Interactive COVID-19 Dashboard

by Diya Anne Tharappan

Submission date: 09-Dec-2021 03:50PM (UTC+0530)
Submission ID: 1725392968
File name: IEEE_Conference-Paper_CovidDashboard-1.pdf (193.68K)
Word count: 1973
Character count: 10412

Interactive COVID-19 Dashboard

Diya Anne Tharappan
Department of Computer Science
PES University
Bengaluru, India
diya.star2002@gmail.com

Ghanashyam Mahesh Bhat
Department of Computer Science
PES University
Bengaluru, India
ghanashyambhat6@gmail.com

Govinda Madhava
Department of Computer Science
PES University
Bengaluru, India
gmbs.madhava001@gmail.com

Likitha P
Department of Computer Science
PES University
Bengaluru, India
likithap@pes.edu

Abstract: ***Owing to the spread of misinformation and false news about COVID-19 we created a dashboard of easily readable charts, graphs, and visualizations right from scratch from the publicly available John Hopkins Repository***

Keywords: Visualization, Covid-19, Dashboard, Vaccine, graph, python, Jupyter, PowerBI, data.

I. INTRODUCTION

Coronaviruses are from a family of RNA viruses in the order of 'Nidovirales'. They are significant human and animal pathogens that cause severe diseases. The word 'Corona' comes from the Latin word 'Crown' owing to its crown-like appearance formed by its membrane surface protein.

In late December 2019, there was a cluster of pneumonia cases in Wuhan, China. Investigations found that the disease was caused by a newly discovered Coronavirus and the disease was subsequently labeled 'COVID-19' which spread like wildfire from China to all corners of the world far and wide. The new Coronavirus, invisible to the naked eye, traveled rapidly from village to village, town to town, and country to country. It spread every time a person so much as sneezed, coughed, or even breathed. Soon WHO declared that the disease acquired public emergency and international concern. To stop the rapid spread of the deadly virus, Many countries plunged into a partial or complete lockdown.

The onset of the Coronavirus outbreak coupled with social isolation rattled many leaving them confused and frightened of the unknown. Mankind was aiming to combat a disease he had never seen before. It left most governments in an insecure and unreliable state. Communication formed a huge barrier. There was and still is, a great thirst for easily decipherable data and clarity on COVID-19 information.

Most of the uncertainty was heightened due to the spread of false information. Awareness about the disease was difficult to spread around as people lacked a comprehensive and easy-to-read source of inference.

Even news channels circulated cumulative stale data instead of up-to-date information which instilled fear in the hearts of many. The Spread of False news was propagated by politicians, celebrities, and various other famous figures on mass media which led to widespread distraught.

To equip the public with apt and reliable information in addition to drawing inferences from the available data, the three of us decided to create a trustworthy and interactive COVID-19 dashboard. We also wanted to study the trends in the data available to help discover some insights and verify the news we receive on a daily basis.

The project utilizes available public data sources, pools information, filters it, and makes it accessible through a dashboard of easily readable charts, graphs, and visualizations.

A dashboard is an effective means to display data as a well-crafted dashboard can, in essence, make a great deal of data be fitted into a single page. This mechanism ensures that information is relayed easily and what otherwise seems like an incomprehensible source of data is transformed into a web page of easily readable charts and graphs

II. METHODS/PROCEDURE FOLLOWED

Our project reads the data from a notable source and performs various operations on it with the help of Python programming.

The file thus generated through working on the data available with the use of python is exported to a Dashboard editor Power
BI where front-end operations are feasible. Power BI helps us generate a trusted and well-formatted COVID-19 Dashboard.

Followed by this, we unpivoted the data using a pd.melt() function to do the same. We added another column for dates in the data frame and collectively put all the dates under one column.

Cleaning the data:

We changed all the values to their corresponding data fields such as deaths, recovered cases, and tried to infer the number of 'NULL' values in our dataset. As we were going to be working with maps we did not want there to be any 'NULL' values and replaced them with other values using the 'fillna()' method.

Merging data:

Once we cleaned the data in all the data frames, we joined the three data frames to obtain the final file which we exported to Power BI for creating effective front-end designs. The visualizations created with the aid of Power BI helped us pave our way through creating a well-organized dashboard.

Another method we explored was converting our Jupyter notebook created into a stand-alone web application using a software called 'voila'. Both web applications served as effective interactive dashboards, however, the visualizations provided with Power BI were easier to work with and provided for a better readable user interface.

