Database and Analytics Programming

Saranya Varshni Roshan Karthikha StudentId: x20154801
Sachin Muttappanavar StudentId: x20144253
Sanket Sonu StudentId: x19206071
MSc in Data Analytics
National College of Ireland Dublin, IRELAND
URL: www.ncirl.ie

***Abstract*—This document is a technical write-up about sequence of data analysis made on three COVID-19 data sets utilizing python programming language. The analysis is carried out predominantly with the AWS cloud services, MongoDB, IBM cloud and Dagster Pipeline and we will take a detailed look at these as well as we proceed with main content. *CRITICAL: While experimenting with the analysis code built, do not perform few operations like DB connection termination and extracting new API data set as those results might differ from existing results.**

***Index Terms*—AWS, S3, MongoDB, Pandas,RDS MySQL, Dagster Pipeline, IBM cloud, Python, GIT**

I. Introduction

Our world is facing recent outbreak of Novel Corona Virus, which was termed as Global Pandemic by World Health Organization. Analysing such data sets of COVID-19 would help us identify hidden insights about the same and guide us to confront this recent sensation with more of prepared resources and well planned strategies.

Our population is badly in need of vaccination to fight against the virus. So, world's top drug regulators have approved two vaccines named Covishield and Covaxin, even though their clinical 3rd phase trail is still ongoing. We were motivated to perform this analysis because of the panic that is spread among us, which is caused due to the mortality rate and the fact that there is no defined or well experimented drugs or medication to treat it so far. All three data sets and their visualization are chosen in relevance to COVID-19. These data sets depict the total active cases and total death cases, total recovered cases, vaccine progress, stringency index and GDP capital variations in different countries.

Some series of steps performed while building this data analysis experiment are the storage of data in AWS S3 service ,reading the same as data frame ,proceeding with the pre-processing in MongoDB cluster and then pushing the processed data into AWS RDS MySql cloud based database with individually visualized results. Further the data sets are combined based on the common attributes and visualized from IBM cloud DB2. This project can be found on GitHub repository (https://github.com/NCI-Projects/DatabaseAndProgramming_Project).

II. Project Specification And Requirements

The primary goal of this data analysis is to create a cloud based storage for data set and access them for analysis through a programming language like python. The storage of data set is made efficient by utilizing AWS S3 (Simple Storage Service). This storage is made reliable by having a global access to the data set rather than using a local storage device. With the help of a NoSql Database, the semi-structured data set is being analysed further. Data pre-processing is a vital transformation done to the data set to make it suitable for data analysis. The whole data flow is being composed into a pipeline which makes it easy and efficient to gather the data from different sources and execute the flow without any interruption. Once the major analysis and data retrieval is complete, visualization of the same is done to identify different patterns and trends that have occurred.

In order to achieve the above mentioned project specification, following requirements are to be planned and included into the flow of data.

A secured, durable and scalable object storage like Amazon's Simple Storage Service is chosen which allows adequate number of get and put requests for our analysis. Handling dynamic schema or semi-structured data is done with a noSQL DB like MongoDB and a pymongo utility was included to query the database using python. Also, the second database is Amazon's relational database service (MySql) to analyse on the processed structured data set. Processes like data cleaning, data transformation, data reduction and data quality assessment were performed to make the data more eligible for analysis. Combination of all three data sets is being made based on country and analysed in IBM cloud DB2 database. The whole flow of data from S3 storage to the data pre-processing in IBM cloud was composed into two pipelines using Dagster utility.

III. Data Flow Architecture

The data flow is architectured in such a way to utilize all possible benefits from AWS ,other cloud services and from the open source databases. **Figure** 1 illustrates a complete flow of data from the initial local storage till data visualization. We have implemented pipeline for automatable tasks in the data flow. This helps in complete data orchestration in the flow of data analysis. All three data sets are stored in a cloud based object storage S3 bucket named 'Slark', which is also secure

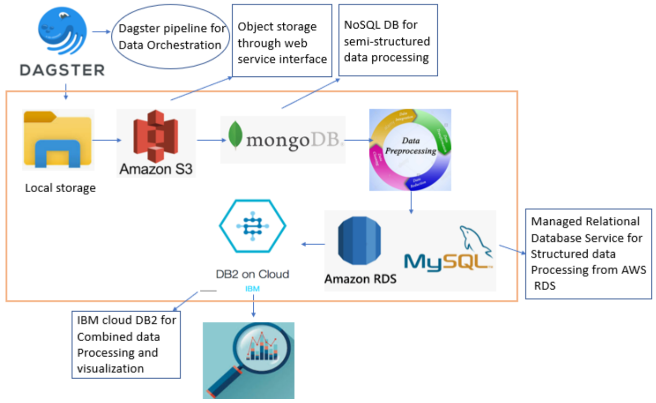

Fig. 1. Data Flow Architecture

and scale-able. Data to the mongoDB database is sourced from AWS S3 where a cluster is created named 'Covid_Collection' that contains different collections which inturn stores the data set from S3 storage. Later, creating dataframe from the list of 3 elements using pandas is performed. Data pre-processing of our data samples is to handle missing data (i.e,) empty strings and tuples, noisy data is the data cleaning. Normalizing (identifying and removing duplicates) and decentralizing data is data transformation. Once the pre-processing is complete, data is again back to its dataframe type and pushed to AWS RDS MySql database. This helps in storage and processing of structured data. Further all 3 data sets are combined with a common column country for combined visualization. Individual visualization of all three data sets is also performed.

IV. Related Work

In this research, Artificial and Big data applications are utilized to find the way to control spread of corona virus and protecting lives from COVID 19. They have achieved through building AI and Big data applications to detect and diagnosis, virus spread tracking and predicting the corona case.

In this paper, machine learning techniques are used to develop models to predict the effect of corona virus attack. They classified the symptoms by tree based algorithm. Built model was able to classify whether person symptoms are low, medium, high, no COVID.They recommended to work on pre-processing for future work to get better insight from data.

V. Methodology

A. Data Source 1:

This dataset is sourced from Kaggle. It has 9576 rows and 15 columns. This data consists of Vaccine dose information from different countries during the time period of 12 Dec 2020 to 3 Apr 2021. It has variables 'country', 'iso_code', 'date', 'total_vaccinations', 'people_vaccinated', 'people_fully_vaccinated', 'daily_vaccinations_raw', 'daily_vaccinations', 'total_vaccinations_per_hundred', 'people_vaccinated_per_hundred', 'people_fully_vaccinated_per_hundred', 'daily_vaccinations_per_million', 'vaccines', 'source_name', 'source_website'.

B. Data Source 2:

Data Description: This data set consists of latest COVID - 19 cases related data across 39 countries provided by Zermelo Technologies. This dataset is picked up to study effect of corona virus on people in different nations. We have scrapped this data from RapidAPI website[3]. GetHistoricalDataForCountryBetweenDates API is used to gather the COVID 19 data. This API returns Datewise dailydeceased, totaldeceased, dailyconfirmed, dailyrecovered, totalconfirmed, totalrecovered count of COVID 19 cases for provided start and end date for mentioned country. We have fetched data with start as 1st December 2020 and end date as 4th april 2021 for 39 different countries. Data-set consist of 4990 records and 9 columns. Using this dataset we analysed how COVID 19 virus affecting people on daily basis in different countries.

