DATA VISUALIZATION PROJECT REPORT

(Project Semester January-May 2023)

COVID19-ANALYTICS DASHBOARD

Submitted by

Satish Kumar

Registration No. - 12020945

Program: B.Tech.(Computer Science and Engineering)

Section: K20DM

Course Code – INT233

Under the Guidance of

Sandeep Kaur; UID:23614

Discipline of CSE/IT

Lovely School of Computer Science and Engineering

Lovely Professional University, Phagwara

2022-23

CERTIFICATE

This is to certify that Satish Kumar bearing Registration no. 12020945 has completed INT233 project titled, **"COVID19-ANALYTICS DASHBOARD"** under my guidance and supervision. To the best of my knowledge, the present work is the result of his original development, effort, and study.

Sandeep Kaur

School of Computer Science and Engineering
Lovely Professional University
Phagwara, Punjab.

2022-23

DECLARATION

I, Satish Kumar, student of B.Tech.(Computer Science and Engineering) under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 20 April'2023

Registration No. 12020945 Satish Kumar

ACKNOWLEDGEMENT

I would like to take this opportunity to express our sincere gratitude and appreciation to all those who have contributed to the successful completion of our project on "**COVID19 - Analytics Dashboard**".

Firstly, I would like to thank our college for providing us with the necessary infrastructure and resources to carry out this project. I am grateful to our project guide, **Sandeep Kaur**, for providing us with valuable guidance and constant support throughout the project.

I extend my heartfelt thanks to all the faculty members of our department who have helped us in various ways during our project. I would also like to thank our friends and colleagues for their valuable feedback and suggestions.

I would like to express our gratitude to "**Kaggle.com**" for providing the data set required in this project which helped me to achieve my goals.

Finally, I express my heartfelt gratitude to the frontline healthcare workers who have been tirelessly working to fight against the COVID19 pandemic. Their selfless service has been a great source of inspiration for us, and I dedicate this project to them.

Thank you all for your invaluable support and encouragement in completing our project on "**COVID19 - Analytics Dashboard**".

Table of Contents

Introduction

The Covid-19 pandemic had a significant impact on the world, affecting every aspect of life. As the pandemic continued to evolve, there was growing need for data-driven insights to help policymakers, researchers, and the public make informed decisions. One way to provide these insights was through an interactive dashboard that can visualize and analyse Covid-19 related data.

The purpose of this report is to explore the creation of an interactive dashboard using data analytics software such as Tableau, that can provide insights and analytics related to the Covid-19 pandemic. The dashboard will focus on visualizing data related to Covid-19 cases, deaths, active, confirmed and recovered rates across the world.

The report will cover various aspects of the creation of the dashboard, including data collection, data exploration, dashboard design, and making stories in Tableau Desktop. By exploring these aspects, this report aims to provide a comprehensive understanding of the process involved in creating an interactive dashboard that can provide data-driven insights related to Covid-19.

Overall, the report aims to highlight the importance of data analytics in providing insights and analytics related to the Covid-19 pandemic. The interactive dashboard will be a valuable resource for policymakers, researchers, and the public to understand the impact of the pandemic on different aspects of life.

Scope of the Analysis

The scope of this analysis is to create an interactive dashboard using Tableau that can provide insights and analytics related to the Covid-19 pandemic. The dashboard will focus on providing users with the ability to visualize data related to Covid-19 cases, deaths, and vaccination rates across the world.

The analysis will cover the following aspects:

1. Data Collection: Collect data from online source Kaggle.com. The data will be cleaned and pre-processed to make it suitable for analysis.

2. Data Exploration: Explore the collected data to identify patterns, trends, and insights related to the Covid-19 pandemic. The exploration will be done using various statistical and visualization techniques.

3. Dashboard Design: Design an interactive dashboard using Tableau that can provide users with the ability to explore and visualize Covid-19 related data. The dashboard will include charts, maps, and other visualizations that will help users understand the impact of the pandemic on countries and WHO Regions.

4. Dashboard Testing: Test the dashboard to ensure that it provides accurate and reliable data. The testing will include user testing, data validation, and dashboard performance testing.

Overall, this analysis will provide users with a powerful tool to explore and understand the impact of Covid-19 on different aspects of life. It will be a valuable resource for policymakers, researchers, and the public who are interested in understanding the pandemic and its effects.

Existing System

The existing systems for visualizing and analysing Covid-19 related data are manual analysis, excel or google sheets.

i. Drawbacks or limitations of existing system

Firstly, Excel, google sheets or manual analysis relies heavily on manual data processing, which can lead to delays in data availability and inconsistencies in data quality. The manual process can be time-consuming and prone to errors, which can impact the accuracy of the analysis.

Secondly, Excel or google sheets or manual analysis is limited in its ability to integrate data from multiple sources. This can make it difficult to gain a comprehensive view of the pandemic's impact and can limit the usefulness of the analysis.

Another major drawback of Excel or google sheets or manual analysis is the lack of interactivity and flexibility. These systems lack the dynamic and interactive visualizations that allow users to explore data in real-time and gain insights quickly. This limits the ability of users to make data-driven decisions.

Additionally, Excel or google sheets or manual analysis does not provide enough context for the data being presented. Raw case counts and death tolls may not provide a complete picture of the pandemic's impact without additional information about the population affected, testing rates, and other relevant factors.

