How to Cite:
Gorle, D. L., & Padala, A. (2022). The impact of COVID-19 deaths, medical analysis & visualization using Plotly. *International Journal of Health Sciences*, *6*(S3), 11957–11971. https://doi.org/10.53730/ijhs.v6nS3.9220

The impact of COVID-19 deaths, medical analysis & visualization using Plotly

Dhana Lakshmi Gorle
Research Scholar, Department of CSE, GITAM Institute of Technology, Visakhapatnam, AP
Email: dgorle2@gitam.in

Dr. Anuradha Padala
Associate Professor, Department of CSE, GITAM Institute of Technology, Visakhapatnam, AP
Email: apadala@gitam.edu

***Abstract*---**In the Recent Pandemic Situation, detecting the initial outcome of pneumonia was in Wuhan (Hu-bei, China) in December 2019. The Noval Coronavirus as Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) was quickly spreading. The epidemic spread to all over of china's mainland province and other 58 countries and its areas, with higher than 87,137 active cases around the world including 79, 968 cases from China. In the view of this present public health emergency, this done a statistical analysis of COVID-19 Death Medical Analysis and Visualized reported cases of coronavirus diseases 2019 (COVID-19) based on the unfolded data collection provided by Centers for Disease Control and Prevention and Coronavirus Worldometer, where the posi-tion and number of total cases and death cases throughout the Global Countries for Predicted re-sults will help healthcare organizations, Medical industries. Using Plotly library package in Python will help to analyze and provide better visualization graphs for massive datasets.

***Keywords*---**COVID-19, statistical analysis, visualization, coronavirus, worldmeters.

Introduction

COVID-19 and Coronavirus, were initially observed in Wuhan, China, in December 2019, after it is spreader throughout the World, affected around 8 million people throughout the globe and causing nearly 300,000 reported deaths by May 2020. Huge amount of data has been collected from total infected cases with regards of symptoms and history to keep track spread of the disease. With

Manuscript submitted: 18 Feb 2022, Manuscript revised: 27 April 2022, Accepted for publication: 9 June 2022

the rapid growth in the novel coronavirus across the world, the World Health Organization (WHO)[2], the International Committee on Taxonomy of Virus (ICTV)[4], the updated results on the impact of COVID-19 over the past few months. The virus that outbreaks the current pandemic is Coronavirus2, as Serious Acute Respiratory Syndrome Coronavirus2 (SARS-CoV-2). The World Health Organization (WHO), The (JHU) Johns Hopkins University Researchers, The Centers for Disease Control and Prevention [1], and the remaining diseases were maintained. All the data held by Johns Hopkins University is freely accessible via the GitHub repository [12].

Plotly tool visualization is available in Python, supporting the number of interactive, high-quality graphs. It is a more effective tool to represent graphs for Data Science, Machine Learning, and Deep Learning. Google Colaboratory or Colab notebook is a platform for writing and executing Python in a browser with zero configurations required, free access to GPUs, and easily shared. The Colab notebook allows combining executable code and rich text in a single document, along with the images, LaTex, HTML, and many more are available. The Colab notebooks are just like Jupyter notebooks hosted by Google Colab. The Colab notebook has a more excellent facility to store the data in Google Drive account and reduce the space in installed drivers in the CPU. Instead of storing the data in the CPU, the Colab notebook can store in GPU. We can easily share the colab notebooks with co-workers or friends, allowing them to comment on them and even edit notebooks. The colab notebooks also go with the latest trends and techniques of Data Science, Machine Learning, and Deep Learning, and many more resources are also available. Unlike unless did not install the libraries in Jyputer notebook, but with the Colab, we have pre-installed libraries, and we can directly import the libraries.

Related Work

The most available solutions for COVID-19 visualizations can categorize into two main areas: (i) Analysis and visualizations by research centers, individual contributors worldwide; (ii) Analysis and visualization. The first category includes visualizations that are picturized in an easily consumable format. These platforms provide insights into the trends. However, they do not include intelligent analytics, predictions for the future and do not have flexibility in interaction. The second category of solutions is in the form of reports generated by leading research firms, such as IBM, Accenture, and Deloitte, to help businesses and their leaders understand the impact of COVID-19 on the economy and businesses.

Proposed Solution

Developing a dashboard design interactively to include world trends. It is also provides analysis and forecasting of cases that have rarely been seen in the most common visualization solutions available for COVID-19. Thus, we propose a solution with various intuitive and interactive visualizations, up-to-date by dynamic loading of data.