C. Data Source 3:

It is the latest corona virus world tracker with total count of 26423 records in it. This is sourced from Kaggle repository. Selection of this dataset was made due to its diverse attribute collections (in total 59 columns) like the icu patients, icu admissions made, stringency index (A composite measure based on nine response indicators including school closures, workplace closures, and travel bans, rescaled to a value from 0 to 100, where 100 is strictest).gdp per capita (A metric that breaks down a country's economic output per person and is calculated by dividing the GDP of a country by its population) and so on. The whole list of attribute information and its types can be viewed in the code output.

Data Source from Web : Latest Corona Virus World Tracker https://www.kaggle.com/codesagnik/latest-coronavirus-world-tracker

D. Data gathering and Handling activities

Dataset 1: We have sourced this dataset from Kaggle: https://www.kaggle.com/gpreda/covid-world-vaccination-progress?select=country_vaccinations.csv This dataset includes World Vaccination Progress based on Country and dates. This dataset contains vaccination observations from period of 12th Dec 2020 to 3rd Apr 2021.

Dataset 2: As mentioned in data description this dataset is scrapped from RapidAPI. We have used Python 'request' library to hit GetHistoricalDataForCoun API end point with request consisting of start date and end date for specific country. Json response from endpoint is saved in csv file using python library 'csv'.

Dataset 3: With reference to the data description in the previous section, this data is sources from an online data repository in relevance to COVID. Total data set count was around eighty thousand records. later, it was filtered for our specific time span for analysis purpose. This file was downloaded in csv format and read into data frames using csv library.

E. Data Processing

Every data sample or observations are defined with different characteristics that is with the help of different attributes.This pre-processing is an essential step before we build a model or perform analysis. Also, this will help in understanding our data sample better before processing results.

Few steps like identifying the data type of features, replacing missing data(empty cells in the sample are either replaced with 0 or manipulated with mean, median or mode) are required to bring out better information from data-set. This in turn boosts consistency in data.

Identification and removal of duplicate tuples are performed. This will remove the noisy data that occurs and distracts the analysis. This can be applicable for some columns that delivers meaningless information or irrelevant data keys.

Finally, clustering of the dataframe after the pre-processing as above mentioned resulted in a clean data sample that can be made more eligible for analysis and visualization.

F. Choice of technologies used

We have used a highly cross-functional python programming as it is one of the best languages when it comes to statistical analysis. There are also Python packages that are specifically tailored for certain functions, including pandas, NumPy, plotly, folium, seaborn, matplotlib, boto3, mysql, Pymongo, csv and json . Also, It is more user friendly and quick to learn. In this work we have availed functional programming techniques of python like lambda expressions, functions, exception handling for data processing and visualization.Also libraries of python to scrap data.

With regards to storage we have used three cloud hosted databases.

1) AWS S3 Cloud Storage - An Amazon S3 bucket is public cloud storage available in Amazon Web Services (AWS). Amazon S3 buckets are like file folders, store objects, which consist of data. We can save data into multiple formats here. First, we will create an AWS S3 Cloud Storage and named the bucket 'slark'. We created "SECRET and ACCESS Keys" from "Your Security Credentials" section by navigating to the IAM (Identity and Access Management). We need these "SECRET and ACCESS Keys" to access our AWS S3 Cloud Storage. These "SECRET and ACCESS Keys" will work as a username and password to approve the connection using python. So, we are first pushing all 3 .csv files into AWS S3 Cloud Storage. We will access AWS S3 Cloud Storage using "SECRET and ACCESS Keys", now we have to select "S3" bucket, which we already created in Cloud Storage from the AWS website. Now, we will push all 3 .csv files directly into AWS S3 Cloud Storage and named them 'CovidData_1','CovidData_2' and 'CovidData_3'. Again, we will access AWS S3 Cloud Storage using "SECRET and ACCESS Keys". We will create 3 DataFrames using pandas and pull all data from AWS S3 Cloud Storage to those 3 DataFrames.

2) AWS MongoDB cloud - MongoDB is open source document oriented database and NoSQL database. All schemas and

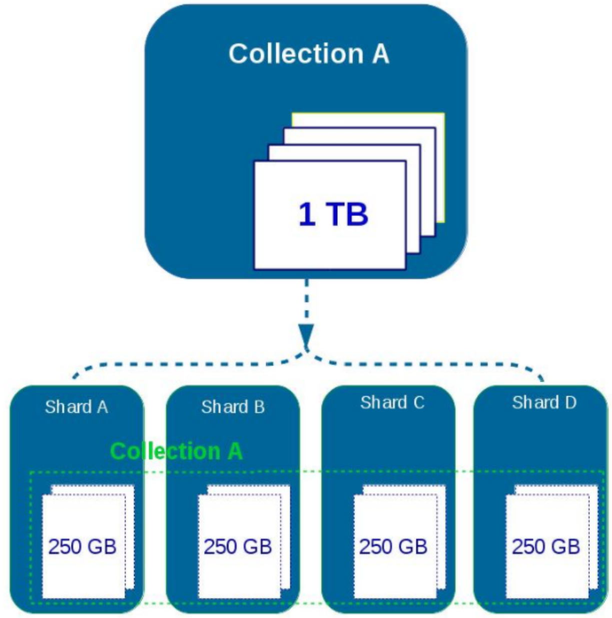

Fig. 2. MongoDB Shared cluster

records are stored in documents using javascript object notation syntax. It is very easy and fast to maintain and handle with MongoDB database compared to traditional databases.In this work we employed AWS cloud hosted MongoDB instance.We have chosen Shared Cluster plan dedicated for deploying small projects which is very secured, easy to maintain and scale our project. We have used shared clusters in which collections are distributed across different shrads nothing but machines as shown in **Figure** 2. This would in-turn helps in achieving high scalability and better execution in read and write activities. In our project we have used MongoDB database to store all three semi structured, unprocessed initial data-sets. To establish communication with MongoDB we have utilized functions of pymongo library.

3) AWS RDS MySql - An Amazon RDS MySQL Database is a public cloud-based server, which can be used to store data. These servers are scalable and cheap. User can take benefits from features like backups, monitoring, software patching, scaling and replication. User can select, different tier for Database server performance. Amazon RDS MySQL Database has the option to select a CPU for the performance boost. However, we are using the Free Tier server for our project. First, we must create a Database into AWS RDS MySQL Database. For this, we access the AWS website to create a cloud database server. We are selecting Ireland Server for AWS so we get a good ping, and our data push and pull will not face many delays. We are selecting MySQL Free Tier for this project. While creating the server, we need to make few changes. We must provide "Credentials" to access our AWS RDS MySQL Database, those are 'user' and 'password'. After, creating AWS RDS MySQL Database, we must change the Security Group which was assigned by default while creating

the server. By default, the Security group has a set of limited access. We need to create a new Security Key by accessing: EC2 -¿ Security Group and while creating a new Security Group, we must create a new "Inbound rule" and have to select "All Traffic" from the "Type" dropdown and then we have to select "Anywhere" from Source dropdown. Later, we need to modify AWS RDS MySQL Database settings and assign this new Security group to the database. First, we will connect to our AWS RDS MySQL Database Server using the hostname (Endpoint) that we can find from the "Connectivity security" section of the database server. Now, using python, we will connect to AWS RDS MySQL Database Server by providing 3 credentials, such as hostname, user and password. After, successful connection, we will create tables for all 3 DataFrames. We will push those 3 DataFrames into 3 tables. Again, we will connect to AWS RDS MySQL Database Server using credentials and will access the database. Now, using Pandas we will pull the data by directly creating 3 new DataFrames. Now, we will use these 3 DataFrames to Visualize each dataset separately in details.