A. DATA COLLECTION:

In order to make our dashboard reliable, trustworthy, and up-to-date we searched exhaustively for an apt source of data. We eventually found the John Hopkins University Repository which had extensive data of COVID-19 cases from all over the world. We cited several sources and found that the data collected by John Hopkins University always incorporated the latest information and was undoubtedly accurate.

In addition to having a COVID-19 cases data set, we were also curious to know as to how the administration of vaccines affected the data we collected and if there was any particular trend we could identify. We used the data from covid.ourworldindata.org to obtain data related to the administration of vaccinations to track vaccination progress.

B. PROCEDURE:

For using Power BI

We used a Jupyter Notebook editor and Python code to read the data available right from the source, so that any changes made to the data such as the addition of new data, would reflect in the Dashboard we were going to build. We imported Python libraries such as NumPy and Pandas to work with the data.

We then read the data and split it into three data frames:- The Deaths, The Confirmed cases and The Recovered cases

For using voila

We proceeded with reading our data from the John Hopkins Repository and performed similar operations as in the above Power BI method except that we performed different operations on the data.

We wanted to infer the top 10 worst-hit countries all around the world and used an effective procedure to highlight the worst-hit countries. We sorted the data and plotted the first 10 data points.

We plotted a bubble chart to effectively represent the countries with the most number of confirmed cases

We plotted a line graph of number of confirmed cases versus dates and deaths.

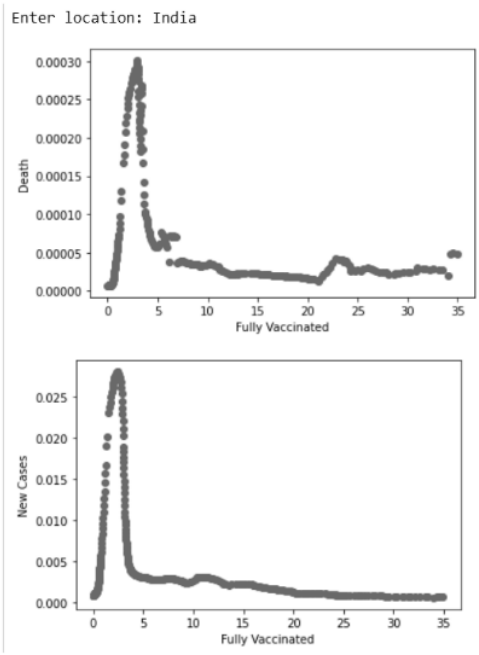
Tracking vaccination progress

We read the data from the covid.ourworldindata.org to obtain data related to the administration of vaccinations to track vaccination progress and then compute the death versus fully vaccinated cases for different countries all around the world.

We created a data input field and upon entering the country, the plot immediately aligns to the number of people vaccinated to the number of deaths in that country. For example, if we were to enter the value in the data field as 'Russia' we would immediately obtain a plot of the deaths versus the fully vaccinated people in Russia.

III. Inferences

Upon Tracking the Vaccination Progress we uncovered that there was a steady decline in the number of cases as the number of people who were administered the vaccine increased.



It was also observed that the daily cases of confirmed cases versus date showed a flatter curve than the cumulative data(which is probably the data portrayed in mass media). At a glance, the cumulative data shows an increase rather than a decrease which leads to the misinterpretation the covid cases are on the rise.

IV. Results

Results obtained using Power BI

There are three main components of the dashboard: the main summary and global map, trends and forecasting and the country/region breakdown and analysis.

A. Main Summary and Global map

The Main Summary entails a calendar that records the number of months up until the current date and changes data interconnected with it upon selecting a particular month

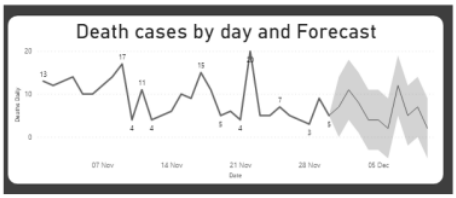
The Date section specifies the number of days from when the data was recorded. Each day can be mapped on the various visualizations

Country/ Region and Province/State specify the various countries which can be mapped onto various visualizations.

The global map highlights confirmed COVID-19 cases all over the world

Alongside the global map is displayed the confirmed cases and the deaths daily.

B. Trends and Forecasting



The above graph is a plot of deaths daily versus the date. We have used the Power BI forecasting feature to predict the likelihood of deaths with a seasonality of 15 and a 95% confidence level(as indicated by the shaded region)

We have also plotted cumulative death cases versus date.

