 1) RMSE=518.761

ARIMA(0, 2, 2) RMSE=509.695

ARIMA(1, 2, 1) RMSE=509.476

ARIMA(2, 2, 0) RMSE=479.825

Best ARIMA(2, 2, 0) RMSE=479.825

The model gave the best results for the values of p=2, q=2 and d=0 and therefore that order was chosen for running all the values.

After deciding the values of the parameters, we divided the dataset into a 66/33 split and ran the ARIMA model for all the counties, noting the RMSE and MSE as well as plotting the expected v/s predicted values for the same. Below are the values of MSE and RMSE along with the final graph showing the projected and expected cases for all the counties.

For county Carlow : MSE -> 2207.463 RMSE-> 46.984

For county Cavan : MSE -> 3920.747 RMSE-> 62.616

For county Clare : MSE -> 3409.201 RMSE-> 58.388

For county Cork : MSE -> 36885.281 RMSE-> 192.055

For county Donegal : MSE -> 25477.220 RMSE-> 159.616

For county Dublin : MSE -> 230232.090 RMSE-> 479.825

For county Galway : MSE -> 14697.977 RMSE-> 121.235

For county Kerry : MSE -> 5251.407 RMSE-> 72.467

For county Kildare : MSE -> 6079.847 RMSE-> 77.973

For county Kilkenny : MSE -> 1161.042 RMSE-> 34.074

For county Laois : MSE -> 3185.941 RMSE-> 56.444

For county Leitrim : MSE -> 943.870 RMSE-> 30.722

For county Limerick : MSE -> 11581.381 RMSE-> 107.617

For county Longford : MSE -> 1086.673 RMSE-> 32.965

For county Louth : MSE -> 9744.466 RMSE-> 98.714

For county Mayo : MSE -> 6998.997 RMSE-> 83.660

For county Meath : MSE -> 7171.258 RMSE-> 84.683

For county Monaghan : MSE -> 5177.140 RMSE-> 71.952

For county Offaly : MSE -> 1214.352 RMSE-> 34.848

For county Roscommon : MSE -> 2161.078 RMSE-> 46.487

For county Sligo : MSE -> 1109.576 RMSE-> 33.310

For county Tipperary : MSE -> 5503.321 RMSE-> 74.184

For county Waterford : MSE -> 4369.105 RMSE-> 66.099

For county Westmeath : MSE -> 4714.585 RMSE-> 68.663

For county Wexford : MSE -> 4490.008 RMSE-> 67.008

For county Wicklow : MSE -> 2423.900 RMSE-> 49.233

Shape

Description automatically generated

The blue line depicts the expected value, and the red line gives the predicted value, as it is visible that the model is over-projecting the number of cases in certain counties but works well for most of the counties.

Then we saved these values in a dictionary and sorted them to find out the counties for which the number of cases in coming four weeks is expected to be the highest and below are the results of the same:

The Counties predicted to have the Highest Cases in next four weeks are :

Dublin -> [7831, 8852, 9597, 10493]

Cork -> [3151, 3452, 3653, 3913]

Galway -> [1322, 1419, 1503, 1596]