Implementation
Packagesused

- Dash: It is an open-source framework that helps to create interactive and dynamic websites using Python programming languages and using these Dash core components to develop tools to add inputs, sliders, drop-downs, graphs to allow users to interact with the data.
- Plotly: Plotly's Python programming library allows for creating interactive and dynamic graphs. In this Dash is used to build the interface, populated by the graphs created from Plotly.
- Pandas: Pandas is a data analytics library in Python and is commonly used for processing tabular data.

Data set

The latest Datasets are available from the Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE)" in Github Page. The dataset contains information about affected, deaths and recovered cases, country/region-wise and province/state-wise, from January 22, 2020. From these three fields, we have derived the following:

- Active cases
- New confirmed cases
- New deaths
- New recovered cases

Data Processing

Data processing is performed dynamically loaded from websites. The input datasets in the form of excel are each read into the application as a pandas DataFrame are merged. After removing the unwanted fields, new fields are derived from mathematical operations. Global Trends: as seen in Fig. 1, it gives an overview of COVID-19 trends across the world.

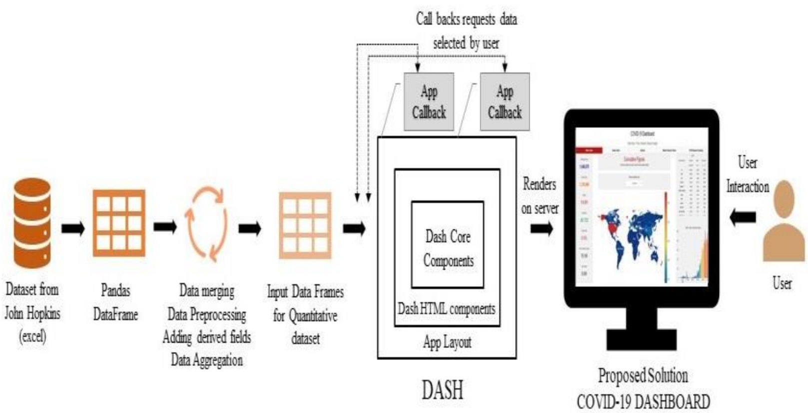


Fig.1. System diagram showing different processes and workflow of the proposed COVID-19 dashboard implementation

- Part 1: It features the heatmap of the world map (Choropleth) with countries highlighted in various color intensities from color scale, based on their figures of confirmed/ death/ recovered cases it can select this type of cases to be visualized from a drop-down.
- Part 2: The second half of the world Trends tab presents a time-series animation of the COVID-19 pandemic spread using an equirectangularchoropleth Map. The figures for the top 10 countries are also showcased with a Doughnut chart and Line-plot to visualize their contribution and trend in cumulative cases.

Daily Reports Data

The Coronavirus Worldmeters will contain the country-based total cases affected, new cases, total deaths, new deaths, recovered cases, active cases. The serious critical, top cases/ 1m pop, deaths 1m pop, entire testes, and population in the CSV file form. This refers to identify the affected information of virus spread. The Coronavirus as COVID-19 is affecting 220 Countries and Territories wise. In this Coronavirus Worldmeters is reset the day after midnight GMT+0. The list of countries and their regional news classification is based on the United Nations geoscheme. The latest news provides the sources.

Table 1
Data Source: COVID-19 Data provided by Worldometers

#	Country, Other	Total Cases	New Cases	Total Deaths	New Deaths	Total Recovered	Active Cases	Serious, Critical	Tot Cases/ 1M pop	Deaths/ 1M pop	Total Tests	Tests/ 1M pop	Population
1	Afghanistan	1,53,375		7,127		1,11,281	34,967	1,124	3,839	178	7,52,190	18,828	3,99,50,726
2	Albania	1,48,222		2,505		1,34,802	10,915	3	51,571	872	10,65,326	3,70,662	28,74,118
3	Algeria	1,96,915		5,339		1,34,073	57,503	42	4,398	119	2,30,861	5,156	4,47,75,274
4	Andorra	15,052		130		14,867	55	4	1,94,445	1,679	1,93,595	25,00,904	77,410
5	Angola	48,004		1,235		43,739	3,030	8	1,410	36	8,30,836	24,396	3,40,56,435
6	Anguilla	256				208	48		16,887		38,936	25,68,338	15,160
7	Antigua and Barbuda	1,742		44		1,349	349	9	17,618	445	17,409	1,76,067	98,877
8	Argentina	51,95,601		1,12,195		48,84,48	1,98,988	2,497	1,13,737	2,456	2,23,07,283	4,88,331	4,56,80,639
9	Armenia	2,43,386		4,876		2,27,726	10,784		81,956	1,642	15,54,292	5,23,379	29,69,725
10	Aruba	14,719		148		13,984	587	14	1,37,181	1,379	1,77,885	16,57,890	1,07,296
11	Australia	58,206	1,641	1,032	13	33,571	23,603	125	2,252	40	3,22,28,090	12,46,945	2,58,45,641

