**Project Report GitHub URL**  
<https://github.com/eoin1979/UCDPA_eoinbyrne.git>

**Abstract**   
(Short overview of the entire project and features)

This project involves creating a functional report based on 3 datasets supplied by the Government of Ireland. You will see the importing of libraries like pandas, NumPy, seaborn, and Matplotlib and datasets in relation to Covid 19, Vaccines and Boosters.,

The datasets will be merged and reviewed to make it easier to create visual reports like Line Charts, Bar Charts, Scatterplots, Histograms, and Boxplots.

I also tested out learnings like sorting and indexing and grouping, dropping columns, creating custom functions, and looping and interows.

**Introduction**  
 (Explain why you chose this project use case)

In Ireland, Covid 19 started on the 29th of February 2020 with our first case. This started a snowball reaction within our country and around the world. I chose this project to review the many factors associated with the pandemic and bring all the different data into 1 report.

My goal is to not only review the up-to-date COVID information but also review the demographics of the people who were infected and also ended up in the Hospital.

We will also review the Vaccine rollout of doses and boosters to see if the boosters are as popular as the original vaccine campaign pushed by the HSE and Irish Government.

Demographics will also play a big part in the review in relation to Age Groups and Gender of the people infected with surprising results.

**Dataset**   
(Provide a description of your dataset and source. Also, justify why you chose this source)

My Project consists of 4 Datasets

I download 3 datasets from <https://data.gov.ie/> which is an Irish Government website. I chose this data as it’s provided by the HSE, it’s up to date and factual.

* Daily Covid Cases - [COVID-19 HPSC Detailed Statistics Profile](https://data.gov.ie/dataset/covid-19-hpsc-detailed-statistics-profile?package_type=dataset) ( [Source](https://data.gov.ie/dataset/covid-19-hpsc-detailed-statistics-profile?package_type=dataset))
* Daily Vac 1 and Vac 2 - [COVID-19 HSE Daily Vaccination Figures](https://data.gov.ie/dataset/covid-19-hse-daily-vaccination-figures?package_type=dataset) ([Source](https://data.gov.ie/dataset/covid-19-hse-daily-vaccination-figures?package_type=dataset))
* Daily Boosters - [COVID-19 HSE Daily Booster Vaccination Figures](https://data.gov.ie/dataset/covid-19-hse-daily-booster-vaccination-figures?package_type=dataset) ( [Source](https://data.gov.ie/dataset/covid-19-hse-daily-booster-vaccination-figures?package_type=dataset))
* When I merged my 3 datasets in python, I downloaded the merged dataset as.csv

## Implementation Process (Describe your entire process in detail) Steps involved in Data Analysis: - Importing required packages - Gathering Data -Transforming Data to our needs. - Exploratory Data Analysis and Visualization STEP-1 Import the following libraries: Pandas, NumPy, Seaborn & Matplotlib Step -2 Importing 3 datasets to make data frames for analysis, Each given their own name. Step -3 Analyzing Data frames to see if they loaded correctly and looking for the common denominator to merge my data frames. I looked at the heads and tails of each data frame to confirm the format and count of each set are matching the .csv files Step -4 Checking if NaN appears in the Date Column of any of the Datasets. This was expected. Vaccines were rolled out a year after our first case and the booster rolled out 8 mons later again. Step-5 Finding the Data is the common denominator in all 3 Data frames. To merge all the data, I want to add the new data as columns instead of rows, The common denominator was Date which was consistent in each set. Step-6 Checking if duplicates appear in the Date Column of any of the Datasets. I didn’t have any duplicate date in my set. Step-7 Merging data frame 1 & 2 together using pd.merge. I couldn’t figure out how to merge all datasets in one go so I did it one by one Step-8 Checking Data and Columns Step-9 Merging data frame 2 & 3 together using pd.merge Step-10 Replacing NaN with 0. I left this until the merge who I would only have to carry this function out once at the end Step-11 Dropping Columns that are not required in this report. There were Object IDs and other columns not relevant to my reporting requirements

## Step 12 Grouping and Sorting some Columns to review Data in ascending order ( Covid Deaths and Confirmed Cases. I wanted to test the sorting and grouping function