Moreover, Excel or google sheets or manual analysis lacks the ability to automate the analysis process, which can be time-consuming and repetitive. This makes it difficult to generate insights quickly and efficiently.

Overall, compared to an interactive dashboard created using data analytics software such as Excel or google sheets or manual analysis has several drawbacks, including limited interactivity, lack of integration, insufficient context, and lack of automation. These limitations highlight the need for a more comprehensive and flexible solution such as an interactive dashboard that can provide real-time insights and analytics related to the pandemic.

Source of dataset

Kaggle is a popular platform for hosting and sharing datasets related to various domains, including Covid-19. The dataset use for this project has been taken from kaggle.com for creating Covid19-Analytics Dashboard.

This data includes 49,068 rows entries and columns “Province/State, Country/Region, Lat, Long, Date, Confirmed, Deaths, Recovered, Active, WHO Region” related to Covid-19. This data shows records of the period 22 Jan’2020 to 27 July’2020. It can be used for conducting research and analysis on various aspects of the pandemic.

Sample view of dataset:

Province/	Country/R	Lat	Long	Date	Confirmed	Deaths	Recovered	Active	WHO Region
	Afghanist	33.93911	67.70995	22-01-2020	0	0	0	0	Eastern Mediterranean
	Albania	41.1533	20.1683	22-01-2020	0	0	0	0	Europe
	Algeria	28.0339	1.6596	22-01-2020	0	0	0	0	Africa
	Andorra	42.5063	1.5218	22-01-2020	0	0	0	0	Europe
	Angola	-11.2027	17.8739	22-01-2020	0	0	0	0	Africa
	Antigua ar	17.0608	-61.7964	22-01-2020	0	0	0	0	Americas
	Argentina	-38.4161	-63.6167	22-01-2020	0	0	0	0	Americas
	Armenia	40.0691	45.0382	22-01-2020	0	0	0	0	Europe
Australian	Australia	-35.4735	149.0124	22-01-2020	0	0	0	0	Western Pacific
New Sout	Australia	-33.8688	151.2093	22-01-2020	0	0	0	0	Western Pacific
Northern	Australia	-12.4634	130.8456	22-01-2020	0	0	0	0	Western Pacific
Queensla	Australia	-27.4698	153.0251	22-01-2020	0	0	0	0	Western Pacific
South Aus	Australia	-34.9285	138.6007	22-01-2020	0	0	0	0	Western Pacific
Tasmania	Australia	-42.8821	147.3272	22-01-2020	0	0	0	0	Western Pacific
Victoria	Australia	-37.8136	144.9631	22-01-2020	0	0	0	0	Western Pacific
Western A	Australia	-31.9505	115.8605	22-01-2020	0	0	0	0	Western Pacific

Analysis on dataset

I. Analysing sum of day wise covid cases like confirmed, active, death and recovered on a particular date by making calendar chart.

i. Introduction: As the covid cases are increasing day by day, it is important to get daily report that how many people with covid cases are being recovered, active, confirmed, or death. Analysing and visualising may help to get daily case count of covid cases. I will describe a method for analyzing the sum of day-wise COVID-19 cases by creating a calendar chart.

ii. General Description: The objective of this analysis is to create a calendar chart that shows the sum of COVID-19 cases on a particular date, including confirmed cases, active cases, deaths, and recoveries.

iii. Specific Requirements, functions, and formulas: To create the calendar chart, it will require a spreadsheet software such as Tableau desktop. The data will need to be structured in a specific way, with each row representing week(Date) and each column weekday(Date). Months and WHO Region are filtered to get month and who region wise cases.

iv. Analysis results: Once the data has been structured and the filters have been applied, the calendar chart is generated. The chart shows each date as a separate square, with the color of the square indicating the total number of cases on that date. Labels indicate the specific number of cases for each date.

By analysing the calendar chart, cases are increasing daily. It can also compare the number of cases on different dates to identify if the number of cases is increasing or decreasing over time.

v. Visualization:

The calendar chart is an effective visualization tool for showing day-wise COVID-19 cases. It provides a clear overview of the data and allows for easy identification of trends and patterns. The chart can be customized by selecting different WHO Region and different months.

II. Analysing sum of death vs. sum of recovered cases in selected WHO Region by making scatter plot and animating it over the period.

i. Introduction: The aim of this analysis is to analyse the sum of death versus the sum of recovered cases in a selected WHO region over the period. The analysis will be done by creating a scatter plot and animating it to show the changes in the data over time. It will help in understanding the relationship between the number of deaths and recoveries in a particular WHO region and to visualize this relationship in an interactive and informative way.