12	Austria	6,91,663		10,778		6,64,582	16,303	142	76,288	1,189	7,72,72,254	85,22,821	90,66,511
13	Azerbaijan	4,32,495		5,722		3,73,475	53,298		42,211	558	44,98,717	4,39,069	1,02,46,025

Time-series Data

Then time-series data contains the data of Total Deaths concerning Date across the countries. The country coordinators the information of time series visualization on geo plots such as Choropleth Maps, bar graphs, scatter plots, and wordcloud to analyze and visualize the global scenarios of COVID-19 and perform the medical analysis of various conditions that accomplish to contribute to death due to COVID-19.

Table 2
Total Death Cases due to COVID-19

Date	Total Deaths	Change in Total	Change in Total (%)
May, 29, 2021	35,48,183	10,887	0%
May, 28, 2021	35,37,296	12,011	0%
May, 27, 2022	35,25,285	11,737	0%
May, 26, 2023	35,13,548	12,404	0%
May, 25, 2024	35,01,144	12,112	0%
May, 24, 2025	34,89,032	8,953	0%
May, 23, 2026	34,80,079	10,073	0%
May, 22, 2027	34,70,006	11,126	0%
May, 21, 2028	34,58,880	12,743	0%
May, 20, 2029	34,46,137	13,124	0%
May, 19, 2030	34,33,013	13,225	0%
May, 18, 2031	34,19,788	13,988	0%

Demerits of existing system

As in an existing system, the problem becomes even worse when the target application involves Twitter messages. Due to short texts (Up to 140 characters), informality, grammatical errors, buzzwords, slang, and the speed with which real-time content, approximately 250 million messages are posted daily. While assigning keywords to the document manually is a

- Very Costly
- Time Consuming
- Tedious Task and
- The number of digitally available documents are growing

Implementation process

To Analysis:

- Importing the Worldmeters dataset and exploring it
- Geographic Data Visualization using Choropleth maps of Total Deaths
- Importing and exploring the medical dataset
- Visualizing the medical data- COVID deaths and Conditions and Age
- WordCloud Text Visualization
- The global impact of COVID-19

Results and Discussions

Importing the Worldmeters dataset and exploring it

Using Google Colab, import all the files from the worldmeters dataset and upload the files.
From google.colab import files
files.upload()
choose CSV file from worldmeters as covid_worldmeters_data.csv
covid_worldometer_data.csv(application/vnd.ms-excel) - 12399 bytes, last modified: 5/28/2021 - 100% done
Saving covid_worldometer_data.csv to covid_worldometer_data.csv
Read the Coovdworldmeters dataset using pandas
df = PD.read_csv('covid_worldometer_data.csv')
df

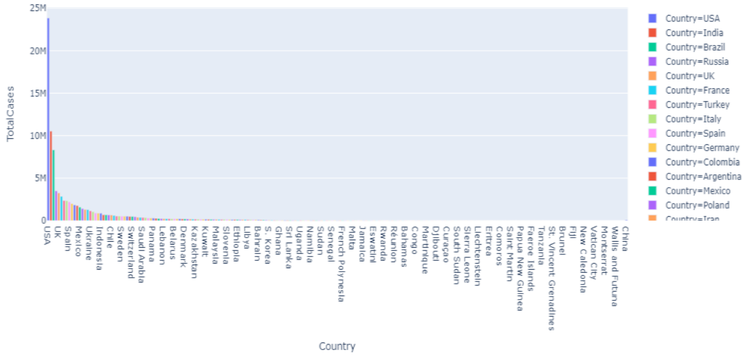
Table 3
Total Cases, Total Deaths, Total Recovered, and Total Active Cases