4) IBM cloud DB2 - IBM Cloud services are faster, secured. IBM providing around 170 products and services on its cloud platform. In this work we availed DB2 service provided IBM cloud which is easy to maintain and highly performant relational database. DB2 database is developed by IBM. Initially DB2 database is developed to work with relational model where data is stored in table format. Later enhanced to give support for non relational data like XML, Json. We initiated DB2 instance on IBM cloud which is free version with 200MB of storage. We used this service to store the combined structured data in traditional table format. This DB2 database will be up and running always we can access it anytime via TCP/IP protocol.

For the Base file Storage we have picked Simple Storage Service offered from Amazon Web Services to form a structure where the data must flow throughout the pipeline only on cloud.

We leveraged all cloud services for our work. Benefit's of using cloud platform are cost saving, high speed, easy backup and restore, reliability, scalability, simple to collaborate.

VI. Combining All 3 Datasets

All three data-sets are combined by common countries. Columns present in final data-set are TOTALCONFIRMED,PEOPLE_FULLY_VACCINATED_PER_HUNDRED, EXTREME_POVERTY, TOTALCONFIRMED, POPULATION_DENSITY, STRINGENCY_INDEX, HOSPITAL_BEDS_PER_THOUSAND, DAILY_VACCINATIONS, DAILYCONFIRMED, DAILYRECOVERED, DAILYDECEASED. There are 33 records in combined data-set and is stored on DB2 ware house on IBM cloud platform.

VII. Dagster Pipeline

Dagster is a data orchestration tool. It helps us define pipelines in terms of the data flow between logical components termed as solids. Each component in a pipeline represents a task which is annotated with a solid that takes input of any type and returns the output for the next operation or solid. We have created 2 pipelines for our data flow.

1) Pipeline1: To upload a data file to S3 object storage in S3 as depicted in **Figure** 14

2) Pipeline2: This performs rest of the flow and analysis as represented in **Figure** 15

Execution of each pipeline and its solids can be driven by the Launch execution in the Playground tab as shown in **Figure** 13

Once the execution of pipeline is successful, all the solids inside that pipeline will be executed and the console output will be displayed in the info tab of during execution.

VIII. Results

A. *Dataset 1 :*

Figure 5 Shows the Total vaccination of India. We can observe from the plot that daily vaccination is increasing day by day on 3rd Apr 2021, a total of 75.97 million people are vaccinated. **Figure** 6 Shows the Total vaccination of the USA. We can observe from the plot that the USA has a good vaccination program, and on 3rd Apr 2021, a total of 161.88 million people are vaccinated. **Figure** 7 Shows the relation of Total vaccination in both India and the USA. It is clearly observed that the USA has a very good Vaccination Program. **Figure** 8 Shows Vaccine Combinations which are mostly used by the greatest number of Countries. It is clearly observed that Oxford/AstraZeneca is the most used combination for most of the Countries. "Moderna, Oxford/AstraZeneca, Pfizer/BioNTech" is the second most combination of vaccine which is used by most countries. **Figure** 9 Shows Vaccine Combinations which are used most by overall all countries. It can be demonstrated from the plot that "Johnson Johnson, Moderna, Pfizer/BioNTech" are the most used combination and a total of 151.28 million people are vaccinated with this Vaccine Combination. **Figure** 10 Shows the top 10 countries with the most vaccinations till 3rd Apr 2021. It can be clearly demonstrated from the plot that the USA has the most successful Vaccination program, and a total of 151.28 million people are vaccinated. China is in 2nd position with total vaccination of 115.47 million. India is in 3rd position with 67.27 million vaccinations. The UK is in 4th position with 35.30 million vaccination. Brazil is in 5th position with 19.28 million vaccination. Turkey is in 6th position with 16.44 million vaccination. Germany is in 7th position with 13.64 million vaccination. Indonesia is in 8th position with 11.44 million vaccination. Russia is in 9th position with 11.39 million vaccination. France is in 10th position with 11.13 million vaccination.

B. *Dataset 2 :*

We first explored active cases distribution from date 01-12-2020 to 04-04-2021. **Figure** 3 is the visualization for active cases distribution. We can see that total confirmed cases are increasing from past three months. By the end of march month, overall confirmed cases are near to 45Million.

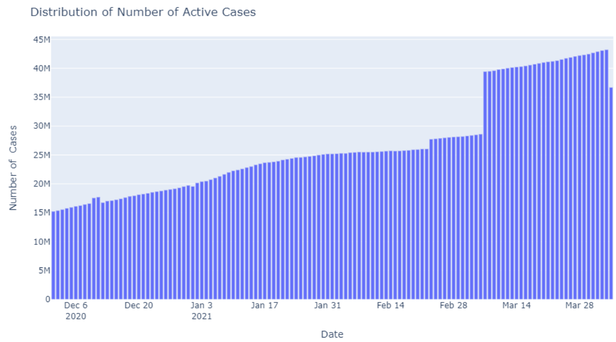

Fig. 3. Active cases distribution

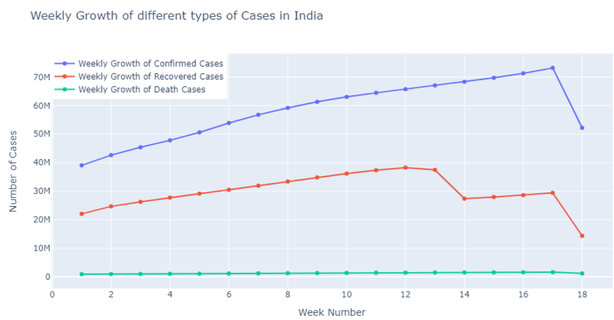

Fig. 4. Weekly growth of cases

Next we visualized weekly increase in different cases across all countries in **Figure** 4. From this graph we can see that, confirmed and recovery cases are increasing at same rate but count of confirmed cases per week are bit higher than recovery cases. Though confirmed cases are increasing, death is constant throughout. This implies corona virus is not deadly virus, we will look into mortality rate in the next graph. In **Figure** 11 we can see recovery rate was almost constant till February month end, thereafter reduced. May be this is sign of second wave of corona virus. While mortality rate is fluctuating between 2.0 to 2.3 percentage from December to April. This gives evidence that deaths are less though corona virus is spreading over world.

In this **Figure** 12 we are showing line graph of total confirmed cases across different countries over last 4 months.If we observe line graph of India in month of march, total confirmed cases are increasing rapidly. That shows second wave of corona virus is started in India.

Here in **Figure** 16, we have visualized deaths rate across different countries on world map using **Folium** library. From graph it is clear that corona virus cases are high in India and United states as they are dark red.

C. Dataset 3 :

Figure 26 illustrates Visualizing new cases count progress with respect to date(Comparing the new cases per million and new cases smoothed per million). Data smoothing is done by using an algorithm to remove noise from a data set. This allows important patterns to more clearly stand out. **Figure** 27 and **Figure** 28 depicts total ICU patients in Ireland and total death count per million for the same country. This compares the correlation between ICU patients and the death count. It can be observed that the pattern between both the graphs are similar.

Figure 29 is a jointplot that can shows how intense the death count along the timeline.

Figure 30 explains how different continents work in progress with total vaccination. Since continent is a categorical variable, stripplot is chosen and it can be observed that Asia and North America are having significant vaccination progress.

Figure 31 displays top 10 countries with highest stringency index when compared to other countries in the same data set. It is evident that Ireland is the fourth country to top the list.