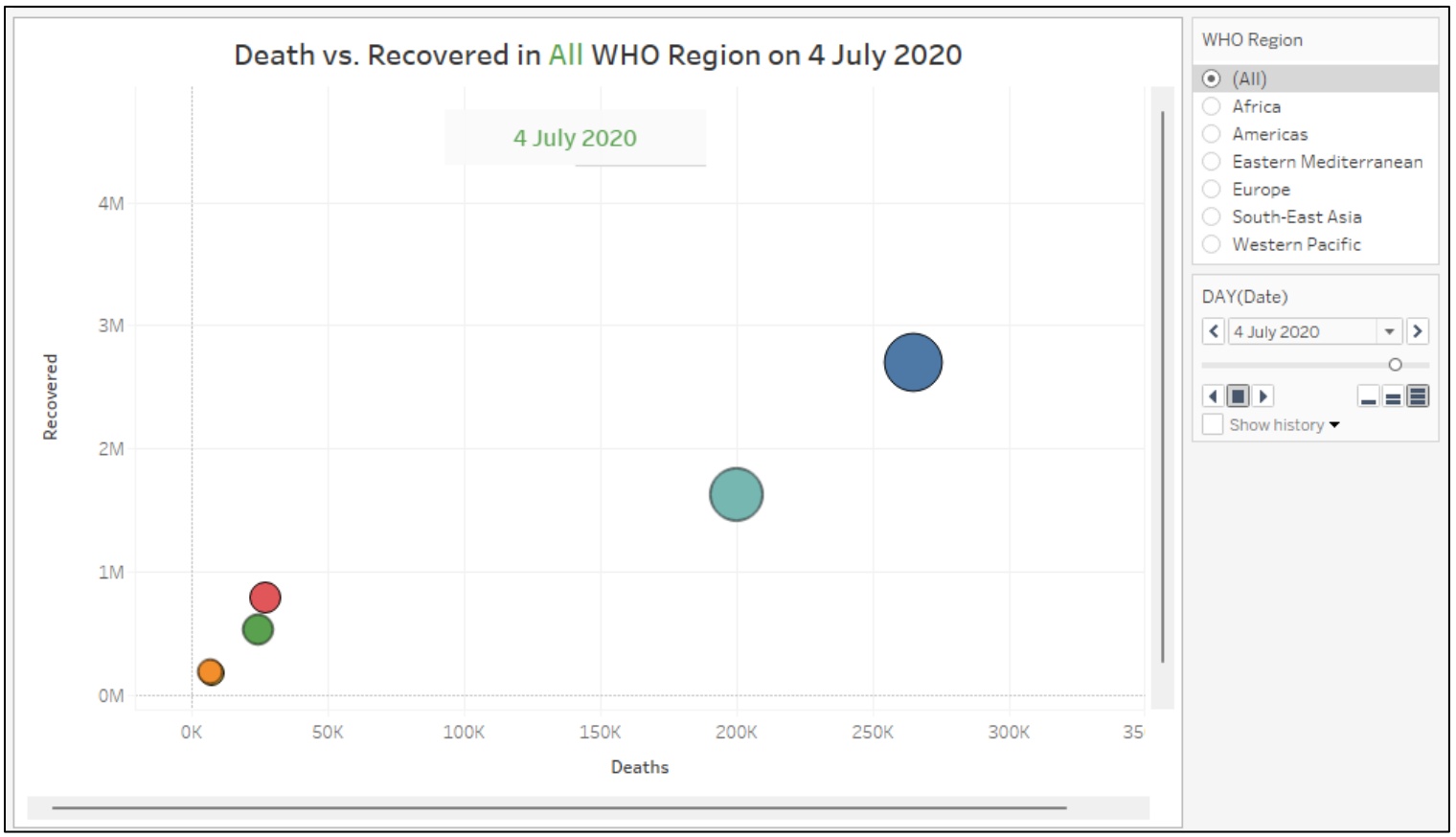
ii. General Description: To carry out this analysis, the number of deaths and recovered cases in the selected WHO region over time will be analyzed and create visualizations.

iii. Specific Requirements, functions, and formulas: To create the scatter plot, firstly data blending is performed on the same data and linked, and row represents recovered cases from primary data source and column contains recovered case from secondary data source WHO Region are filtered, WHO Regions are kept in color marks. Now, the scatter plot is ready. To

animate it over the period, DAY(Date) is kept in Pages area. Now scatter plot is ready to animate over the period WHO Region wise.

iv. Analysis results: The analysis reveals the relationship between the number of deaths and recovered cases in the selected WHO region over time. We will be able to identify any patterns or trends in the data and draw conclusions about the effectiveness of the region's response to the pandemic.

v. Visualization:



The visualization consists of a scatter plot of the sum of deaths versus the sum of recovered cases, with time as the animation parameter. The scatter plot will be animated to show how the relationship between the two variables changes over time. The animation will provide an interactive and informative way of visualizing the data and understanding the trends and patterns over time.

III. Analysing sum of confirmed vs. sum of active cases in selected countries by making scatter plot and animating it over the period.

