COVID-19 ANALYSIS REPORT

(Project Semester August-December 2024)

Submitted by:

Name: Hemanth Sai Kumar Pentakota

Registration No: 12220407

Program and Section: B. Tech CSE and K22YP

Course Code: INT233

Under the Guidance of

Pardeep Kumar 25237

Assistant Professor

Discipline of CSE/IT

Lovely School of Computer Science and Engineering

Lovely Professional University, Phagwara

CERTIFICATE

This is to certify **Hemanth Sai Kumar** that bearing Registration no. **12220407** has completed the INT233 project titled, **"Covid -19 Analysis"** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

Pardeep Kumar
Assistant Professor
School of Computer Science and Engineering
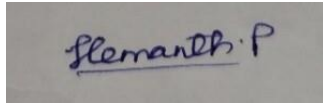Lovely Professional University
Phagwara, Punjab.

Date: 10-10-2024

DECLARATION

I, student of CSE 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.

Signature: Hemanth. P

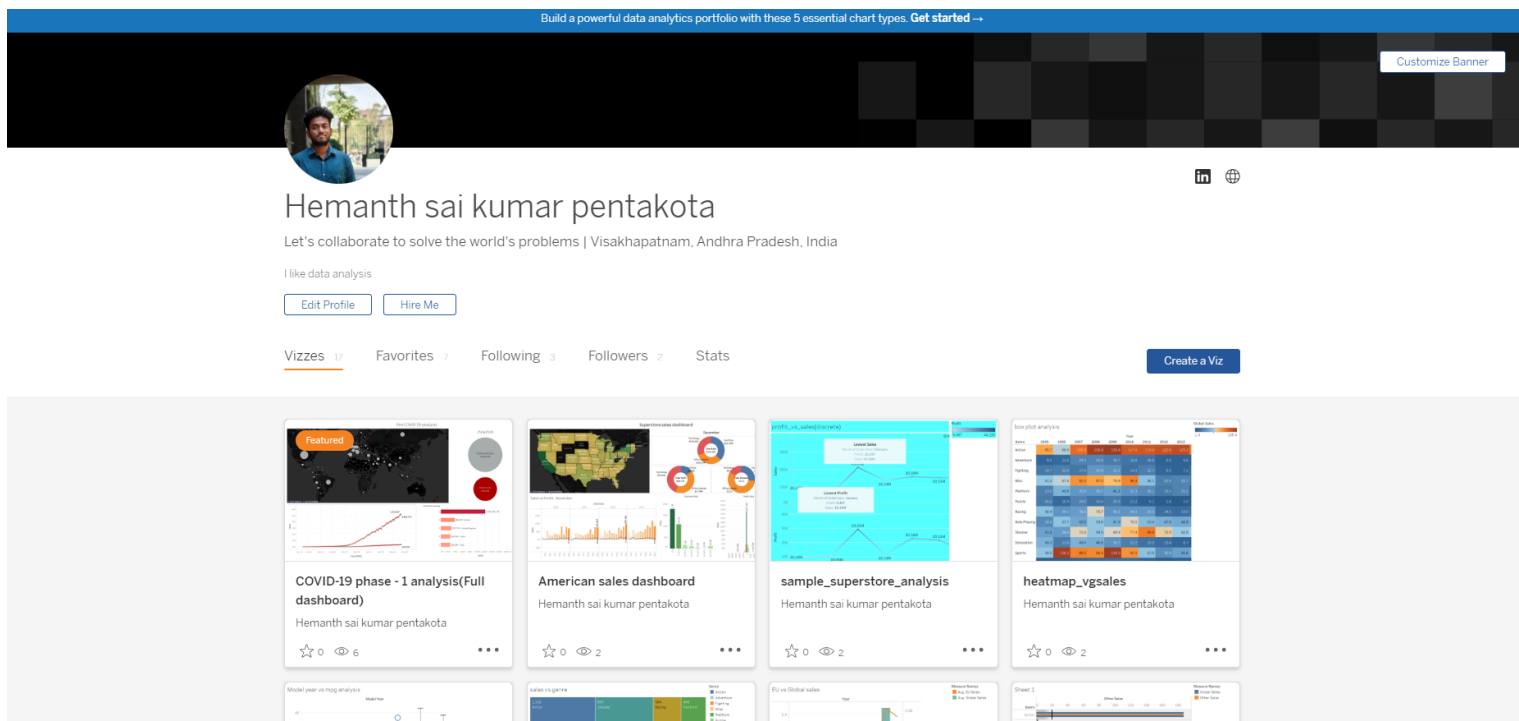**Date:** 10-10-2024

Registration No: 12212206

Name of Student: Hemanth Sai Kumar Pentakota

<u>UPLOADATION</u>

I also declare that all my work done is justifiable and has been uploaded to 'Tableau Public' and as a post in 'LinkedIn'.

Here are the screenshots of the posts:

HEMANTH SAI KUMAR PENTAKOTA (He/Him) • You
I like data analysis
1mo • Edited • 🌐

🚀 It's my first Tableau project! 🚀
During my summer training, I worked on analyzing COVID-19 data from the first six months of the pandemic. This project helped me sharpen my data visualization and analytical skills by exploring trends in global cases and deaths during the early days of COVID-19.

Data set is created by #GeeksforGeeks

📊 Key Features:
Choropleth Map: Visualizing global case distributions.
Dual Axis Line Chart: Tracking confirmed cases vs. death trends over time.
Dynamic Race Bar Chart: Ranking countries by the number of cases.
Filters: Dynamic interaction to explore confirmed and death cases for different dates.
Maps & Dashboards: Used to effectively communicate key metrics like total confirmed and death cases.
Pages shelf: Used to dynamically visualize the data across certain specified timeline.

This project was a great learning experience in data storytelling using Tableau. It enabled me to present complex data in a simple and interactive way. With the guidance of Mr.Pardeep Kumar sir, I was able to eliminate some outliers and he supported me throughout my learning curve.

Well this could be the beta version of my project. I would like to even expand my project and add more compact visualizations for better analysis. Stay tuned!

Feel free to check out my project and share your feedback! 😊