Figure 32 shows the correlation between the total cases and the stringency index. It is clearly visible that with high stringency the total cases count is relatively less and with a lower stringency, later total cases count is comparetively high.

Figure 33 depics the average GDP per capita of a country where the average GDP remains to fluctuate minimal during pandemic.

D. Results for Combined Dataset :

We have combined all three dataset by country name and processed it to visualize. From **Figure** 18 and **Figure** 19 we can observe that poor countries are not able to manufacture COVID vaccines and vaccinate citizens. Malawi, Rwanda, Uganda are countries with highest poverty rate where vaccination rate per hundred rate is low.

Stringency rate is measured based on shop closure, travel bans, school closure other kind of lockdown. Stringency index is high in Australia and Singapore as per **Figure** 21. Singapore is having high population density(**Figure** 20) they should have high stringency rate to stop corona spread. Due to high restrictions in these countries , total confirmed cases are low compare to other countries(**Figure** 22).

Daily vaccination rate on an average is high in United states followed by India, Italy and China as per **Figure** 23. However, daily confirmed, death cases, recovery cases are at the same pace in these countries(**Figure** 24 and **Figure** 25). This indicates that vaccine is not so strong in preventing people from getting infected.

IX. Future Enhancements

This project is getting bigger day by day as COVID cases are increasing because of second wave. In future, we can deploy Machine Learning models, which can provide better results. We can use SMOTE to balance our imbalanced dataset.

We can use Neural Networks and Deep Learning to predict the results based on variables like cells, blood genes, age, health condition, virus genome type. We can predict a good results after getting those variables which will eventually help us predict whether a person is COVID positive or negative.

REFERENCES

[1] *Covid Symptom Severity Using Decision TreeNaim Rochmawati; Hanik Badriyah Hidayati; Yuni Yamasari; Wiyli Yustanti; Lusia Rakhmawati; Hapsari P. A. Tjahyaningtijas*

[2] *A Survey on the State-of-the-Arts: Artificial Intelligence (AI) and Big Data for Coronavirus (COVID-19) Pandemic*, *Quoc-Viet Pham, Dinh C. Nguyen,Thien Huynh-The, Won-Joo Hwang, Pubudu N. Pathirana*, *FOUNDER, ETHEREUM & ETHCORE*, *GAVIN@ETHCORE.IO*

[3] https://rapidapi.com/ZermeloTechnologies/api/covid-19-world-data-by-zt/detailsRapidAPI , *Quoc-Viet Pham, Dinh C. Nguyen,Thien Huynh-The, Won-Joo Hwang, Pubudu N. Pathirana*, *FOUNDER, ETHEREUM & ETHCORE*, *GAVIN@ETHCORE.IO*

[4] *Finding an Accurate Early Forecasting Model from Small Dataset : A Case of 2019-nCoV Novel Coronavirus Outbreak" by Simon James Fong, Gloria Li, Nilanjan Dey, Rubén González Crespo, Enrique Herrera-Viedma. International Journal of Interactive Multimedia and Artificial Intelligence 6.1 (2020): 132-40*,

[5] *A review of mathematical modeling, artificial intelligence and datasets used in the study, prediction and management of COVID-19 by Youssoufa Mohamadou, Aminou Halidou Pascalin Tiam Kapen. Applied Intelligence volume 50, pages3913–3925(2020).*,

[6] *Visualising COVID-19 Research by Pierre Le Bras, Azimeh Gharavi, David A. Robb, Ana F. Vidal, Stefano Padilla, and Mike J. Chantler. Strategic Futures Laboratory, School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh, UK.*,

[7] *Role of biological Data Mining and Machine Learning Techniques in Detecting and Diagnosing the Novel Coronavirus (COVID-19): A Systematic Review: A Systematic Review" by A. S. Albahri2 Rula A. Hamid3 Jwan k. Alwan4 Z.T. Al-qays5. 14 March 2020 /Accepted: 27 April 2020/Published online: 25 May 2020.*,

[8] *by Umair Qazi, Muhammad Imran and Ferda Ofli.*,

[9] *Statistical Explorations and Univariate Timeseries Analysis on COVID-19 Datasets to Understand the Trend of Disease Spreading and Death by Ayan Chatterjee 1,*OrcID,Martin W. Gerdes 1 and Santiago G. Martinez 2. Sensors 2020, 20(11), 3089.* ,

[10] *A study on the quality of novel coronavirus (COVID-19) official datasets by Ashofteh, Afshina; * — Bravo, Jorge M. Statistical Journal of the IAOS, vol. 36, no. 2, pp. 291-301, 2020*