S.no	Country	Total Cases	Total Deaths	Total Recovered	Active cases	Total cases/ 1M	Total Death/ 1M	iso_alpha
1	USA	23,848,410	397,994	14,112,119	9,338,297	71,822	1,199	USA
2	India	10,528,508	151,954	10,162,738	213,816	7,589	110	IND
3	Brazil	8,326,115	207,160	7,339,703	779,252	39,021	971	BRA
4	Russia	3,495,816	63,940	2,882,044	549,832	23,949	438	RUS
5	UK	3,260,258	86,015	1,406,967	1,767,276	47,890	1,263	UKR
…	…	…	…	…	…	…	…	…

S.no	Country	Total Cases	Total Deaths	Total Recovered	Active cases	Total cases/ 1M	Total Death/ 1M	iso_alpha
217	Wallis and Futuna	4		1	3	359	NaN	WLF
218	Samoa	2		2	0	10	NaN	WSM
219	Micronesia	1		NaN	1	9	NaN	FSM
220	Vanuatu	1		1	0	3	NaN	VUT
221	China	87,988	4,635	82,352	1,001	61	3	CHN

The above table list out the details of Covid-19 such as Country, TotalCases, TotalDeaths, Total Recovered Active Cases, Totalcases/1M, TotalDeaths/1M, and iso_alpha, which show the countries affected

px.bar(df, x='Country', y='TotalCases', color='Country')



Fig, 2. Graph- Total Cases across Countries

The above graph depicts that the USA has a higher rate of infected cases, greater than 10M than other countries. Next to the UK, Spain shows an increased number of infected cases, for every nation has specified a specific color to differentiate the countries.

px.bar(df, x='Country', y='TotalDeaths', color='Country')

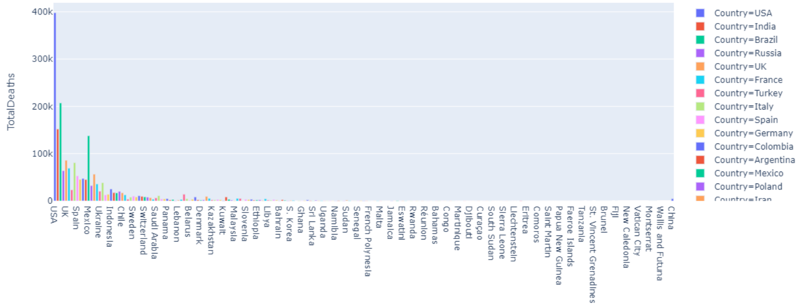


Fig. 3. Graph- Total Deaths across Countries

The above graph shows that the UK has highly affected death cases when compared to all other countries. It is nearly 200k of death cases; next to the USA, Mexico is shown death cases highly, to differentiate with specific colors to different countries.

Index the columns using:

df.columns

Index(['S.no', 'Country', 'TotalCases', 'TotalDeaths', 'Total Recovered', 'Active cases', 'Totalcases/1M', 'TotalDeath/1M', 'iso_alpha'], dtype='object')

Scatter the plots from Plotly with the coordinates of Total cases affected and Total Death Cases

px.scatter(df, x='Totalcases/1M', y='TotalDeath/1M', color='Country')

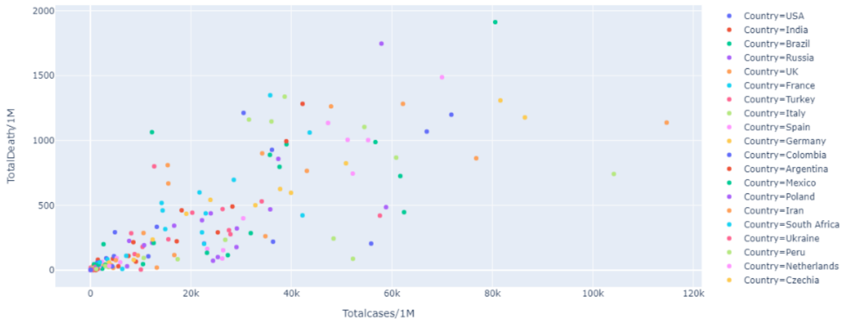


Fig. 4. Graph- Total affected cases and Total Deaths Cases over 1M

The above graph depicts that the Total affected and Total Death cases over 1M.