#Tableau #DataVisualization #COVID19Analysis #DataAnalytics #Dashboard #GeeksforGeeks #Salesforce #Excel

HEMANTH SAI KU...
I like data analysis
Visakhapatnam, Andhra Pradesh
Lovely Professional University

Grow your career with Premium
Try for ₹0

Profile viewers   26
Post impressions   37

✅ Changes saved.   ✕

Messaging

<u>Acknowledgment</u>

The satisfaction that accompanies the successful completion of this project would be in complete without the mention of the people who made it possible, without whose constant guidance and encouragement would have made efforts go in vain. I consider myself privileged to express gratitude and respect towards all those who guided us through the completion of this project.

I convey thanks to my project guide **Pardeep Kumar** of the Computer Science and Engineering Department for providing encouragement, constant support, and guidance whichwas of great help in completing this project successfully.

Last but not least, we wish to thank our parents for financing our studies in this college as well as for constantly encouraging us to learn engineering. Their personal sacrifice in providing this opportunity to learn engineering is gratefully acknowledged.

1. INTRODUCTION

This project focuses on analyzing the initial six months of the COVID-19 pandemic, tracking the spread and impact of the virus globally. Using data visualization techniques in Tableau, the analysis provides insights into the rapid escalation of confirmed cases and death tolls across different countries.

Objective

The objective of this project is to visually represent the progression of COVID-19 from its onset, helping understand the scale, speed, and geographic distribution of the virus. By presenting data from January to June 2020, my dashboard aids in identifying key patterns and comparing the outbreak's severity across different regions.

Key Features

1. Global Spread Visualization (Choropleth Map):
    - This map shows the spread of COVID-19 across countries, with larger circles representing higher case counts. It gives us a quick snapshot of which regions were hardest hit in those first six months.
    - The color-coded map differentiates between confirmed cases and deaths, making it easy to see both the intensity and impact of the outbreak in each area.

2. Growth Trajectory of Cases and Deaths (Dual-Axis Line Chart):
    - A dual-axis line chart shows how confirmed cases and deaths grew over time, illustrating the pandemic's exponential spread.
    - Key milestones are highlighted, showing when major spikes happened —possibly in response to events like lockdowns or public health measures.