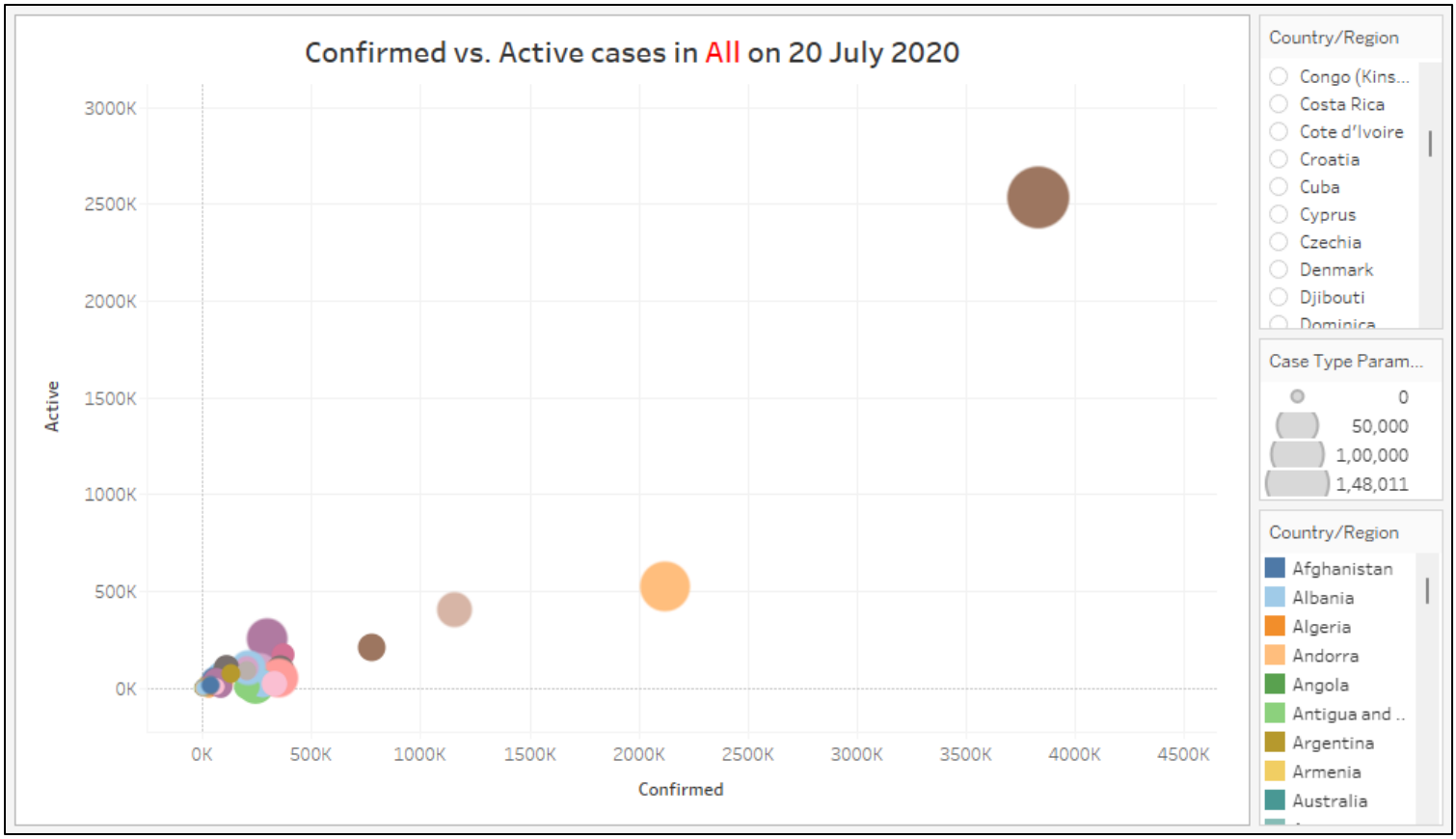
i. Introduction: The aim of this analysis is to analyse the sum of confirmed vs. sum of active cases in a selected countries over the period. The analysis will be done by creating a scatter plot and animating it to show the changes in the data over time. It will help in understanding the relationship between the number of confirmed and active cases in selected countries and to visualize this relationship in an interactive and informative way.

ii. General Description: To carry out this analysis, the number of confirmed and active cases in the selected countries over time will be analyzed and create visualizations.

iii. Specific Requirements, functions, and formulas: To create the scatter plot, row represents active cases and column contains confirmed cases. Countries are filtered and kept in color marks. Now, the scatter plot is ready. To animate it over the period, DAY(Date) is kept in Pages area. Now scatter plot is ready to animate over the period countries wise.

iv. Analysis results: The analysis reveals the relationship between the number of active and confirmed cases in the selected country over time. We will be able to identify any patterns or trends in the data and draw conclusions about the effectiveness of the region's response to the pandemic.

v. Visualization: The visualization consists of a scatter plot of the sum of active versus the sum of confirmed cases, with time as the animation parameter. The scatter plot will be animated to show how the relationship between the two variables changes over time. The animation will provide an interactive and informative way of visualizing the data and understanding the trends and patterns over time.



IV. Finding countries in selected WHO Region in which death due to covid is maximum till selected date and representing on world map.

i. Introduction: The World Health Organization (WHO) has divided the world into six regions for monitoring and coordination purposes. In this analysis, focus will be on finding the countries in a selected WHO region and in selected countries where death due to COVID-19 is maximum until a selected date and representing them on a world map.

ii. General Description: To achieve the objective of this analysis, it needs to perform the following tasks:

Identify the selected WHO region and countries for analysis.

Filter the data to include only deaths up to the selected date.

Determine the countries within the selected region with the maximum number of COVID-19 deaths.