Geographic Visualization using Choropleth maps

df.columns

Index(['S.no', 'Country', 'TotalCases', 'TotalDeaths', 'Total Recovered',
'Active cases', 'Totalcases/1M', 'TotalDeath/1M', 'iso_alpha'],
dtype='object')

A choropleth map displays divided geographical areas or regions that are coloured, shaded or patterned in relation to a data variable. Amazing Representation of data in a map.Choropleth maps provide to visualize how a measurement varies across a geographic area.

```
#Dataset 1
#parameters= dataset, locations= ISOALPHA, color, hover_name, color_continuou
s_scale= [RdYlGn, Blues, Viridis...],
px.choropleth(df, locations='iso_alpha',
hover_name='Country',
color='Totalcases/1M',
projection='orthographic',
color_continuous_scale='viridis')
```

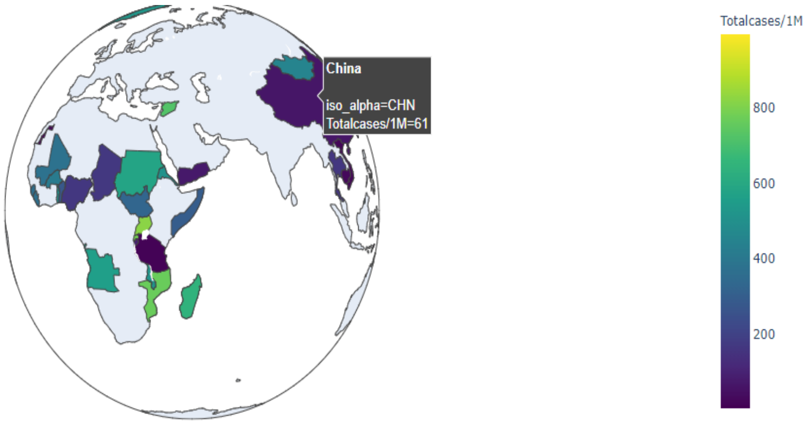


Fig. 5. The Choropleth Maps Chain over 61 Total Cases

The Geographic Visualization using Choropleth maps shows china has 61 Total Cases over 1M. Choropleth maps uses the projection has Orthographic, and the color continuous scale has viridis.

```
px.choropleth(df, locations='iso_alpha',
hover_name='Country',
color='Totalcases/1M',
projection='equirectangular',
color_continuous_scale='plasma')
```

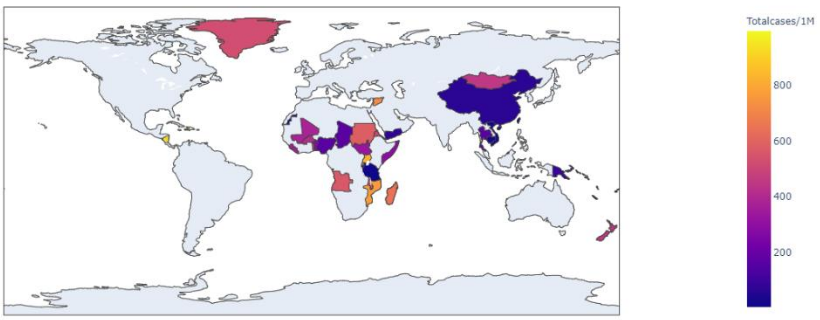


Fig. 6. The choropleth Maps Total Cases affected throughout the World

It is very crucial to show how the virus has spread to the countries in a brief period. The geographical and choropleth map with different colors as continuous color scale has plasma and projection has equirectangular.

```
px.choropleth(df, locations='iso_alpha',
hover_name='Country',
color='Totalcases/1M',
scope='asia',
color_continuous_scale='plasma')
```

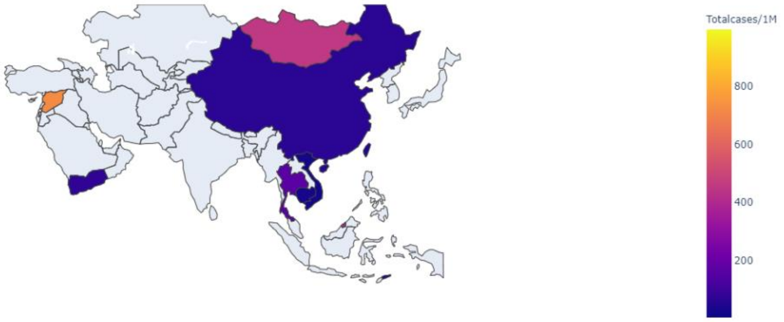


Fig. 7. The choropleth Maps Total Cases affected through worldmeter data

Analyzing total cases would require pre-processing the worldmeters data and visualization in Plotly, which is in the Colab notebook. The choropleth graphs show the same as the above while using the consolidated view specified as the scope of Asia with the continuous color scale as plasma.