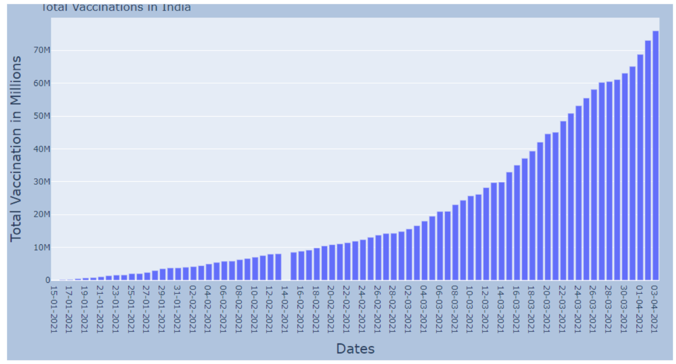

Fig. 5. Total Vaccinations in India

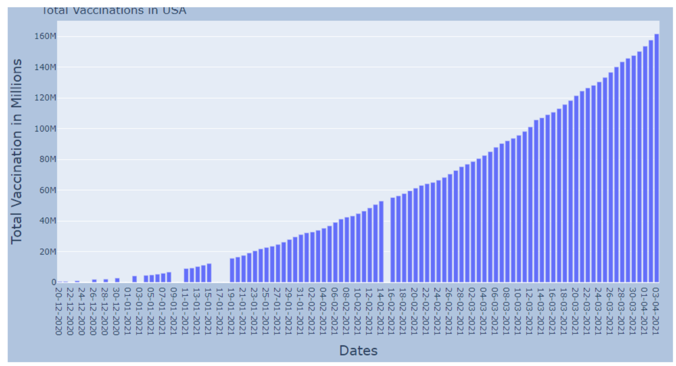

Fig. 6. Total Vaccinations in USA

X. APPENDIX (DATA VISUALIZATIONS PLOTS)

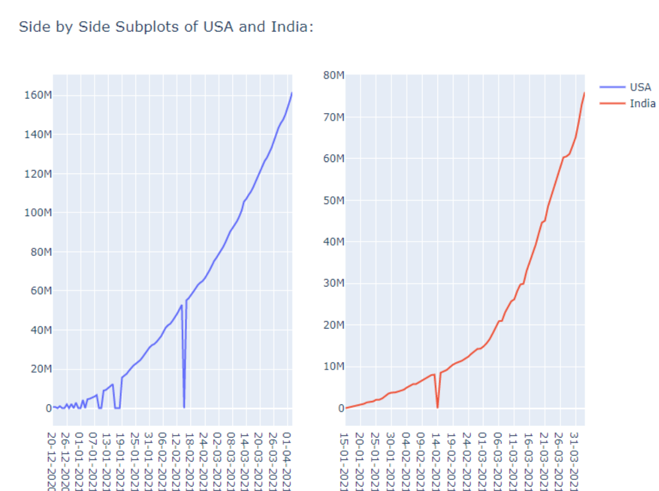

Fig. 7. Progress of Total Vaccinations for USA and India

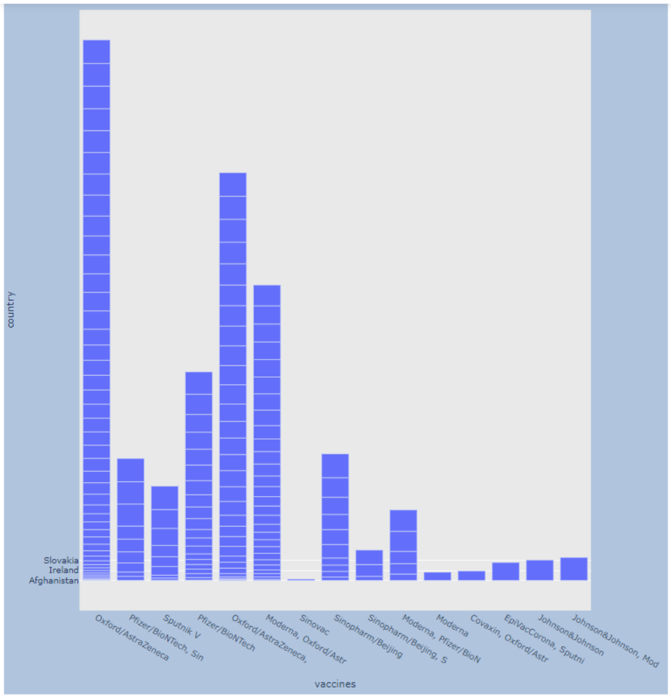

Fig. 8. Vaccine Combinations which are mostly used by most number of Countries

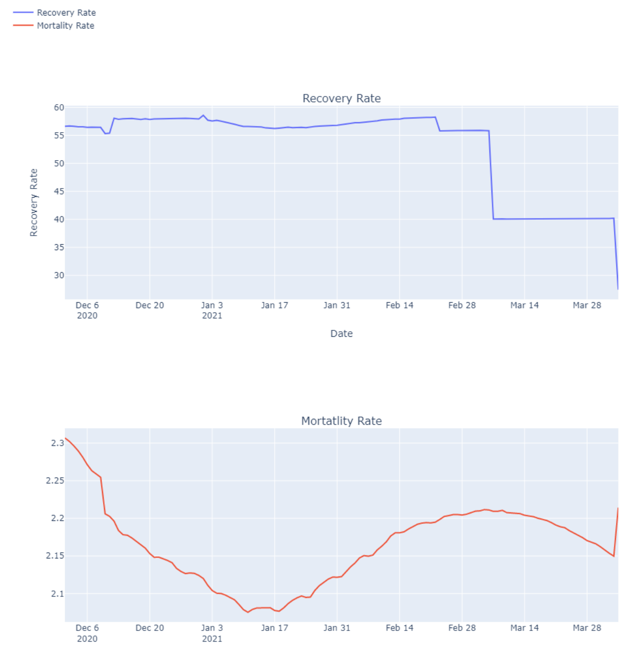

Fig. 11. Recovery and mortality rate

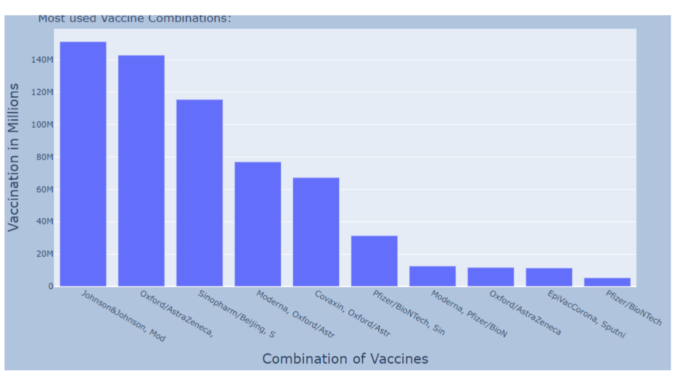

Fig. 9. Vaccine Combinations mostly used

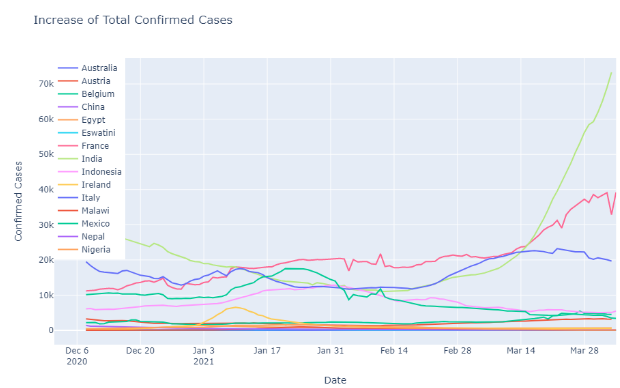