Create a world map visualization and highlighting the countries with the maximum number of COVID-19 deaths.

iii. Specific Requirements, functions, and formulas: To accomplish the tasks outlined above, the following things requires:

- Knowledge to create a world map visualization.
- Familiarity with filtering data in Tableau to obtain data for specific time periods.

iv. Analysis results: Countries within the selected WHO region with the maximum number of COVID-19 deaths up to the selected date are represented with darker color. These results can assess the impact of the pandemic on different regions and countries and inform public health policy and interventions and government to increase supplies in proper amount.

v. Visualization: To present our analysis results visually, created a world map visualization. The visualization includes color markers, darker color shows indicating the countries with the maximum number of COVID-19 deaths within the selected region.

V. Analysing world map by density graph which show covid confirmed cases till selected date in selected WHO Region.

i. Introduction: The World Health Organization (WHO) has divided the world into six regions for monitoring and coordination purposes. In this analysis, focus will be on analyzing the confirmed COVID-19 cases in a selected WHO region until a selected date and representing them on a world map using a density graph.

ii. General Description: To achieve the objective of this project, we will need to perform the following tasks:

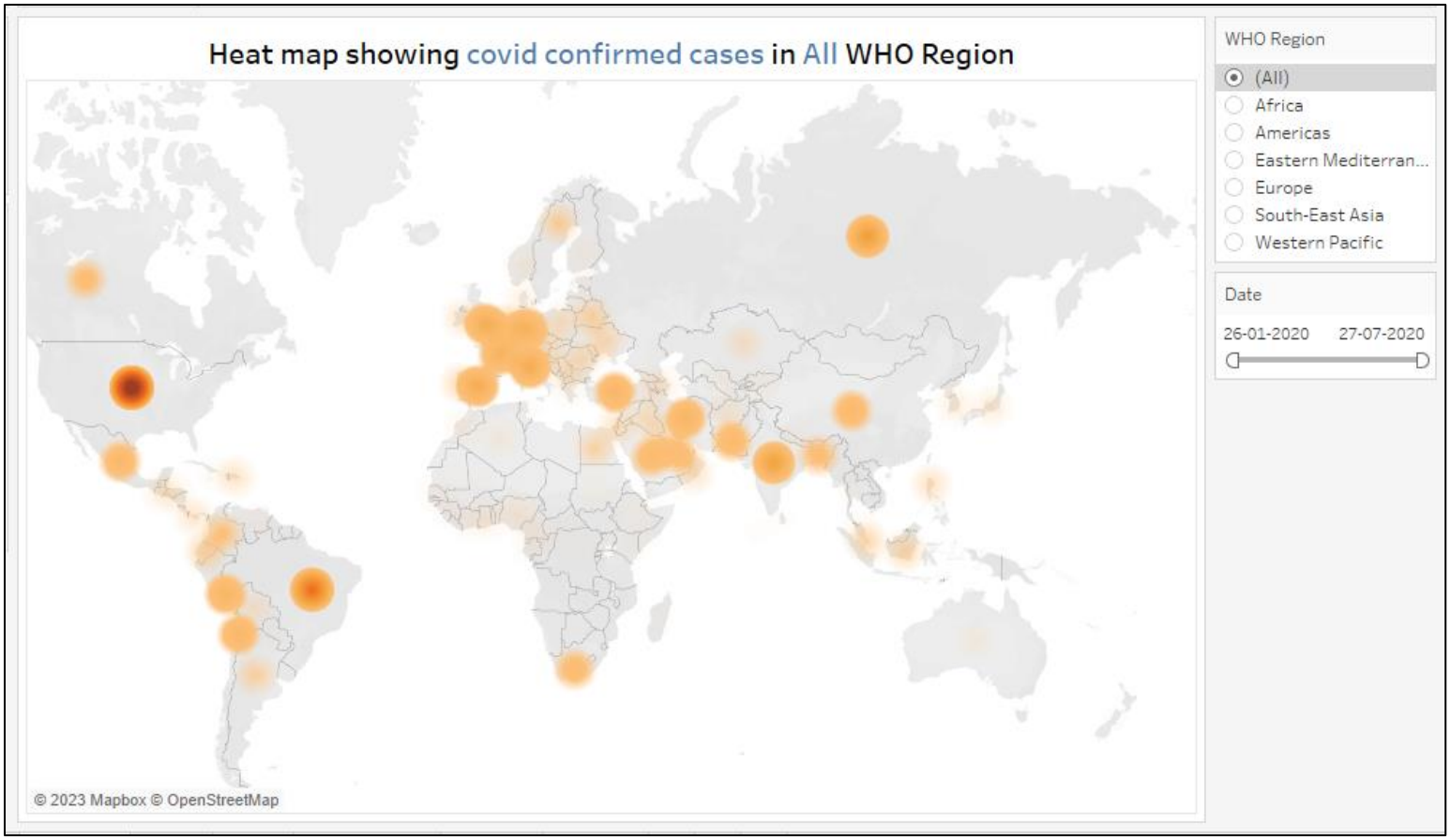
- Identify the selected WHO region for analysis.
- Filter the data to include only cases up to the selected date.
- Use a density graph to represent the distribution of confirmed cases across the world map.
- Analyse the density graph to identify high-risk areas within the selected region.

iii. Specific Requirements, functions, and formulas: To accomplish the tasks outlined above, we will require the following:

- Knowledge to create a density graph on world map.
- Familiarity with filtering data to obtain data for specific time periods.

iv. Analysis results: The density graph shows the distribution of confirmed COVID-19 cases across the world map within the selected WHO region up to the selected date. This analysis can be used within the region and inform public health policy and government to take actions like supplying more vaccines, applying lockdown and improve medical health centres.

v. Visualization: To present our analysis results visually, created a density graph on world map, showing the distribution of confirmed COVID-19 cases across the world map within the selected WHO region. The density graph will use a colour scale to represent the density of confirmed cases, with higher densities indicated by darker colours.