Importing and Analyzing the dataset(Dataset 2: Covid-19 Pre-existing condition Medical dataset; source: CDC, US)

```
from google.colab import files
files.upload()
```

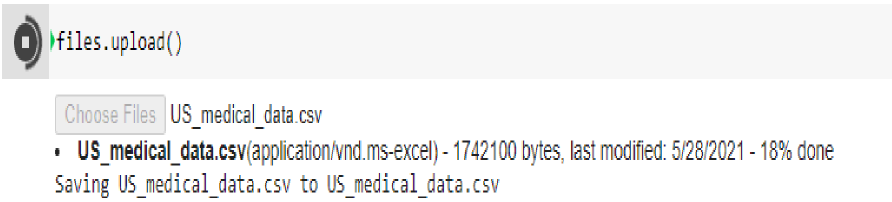


Fig. 8. Uploading US_medical_data dataset

```
df1 = pd.read_csv('US_medical_data.csv')
df1
```

Table 4
Total Cases Total Deaths, Total Recovered Cases and Total Active Cases over 1M to the US_medical_data.csv dataset

S.No	Data as of	Start Week	End Week	State	Condition Group	Condition	ICD10_codes	Age Group	COVID-19 Deaths	Number of Mentions	Flag
0	1/10/2021	1/4/2020	1/9/2021	US	Respiratory diseases	Influenza and pneumonia	J09-J18	0-24	198.0	205.0	NaN
1	1/10/2021	1/4/2020	1/9/2021	US	Respiratory diseases	Influenza and pneumonia	J09-J18	25-34	968.0	985.0	NaN
2	1/10/2021	1/4/2020	1/9/2021	US	Respiratory diseases	Influenza and pneumonia	J09-J18	35-44	2537.0	2584.0	NaN
3	1/10/2021	1/4/2020	1/9/2021	US	Respiratory diseases	Influenza and pneumonia	J09-J18	45-54	7369.0	7555.0	NaN
4	1/10/2021	1/4/2020	1/9/2021	US	Respiratory diseases	Influenza and pneumonia	J09-J18	55-64	18378.0	18829.0	NaN
...	...	...	...	...	...	...	...	...	...	...	...
12415	1/10/2021	1/4/2020	1/9/2021	PR	COVID-19	COVID-19	U071	65-74	321.0	321.0	NaN
12416	1/10/2021	1/4/2020	1/9/2021	PR	COVID-19	COVID-19	U071	75-84	366.0	366.0	NaN
12417	1/10/2021	1/4/2020	1/9/2021	PR	COVID-19	COVID-19	U071	85+	264.0	264.0	NaN
12418	1/10/2021	1/4/2020	1/9/2021	PR	COVID-19	COVID-19	U071	Not stated	0.0	0.0	NaN
12419	1/10/2021	1/4/2020	1/9/2021	PR	COVID-19	COVID-19	U071	All Ages	1290.0	1290.0	NaN

The above table list out the details of Covid-19 such as Country, TotalCases, TotalDeaths, Total Recovered Active Cases, Totalcases/1M, TotalDeaths/1M, and iso_alpha, which show the countries affected. Using groupby clause grouping out all the states based on the count condition
df1.groupby(['State']).count()

It will dive into medical analysis, import the medical data set into the Google Collab environment, and explore it. So the import the data, we will use the following syntax from google dot collab. We are going to import finds and start the upload. We can find our medical data dot CSB. We will load this data set, so please double click on it, and the Data is being uploaded. Data has into the

Google Collab environment. Now let us use read unless ucsb constant from pandas to take a look at this Data. So it is BF one the name of the data frame, and pity for pandas. Moreover, we need to input the file name. We will cooperate from here to avoid the type of mystic.