Fig. 12. Countrywise raise in corona cases

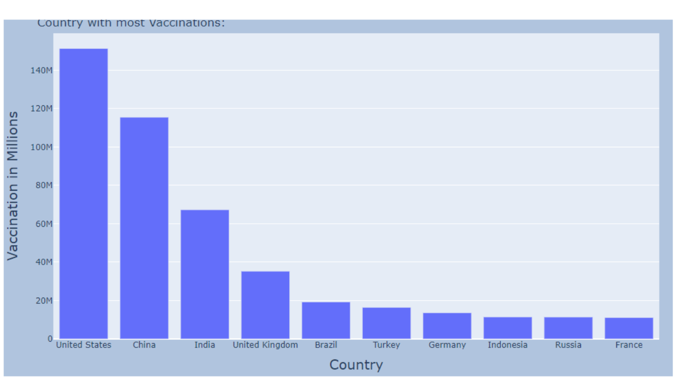

Fig. 10. Top 10 Countries with most Vaccinations

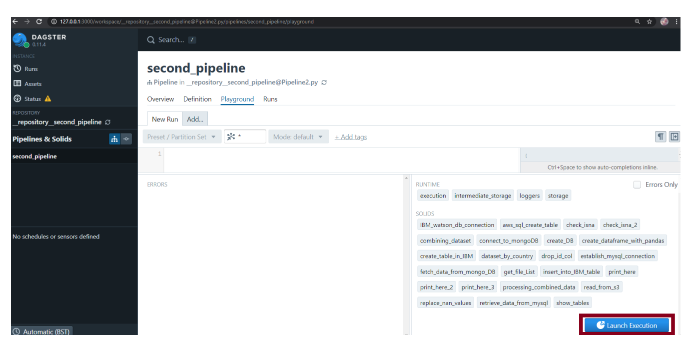

Fig. 13. Launch execution inside Playground tab

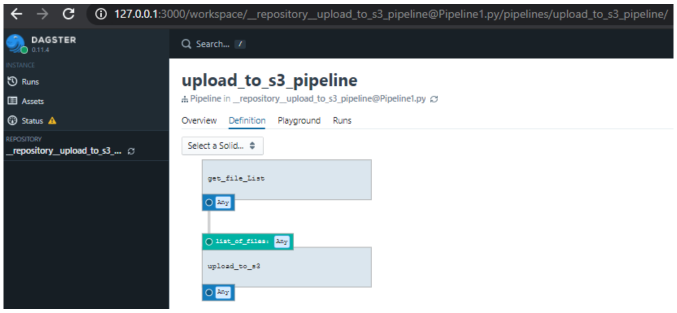

Fig. 14. Upload to S3 pipeline

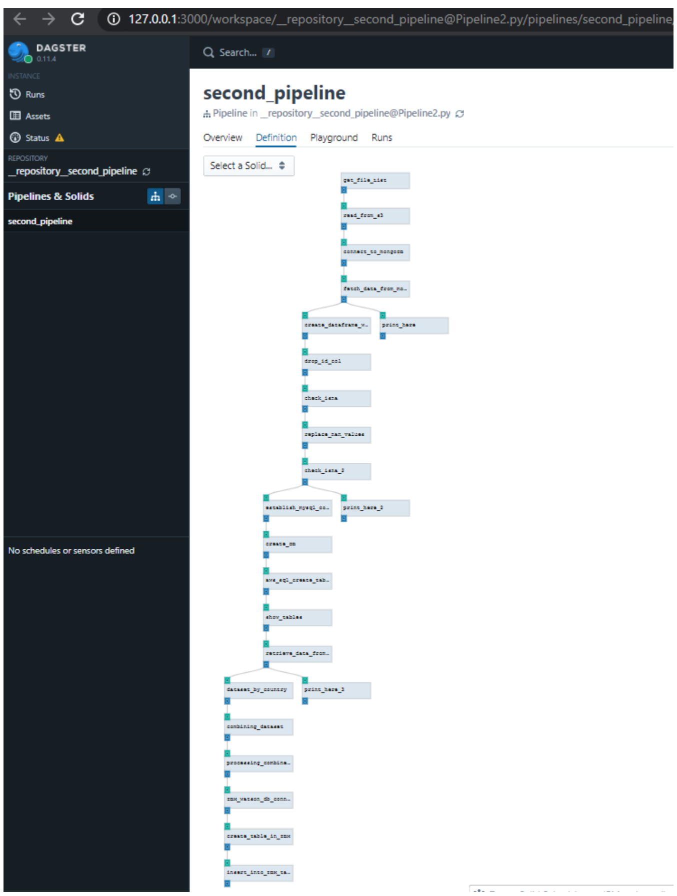

Fig. 15. Pipeline for rest of the flow

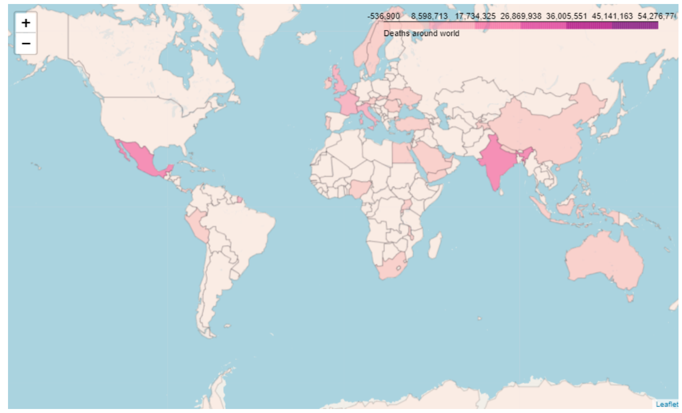

Fig. 16. World map with deaths across different countries

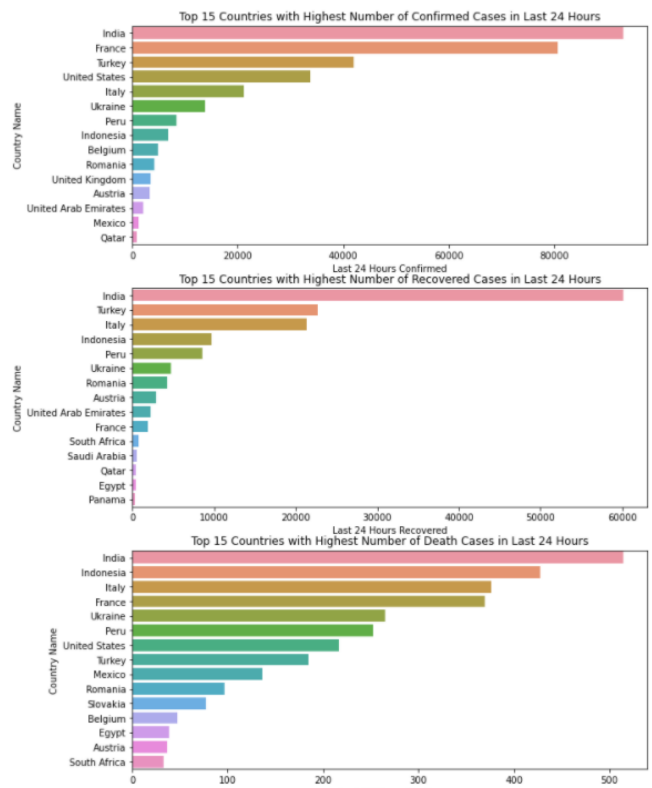

Fig. 17. Different type of cases in Last 24 hours

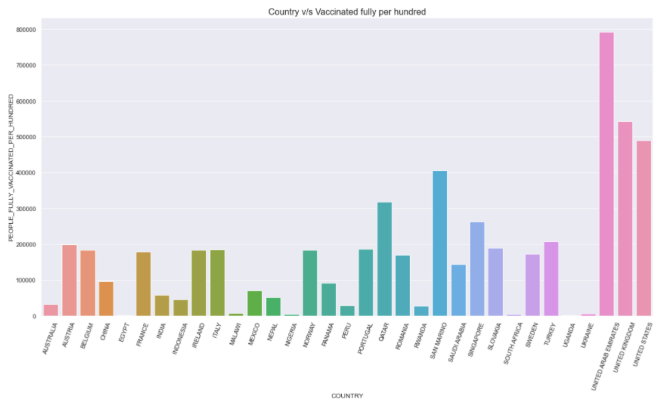

Fig. 18. World map with deaths across different countries