Similarly we have plotted, Confirmed cases by day and Cumulative confirmed cases and by using the forecasting feature we can predict the confirmed cases for the next ten days(as indicated by the grey region)

C. Country/Region breakdown analytics

The Country/Region analytics is broken down into the confirmed daily cases, the death cases, the Recovered cases, and a new measure called 'CFR'- Case Fatality Rate. The CFR is given by dividing the sum of confirmed daily cases by the daily death cases. If the CFR is more than 0.04, the data is highlighted with a darker color.
Furthermore, if a particular country is chosen in particular, (in this example we have chosen Australia) it will display all the totals corresponding to the parameters in the ch

(ii)*Using voila*
Incorporating the top 10 worst-hit countries:-

We read the data from the John Hopkins Repository just as used in the above method and collect the deaths, the

confirmed cases and the recovered under different data frames. We then use Python code to traverse through the confirmed cases and sort the cases. We then print the countries corresponding to the number of cases.

Making a Bubble Chart of the worst-hit countries:-

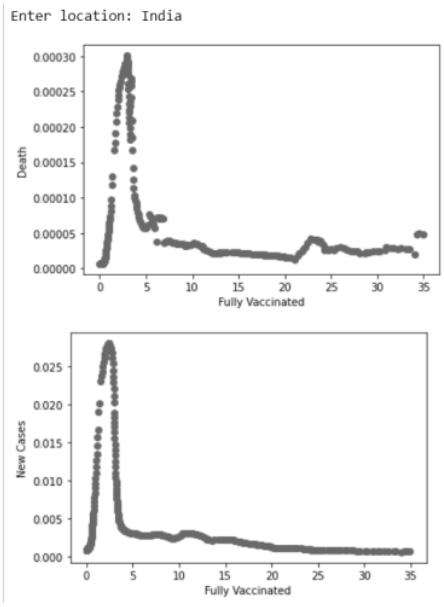
We created an interactive Bubble chart and defined a function for it in Python that traverses through the data and finds the Country with the most number of confirmed cases and creates a radius for the bubble accordingly. A Bubble with a bigger radius would imply that the country has more number of confirmed cases.

Making a line graph to check the details of each Corresponding Country:-

We used Python to traverse through countries and plot a line graph for the same.

Vaccination progress:

We studied the trends in vaccinations and found a negative correlation between the confirmed cases and vaccines administered. This process especially was very enlightening as we could verify for ourselves the positive consequences of administering the vaccine to large amounts of people.



D. CONCLUSION

Our project is based upon the public data available from the COVID-19 Data Repository by John Hopkins CSSE as our main database. We have built upon the available data in the 'time-series' csv files and merged them to form a single excel file from which we have drawn the data visualizations and inferences. We, in effect, have unearthed several trends and have understood the data better with various visualizations.

REFERENCES

[1] WHO
https://www.who.int/emergencies/diseases/novel-coronavirus-2019?adgroupsurvey={adgroupsurvey}&gclid=Cj0KCQiAqbyNBhC2ARIsALDwAsBNXnNd8ED0A64yM0qNtc65Gg9JBfL-xtVp_1hq8BglJfjXLTM0QuAaAgEDEALw_wcB

[2] Wikipedia:-

https://en.wikipedia.org/wiki/COVID-19_misinformation

[3] Dataset on Novel Corona Virus Disease 2019 in India: This dataset has information from the states and union territories of India at a daily level.
https://www.kaggle.com/sudalairajkumar/covid19-in-india

[4] Coronavirus Pandemic (COVID-19):
207 country profiles which allow for exploring the statistics on the coronavirus pandemic for every country in the world.
https://ourworldindata.org/coronavirus

[5] https://www.mygov.in/covid-19/

[6] https://towardsdatascience.com/

[7] https://covid.ourworldindata.org/data/owid-covid-data

Interactive COVID-19 Dashboard

ORIGINALITY REPORT

2%	0%	2%	1%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Renu S. Hiremath, Shreya Bhat, H. R. Srikanth. "An automated evaluator for a classical dance — Bharatanatyam (Nritta)", 2017 Second International Conference on Electrical, Computer and Communication Technologies (ICECCT), 2017 Publication	1%
2	"COVID-19: Systemic Risk and Resilience", Springer Science and Business Media LLC, 2021 Publication	1%
3	Samrat K. Dey, Md. Mahbubur Rahman, Umme R. Siddiqi, Arpita Howlader. "Analyzing the epidemiological outbreak of COVID‐19: A visual exploratory data analysis approach", Journal of Medical Virology, 2020 Publication	<1%

Exclude quotes	On	Exclude matches	< 5 words
Exclude bibliography	On		