VI. Trend analysis of each type of covid cases over selected period in selected WHO Regions and countries.

i. Introduction: In this analysis, focus will be on trend analysis of each type of COVID-19 cases (confirmed, recovered, active and deaths) over a selected period in a selected WHO region and countries in the selected period.

ii. General Description: To achieve the objective of this project, we will need to perform the following tasks:

- Identify the selected WHO region and countries for analysis.
- Use trend analysis to represent the different types of cases over the selected period for each country within the selected region.
- Analyse the trend analysis to identify changes and patterns in the different types of cases over the selected period.

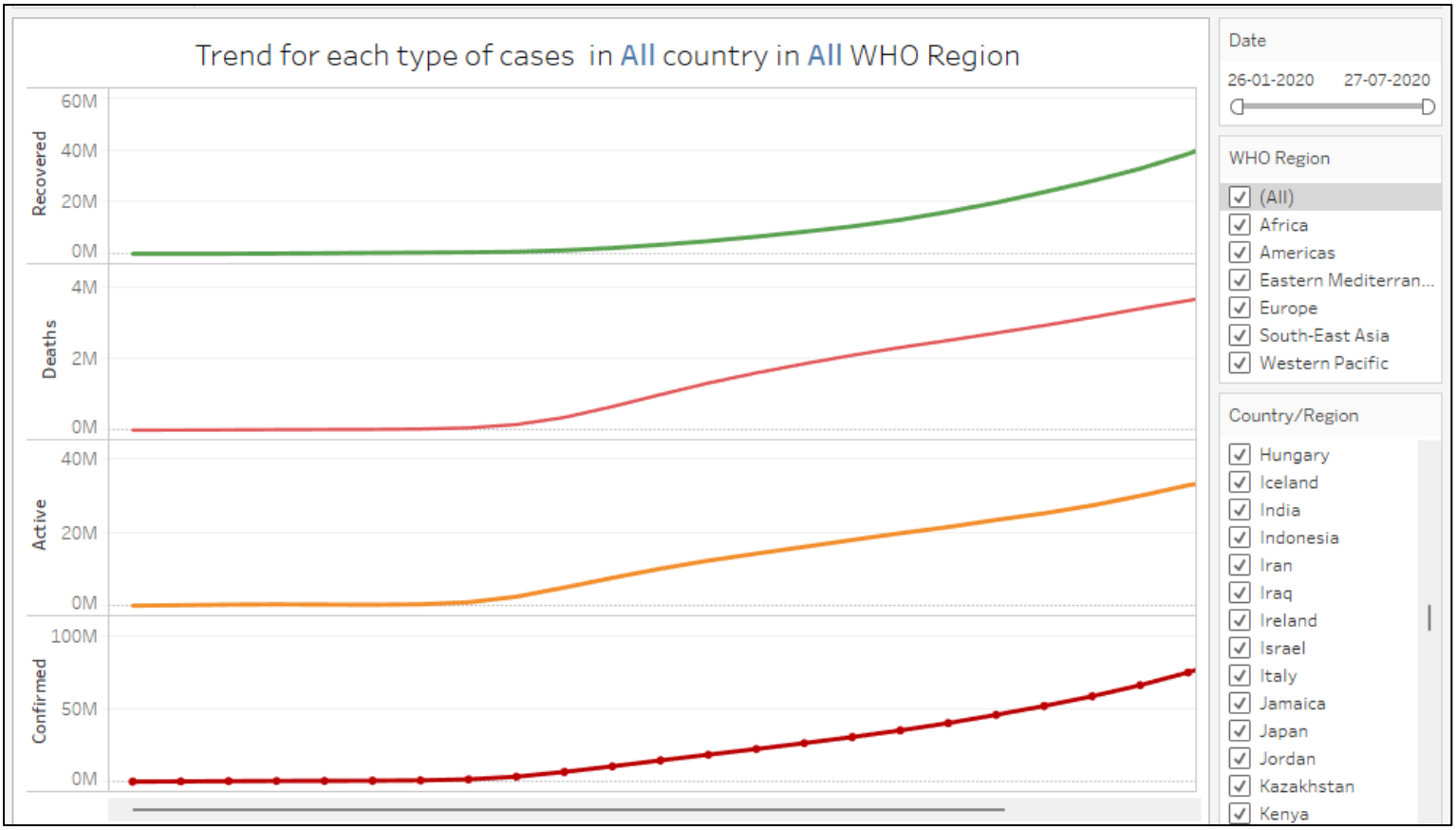
iii. Specific Requirements, functions, and formulas:

- Access to reliable and up-to-date data on COVID-19 cases by country and region.

- Knowledge of perform trend analysis.
- Familiarity with filtering data in Tableau to obtain data for specific time periods.

iv. Analysis results: This trend analysis shows different types of COVID-19 cases (confirmed, recovered, active and deaths) across countries within the selected region over the selected period.

v. Visualization:



To present our analysis results visually, created a trend analysis for each type of COVID-19 case (confirmed, recovered, active, and deaths) over the selected period for each country within the selected WHO region. The trend analysis will use a line chart to represent the changes and patterns in each type of case over time.