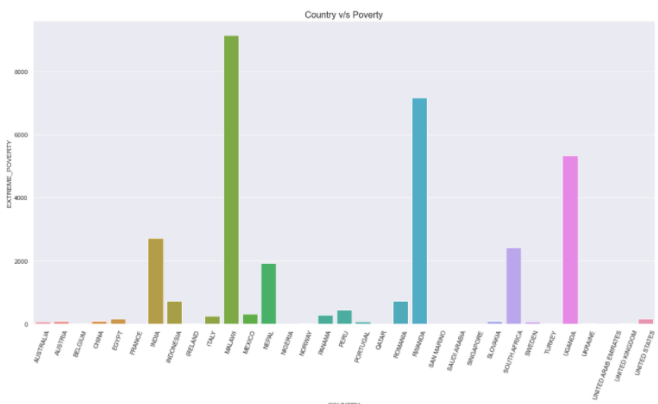

Fig. 19. Poverty rate of different countries

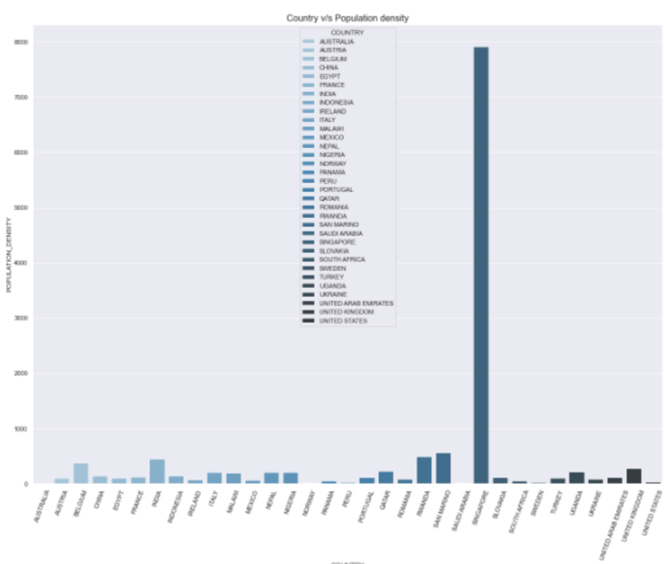

Fig. 20. Population density across different countries

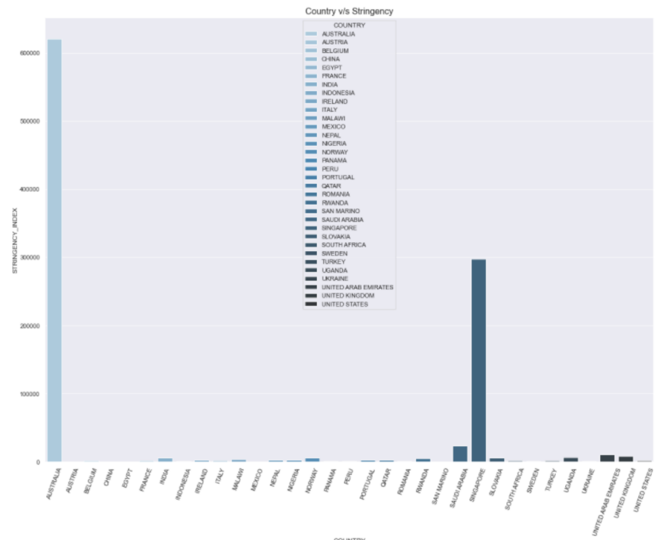

Fig. 21. Stringency Index across different countries

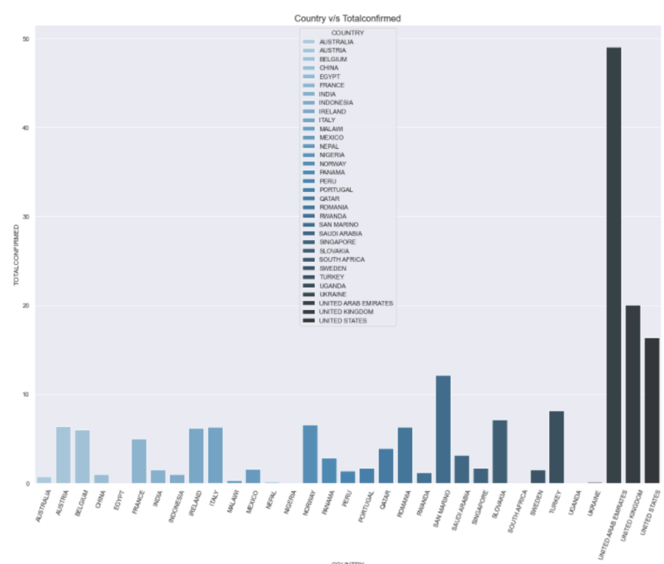

Fig. 22. Total confirmed cases in the

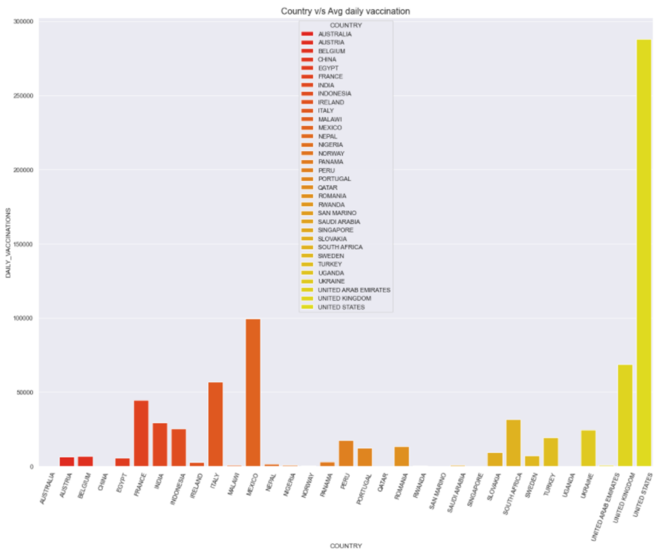

Fig. 23. Average daily vaccination count

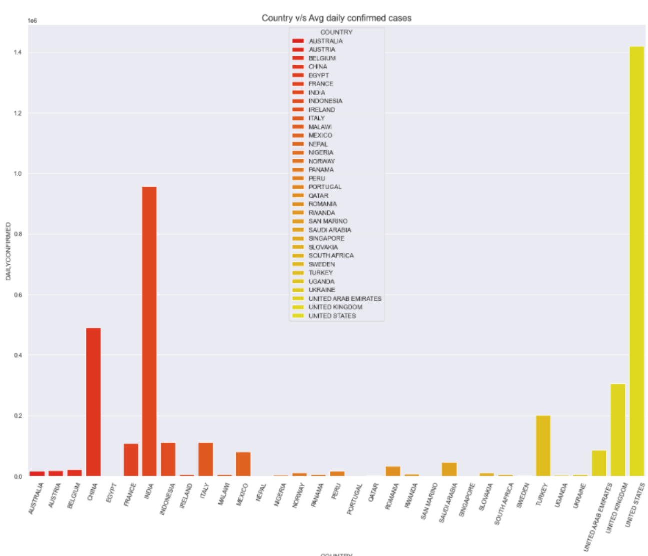

Fig. 24. Average daily Confirmed count

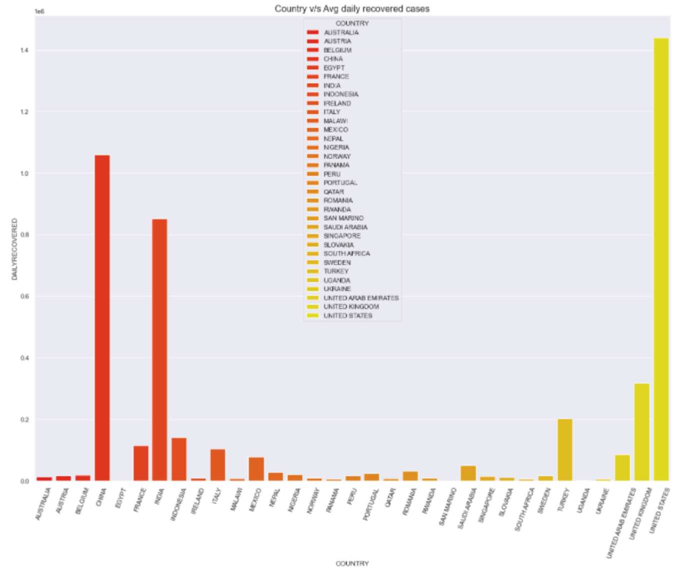

Fig. 25. Average daily recovered count

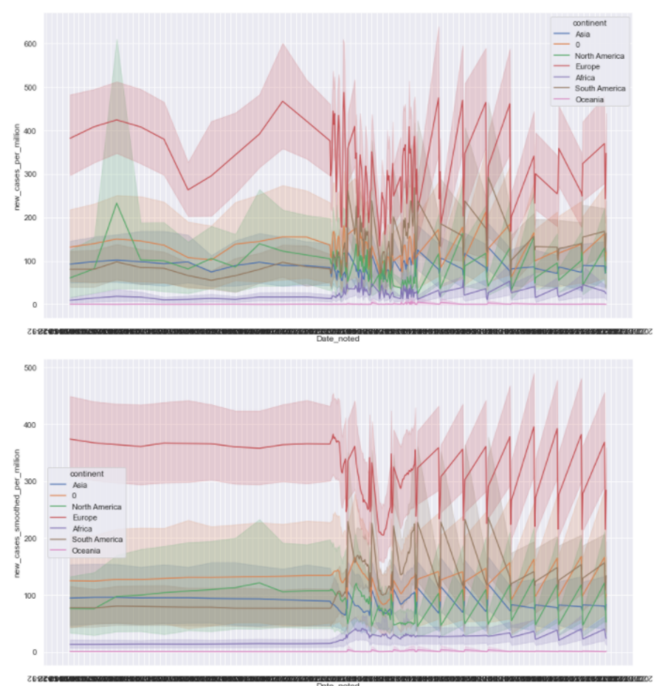