## Step-13 Now we will focus on rounding up the numbers, for this, we will change the floats to Ints. Some of my columns had decimal points as they were showing up as floats, I changed this to ints which cases the decimal points to disappear Step-14 I will also change my Date from object to datetime Step-15 Creating custom functions to see which Days were good or bad based on Covid cases over or under 1000. I wanted a quick review of this, but I didn’t want to add a new column to my data frame. Step-16 Looping over my data frame Step17 NumPy Converting df to NumPy, checking min and max of df and certain columns

## Step-18 Exporting the Merged data frame to .csv got the visualization part. I did this as I was having real difficulty with pie charts. exporting my data frame and adding it back resolved this issue and adding it Step-19 Adding %matplotlib notebook to view graphs. I was concerned my charts would not show up in notebook Step -20 Importing renamed merged .csv Step – 21 Creating Line Charts to visualize Daily Case Volume, Cumulative Case volume, Vaccine Dose rollout, Hospitalizations, and ICU volumes Step – 22 Creating Bar Charts to visualize Daily Covid Deaths, Cumulative Deaths, Daily Vaccine Rollout Step-23 Creating Histograms to visualize Daily Deaths and Daily Cases Step- 24 Creating Scatter Plot to visualize Daily Deaths and Daily Cases Step 25 I needed to create new Data frames for pie chats as I needed to only select multiple Columns but a single row (751) Step-26 Creating Pie Charts to visualize infections by Gender, Infections by Age Group, Hospitalizations by Age group. Step-27 Creating Boxplots to Mean of Total Hospitalizations and Patients requiring ICU.

## Results

## Chart Description automatically generated Irelands Daily confirmed cases 2020 – 2022. This is a reporting showing the Daily Infection Volume in Ireland. With restrictions lifted you can see in 2022 Covid cases are increasing

## Chart, line chart Description automatically generated

## Cumulative Confirmed Covid Cases from 2020 – 2022 Irelands increase in cases in Q4 2021 – Q1 2022 is a worrying trend which may result in new restrictions being implemented

## Chart, line chart Description automatically generated

## Cumulative Vaccine Roll out 2020 – 2022 Report based on First and Second Vaccine Single dose ( J&J ) and Boosters. Worrying trend of Booster levelling off in Feb and Mar 2022 which means about 1/3 of people don’t have a booster shot which Covid cases are increasing.

## Chart, line chart Description automatically generated

## Hospitalization trend 2020- 2022 Report shows the cumulative number of Hospitalizations . We are seeing a vasr increase of hospitalizations in Q1 2022.

## Chart, line chart Description automatically generated

## Hospitalizations requiring ICU.

## This is good news, In 2022 while we have an increase in Confirmed Cases and Hospitalizations, The number of people requiring ICU has been in decline.

## Chart, histogram Description automatically generated

## Covid deaths per day. Based on Q2 2022 results from other reports it looks that Deaths are in falling. You can also see from July 2021 where the HSE hack (ransomware) caused reporting issues from cases in Hospitals.

## Chart, bar chart Description automatically generated

## Here you can see the Vaccine Rollout and when they occurred, First , Second, Single dose and Boosters.

Chart, pie chart

Description automatically generated

Infections based on Gender 2020 – 2022 .

Chart, pie chart

Description automatically generated

Infections based on Age Group

Chart, pie chart

Description automatically generated

Hospitalisations based on Age Group

Chart, scatter chart

Description automatically generated

Scatter Chart of Dose 1 & Dose 2 Infections

Insights  
 (Point out at least 5 insights in bullet points)

* There were more Females than Males infected with COVID 19
* While Covid Infections and Hospitalizations are going Up, ICU and Deaths are going down
* Where there was a public exception of getting fully vaccinated, it seems that the interest of getting the booster has dropped of while infections have increased
* Over 55s made up 15% of Infections yet the same age groups made up 60% of hospitalizations.
* Based on all the data from 2020 -2022 . Vaccines are stopping people going to ICU or Dying from Covid, but Restrictions reduce the chance of contracting the virus based on 2020 to 2021 data.

Machine Learning   
Based on the data, Regression would suit this type of data. You can do Date and Volume analysis to Identify what restrictions would be required based on a threshold of infections, hospitalizations, and deaths.

We already know what restrictions were in place that worked to reduce infection volumes so this would give guidance to the HSE and Government on the timing to implement restrictions again.