VII. Analysing each case type side by side using bar chart over selected period in selected WHO Region and countries.

i. Introduction: In this analysis, focus will on analyzing the different types of COVID-19 cases (confirmed, recovered, active and deaths) side by side using a bar chart over a selected period in a selected WHO region and countries.

ii. General Description: To achieve the objective, it will need to perform the following tasks:

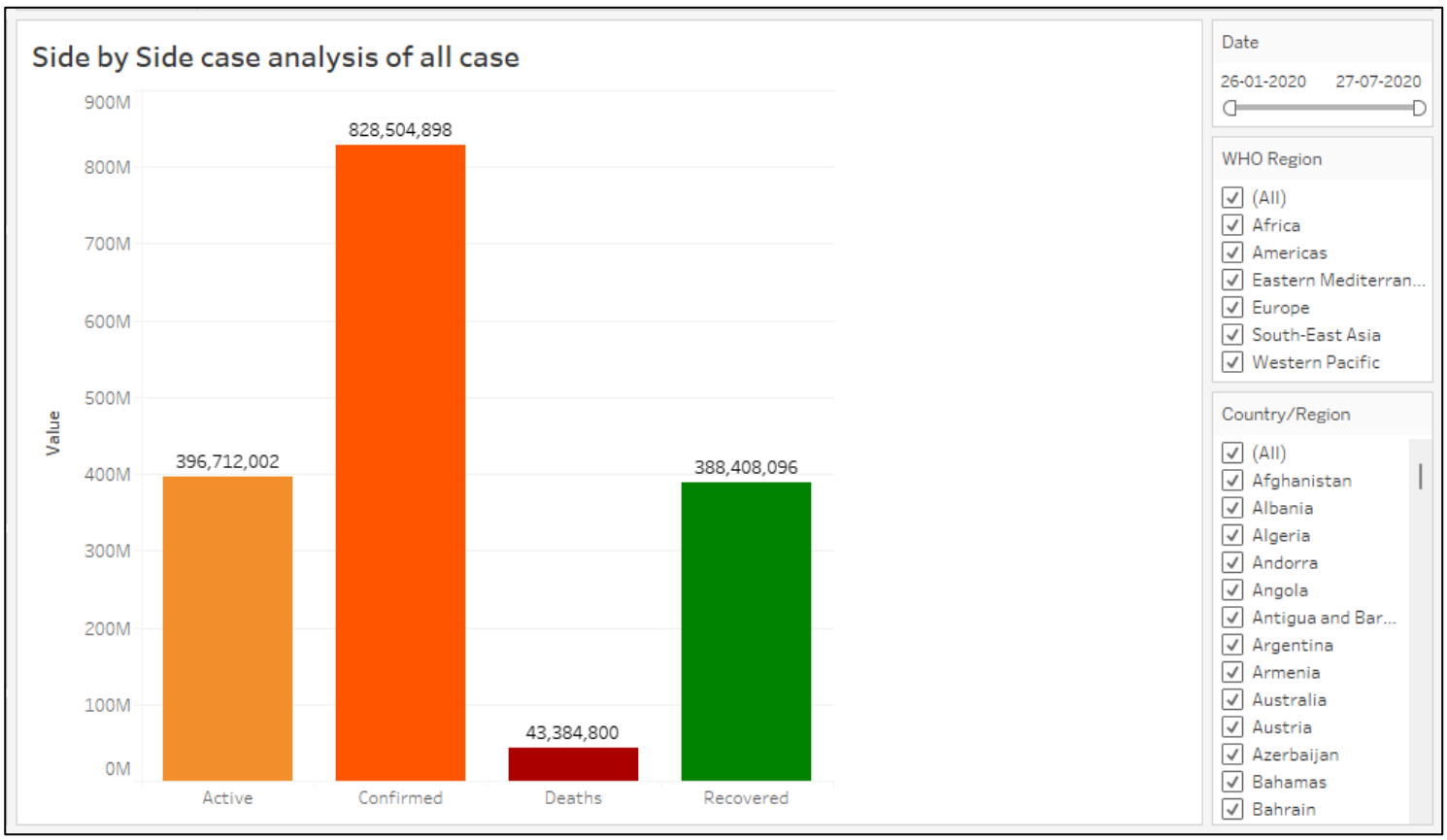
- Identify the selected WHO region and countries for analysis.
- Obtain the latest available data on confirmed, recovered, and death COVID-19 cases in each country within the selected region.
- Filter the data to include only cases for the selected period.
- Use a bar chart to represent the different types of cases side by side for each country within the selected region.
- Analyse the bar chart to identify trends and patterns in the different types of cases across countries within the selected region.

iii. Specific Requirements, functions, and formulas:

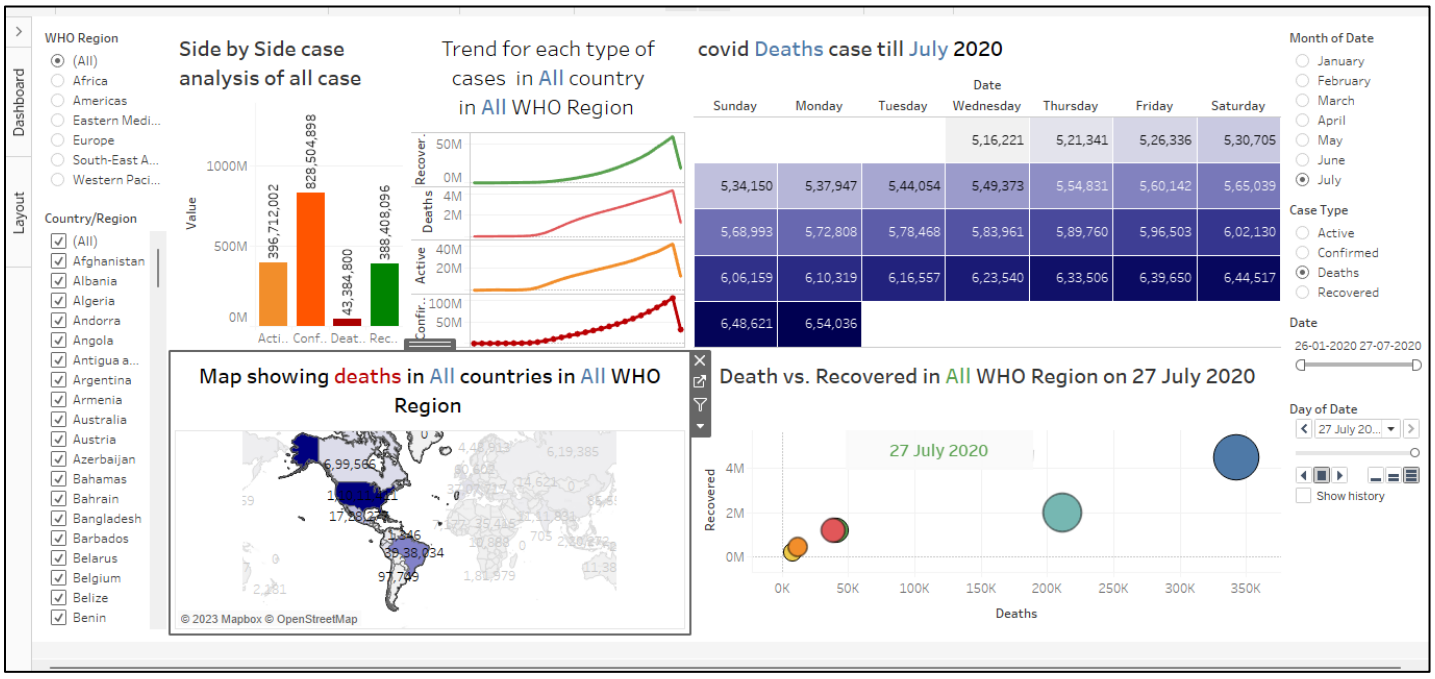
- Access to reliable and up-to-date data on COVID-19 cases by country and region.
- Knowledge of creating a bar chart.
- Familiarity with filtering data in Tableau to obtain data for specific time periods.

iv. Analysis results: This side-by-side bar chart shows the different types of COVID-19 cases (confirmed, recovered, active and deaths) across countries within the selected region over the selected period. This analysis can be used to get the latest results of covid.

v. Visualization: This bar chart is showing the different types of COVID-19 cases (confirmed, recovered, and deaths) side by side for each country within the selected WHO region. The bar chart will use a colour scale to represent each type of case, with different colours indicating confirmed, active, recovered, and death cases.



COVID19-ANALYTICS DASHBOARD



Future Scope

COVID-19 pandemic impacted societies around the world, the importance of data analysis in understanding its spread and impact has become increasingly clear. This analytics dashboard will provide insights into COVID-19 cases and can be an invaluable tool for public health officials, healthcare providers, and governments.

In the future, there are several potential directions in which research on COVID-19 analytics dashboards could be expanded:

1. Further development of predictive analytics: As the data is collected over time, it will become possible to develop models that can predict future trends in COVID-19 cases. This could help public health officials, governments and policymakers make more informed decisions about how to allocate resources like meals, vaccines and implement interventions.
2. Integration of demographic data: Understanding how COVID-19 affects different demographic groups is important for developing targeted interventions and addressing health inequities. By integrating demographic data into analytics dashboards, it will be possible to track disparities in COVID-19 outcomes and identify areas where additional resources are needed.
3. Exploration of new data sources: While existing data sources are valuable for understanding the spread of COVID-19, there may be additional data sources that could provide insights into the pandemic. For example, social media data could be used to track public sentiment around COVID-19 and identify areas where misinformation is spreading.

Overall, research on COVID-19 analytics dashboards is likely to continue to be an important area of study in the coming years as the pandemic evolves and new challenges arise. By developing more sophisticated analytics tools and integrating new data sources, it may be possible to track the spread of the virus and mitigate its impact on communities around the world more effectively.

References

[1] https://www.kaggle.com/datasets/imdevskp/corona-virus-report

[2] https://help.tableau.com/current/pro/desktop/en-us/dashboards_create.htm

Bibliography

[1] Microsoft Excel *(Software)*

[2] Kaggle.com *(website)*

[3] Tableau Desktop *(Software)*

[4] Tableau.com *(website)*