Visualizing the Pre-existing condition Medical dataset

```
filter1=df1[(df1['State']=='US')]
px.bar(df1,
x='Age Group',
y='COVID-19 Deaths',
color='Condition',
hover_name='Condition',
log_y=True)
```

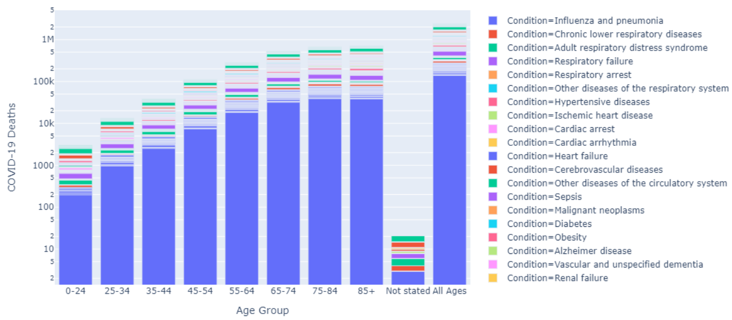


Fig. 9. Graph – Visualize the pre-existing US_medical _data.csv dataset over the Age Group

```
px.bar(df1,
x='Age Group',
y='COVID-19 Deaths',
color='Condition Group',
hover_name='Condition')
```

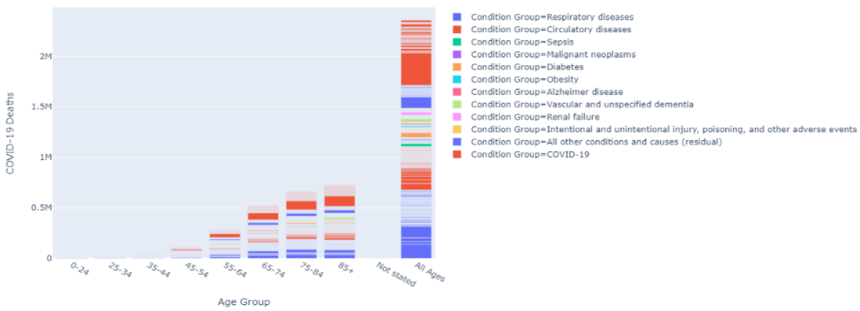


Fig. 10. Fig. 9: Graph – Visualize the pre-existing US_medical _data.csv dataset over the Age Group throughout the World

```
filter2 = df1[(df1['State']=='US') & (df1['Age Group']=='All Ages')]
```

```
filter2.columns
Index(['Data as of', 'Start Week', 'End Week', 'State', 'Condition Group',
'Condition', 'ICD10_codes', 'Age Group', 'COVID-19 Deaths',
'Number of Mentions', 'Flag'],
dtype='object')
px.bar(filter2, x='Condition Group', y='COVID-19 Deaths', color='COVID-
19 Deaths', hover_name='Condition')
```

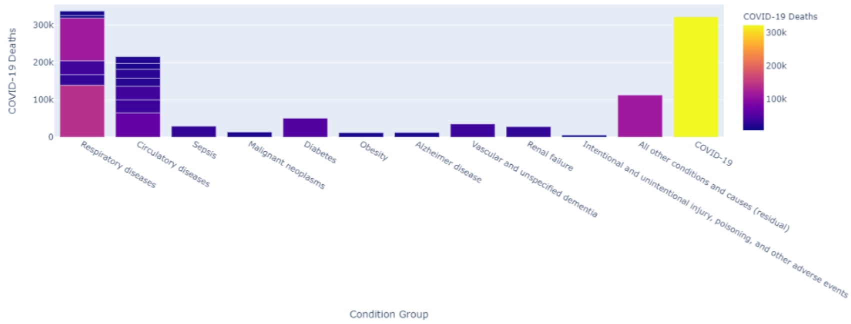


Fig. 11. Graph – Visualize the pre-existing US_medical _data.csv dataset over the Age Group Deaths

WordCloud Text Visualization

Step a: Convert the columns with conditions and condition group into the list using the tolist() function.
Step b: Convert the list to one single string
Step c: Convert the string into WordCloud

```
WordCloud imported in task 1. # WordCloud visualization needs matplotlib
list1=filter2['Condition Group'].tolist()
string1="".join(list1)
plt.figure(figsize=(20,20))
plt.imshow(WordCloud().generate(string2))
```