Fig. 26. New cases per million Vs New cases smoothed per million

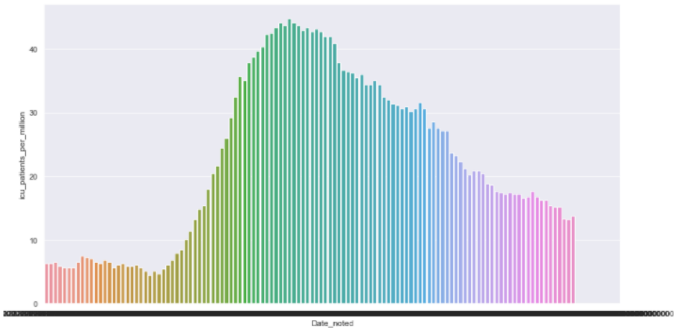

Fig. 27. ICU patients in Ireland

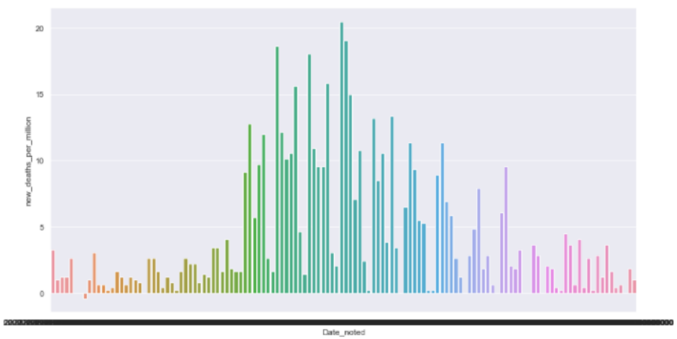

Fig. 28. Total Death count per million

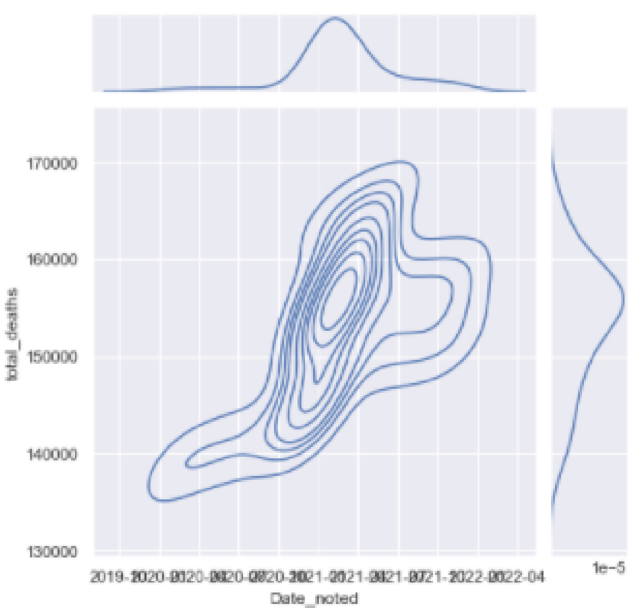

Fig. 29. Joint plot for total deaths

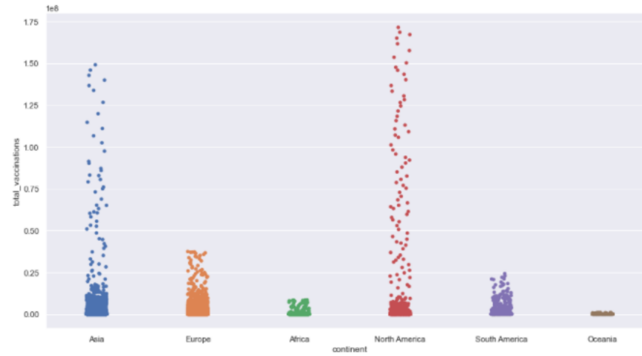

Fig. 30. vaccination progress

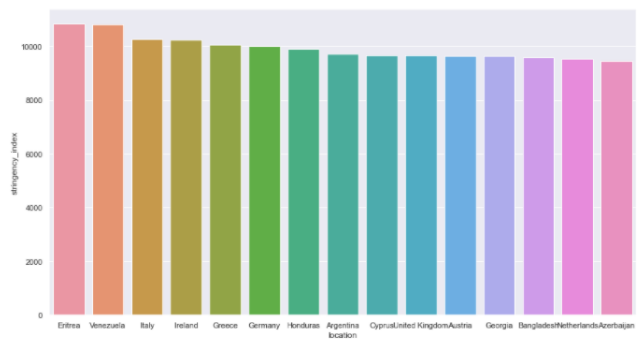

Fig. 31. Top 10 countries with highest stringency index

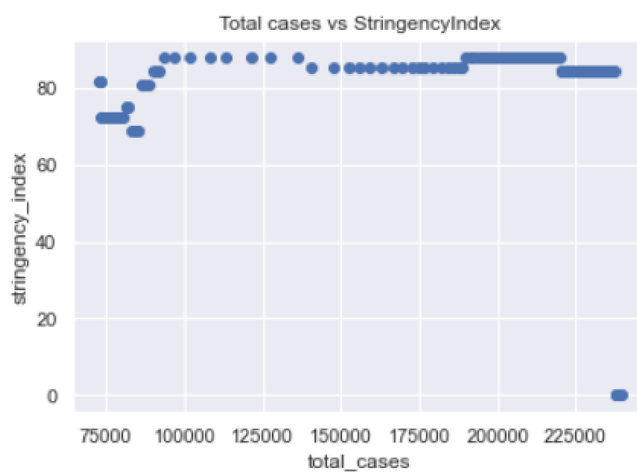

Fig. 32. Total cases Vs Stringency index

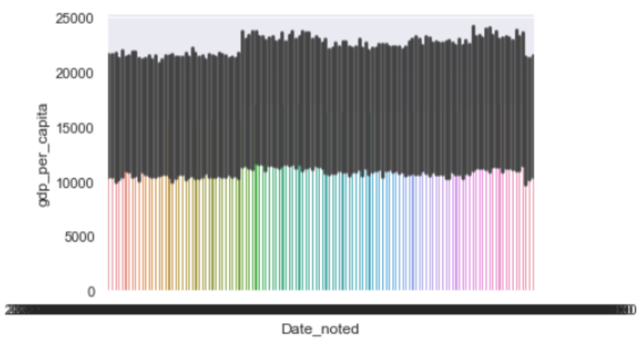

Fig. 33. gdp per capita