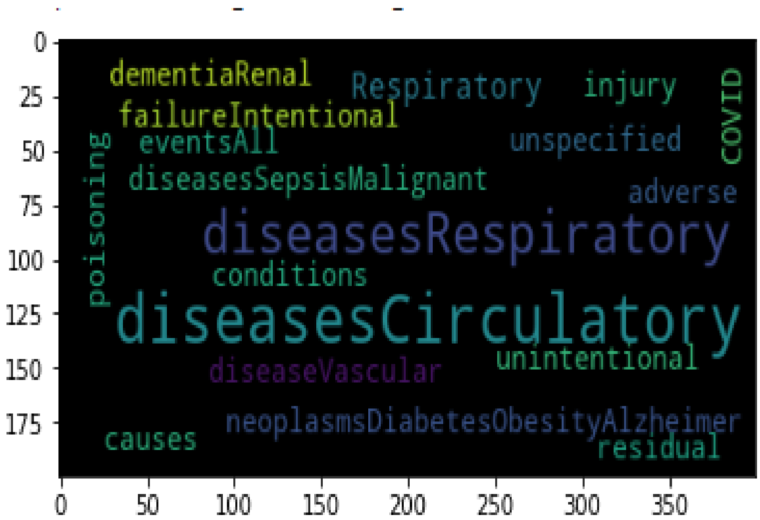


Fig. 12. Graph – Text Visualization of pre-existing US_medical _data.csv dataset

Conclusion

The experiment has been done with data addition, and interactive dashboards are allowed to visualize the information and it is easier for the human understanding. It is Designed an interactive data-driven visualization, analytics, and forecast the data through dashboard. It reduces the dynamic calculations and predictions for each page load about 89%, which can be achieved by adding another worker to pre-calculate the visuals' required data at regular intervals. In this way,the main dashboard can run faster about90% without worrying about recalculating the pictures for each user. We would also like to include additional visualizations like country-specific lock down information and its effect on virus containment and more in-depth analytics involving characteristics of patients like age, gender, and country-specific testing statisticsand also list out the countries classification is based on United Nations Geoscheme. The latest sources are provided by the dashboard worldmeter's COVID-19 data.

References

1. "COVID-19 dashboard," Johns Hopkins University, https://coronavirus. jhu.edu/map.html, Accessed: 2020-04-18.
2. "COVID-19 dashboard," World Health Organization, https://covid19.
3. "COVID-19 dashboard" Worldmeters Coronavirus, https://www.worldometers.info/coronavirus
4. World Health Organization, dipicts the coronavirus disease (COVID-19) and the virus that causes it. The ICTV's page is here: International Committee on Taxonomy of Viruses (ICTV).
5. COVID-19 dashboard " The Center for Disease Control and Prevention, https://www.cdc.gov/coronavirus/2019-ncov/index.html
6. Sahitya, "The impact of COVID-19-Data Visulaization using Plotly and Comparative Aanalysis with SAR" Mar 6 2020.

7. Frincy Clement, AsketKaur, Maryam Sedgh, DeepaKrishnaswamy, KumardevanPunithakumar, "Interactive Data-Driven Visualization for COVID-19 with Trends, Analytics and Forecasting" 24th International Conference Information Visualisation(IV), 978-7281-9134@2020 IEEE
8. R. Muthusamy, K. Saritha, "Statistical Analysis and Visualization of the Potential Cases of Pandemic Coronavirus" Springer Publications, 204-208, 10 June 2020, Indian virological Society 2020
9. World Health Organization. Dipicts the coronavirus disease (COVID-19). The ICTV's page is here: International Committee on Taxonomy of Viruses (ICTV). 2020. https://talk.ictvonline.org/. Accessed 20 March 2020.
10. World Health Organization. Rolling updates on coronavirus disease (COVID-19). 2020. https://www.who.int/emergencies/dis eases/novel-coronavirus-2019/events-as-they-happen. Accessed 20 Mar 2020.
11. 2019 Novel CoronaVirus CoViD-19 (2019-nCoV) Data Repository by Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE). 2020. https://github.com/CSSEGISandData/COVID-19. We accessed 17 April 2020.
12. Suryasa, I. W., Rodríguez-Gámez, M., & Koldoris, T. (2021). Health and treatment of diabetes mellitus. International Journal of Health Sciences, 5(1), i-v. https://doi.org/10.53730/ijhs.v5n1.2864
13. Suryasa, I. W., Rodríguez-Gámez, M., & Koldoris, T. (2021). The COVID-19 pandemic. International Journal of Health Sciences, 5(2), vi-ix. https://doi.org/10.53730/ijhs.v5n2.2937
14. Mahardika, I. M. R., Suyasa, I. G. P. D., Kamaryati, N. P., & Wulandari, S. K. (2021). Health literacy is strongest determinant on self-monitoring blood glucose (SMBG) type 2 DM patients during COVID-19 pandemic at public health centre in Tabanan Regency. International Journal of Health & Medical Sciences, 4(3), 288-297. https://doi.org/10.31295/ijhms.v4n3.1752