3. Top 5 Affected Countries Ranking (Horizontal Race Bar Chart):
    - This chart ranks the top 5 countries by confirmed cases and deaths, showing which places were struggling the most during the initial wave. The U.S., Brazil, and the U.K. lead, giving a clear picture of where COVID-19 had its biggest impact.

4. Dynamic Case and Death Overview (Summary Circles):
   o Large, color-coded circles give the total confirmed cases and deaths as of a selected date, making it easy to get a quick snapshot of the situation at any point in time.
   o These figures update dynamically based on the date you choose, so you can see how things changed as the pandemic unfolded.

5. Interactive Date Filter:
   o The date slider let anyone pick a specific date to view the data for that day. This feature allows you to focus on key moments, like peaks or new waves, and see how cases and deaths evolved over time.
   o It makes the dashboard interactive and helps you spot patterns or shifts in case and death trends.

6. Comparison of Case Types:
   o A simple filter allows switching between viewing confirmed cases and death cases, so users can dive into either metric separately. This is useful for understanding the relationship between infection rates and mortality rates in each region.

7. Cumulative Case Growth by Date:
   o Another line chart shows the accumulation of cases over time, emphasizing the rapid rise in numbers as COVID-19 spread globally. This gives a big-picture view of the pandemic's growth, highlighting how quickly it escalated.

2. SCOPE OF ANALYSIS

This COVID-19 analysis project focuses on exploring the global spread and impact of COVID-19 during the initial six months of the pandemic. By examining data up to May 2020, the project provides insights into how the virus quickly affected various regions worldwide, and which countries were impacted the most. The analysis includes multiple dimensions to give a comprehensive view of the pandemic's early stages. Key areas of analysis include:

- **Geographical Spread**:

Visualizing confirmed cases and deaths across countries to understand the global reach and intensity of the virus. This allows for easy identification of COVID-19 hotspots and regional patterns.

- **Temporal Growth Patterns**:

Tracking the cumulative growth of cases and deaths over time using line charts to observe the rapid escalation in numbers. The analysis highlights significant spikes, helping to correlate with possible external factors like lockdowns or policy changes.

- **Country Comparison**:

Ranking the top 5 most affected countries by confirmed cases and death counts, providing a comparative view of how different regions experienced the pandemic. This section aims to spotlight the countries under the highest strain during this period.

- **Case Type Breakdown**:

Offering flexibility to focus on either confirmed cases or death counts through filters, which helps in distinguishing between infection rates and fatality rates. This breakdown helps us to understand the severity of the virus in different contexts.

- **Date-Specific Observations**:

This feature helps to select specific dates for more detailed examination of COVID-19 status on that day, helping to capture the day-by-day evolution of the virus. This feature supports analysis of specific points in time, like peaks or the initial outbreak days in each country.

- **Growth Trend Analysis**:

Observing the overall trend of COVID-19 case accumulation, highlighting the exponential nature of the virus's spread. This section emphasizes the urgent need for containment and public health measures by showing how quickly cases surged in a short period.

3. SOURCE OF DATASET

This project leverages a COVID-19 dataset provided by **Geeks for Geeks** through their course *Complete Machine Learning and ML.* Geeks for Geeks, widely recognized as a trusted platform for programming and computer science education, offers resources and courses in various technical domains, catering to learners from beginner to advanced levels. Their dataset includes crucial data points from the early stages of the COVID-19 pandemic, allowing for a detailed analysis of the virus's global spread and impact.

Dataset Details

The dataset contains some of the following important key fields:

i.   **Case_Type**: Specifies whether the record pertains to confirmed cases or deaths, providing insights into the type of case data.

ii.   **Cases**: Lists of the number of cases reported on each date for a given country or region.

iii.   **Country_Region** and **Province_State**: These fields identify the location of each record, allowing for region-specific analysis.

iv.   **Lat** and **Long**: The latitude and longitude coordinates for each location, facilitating accurate geospatial plotting on maps.

v.   **Population**: Reflects the total population of each country or region, which is crucial for calculating infection and death rates relative to population size.

vi.   **Data_Source**: Indicates the origin of the data, which in this case is "2019 Novel Coronavirus" as recorded in each entry, denoting the focus on the novel coronavirus outbreak.

4.1 REGION WISE CASE ANALYSIS
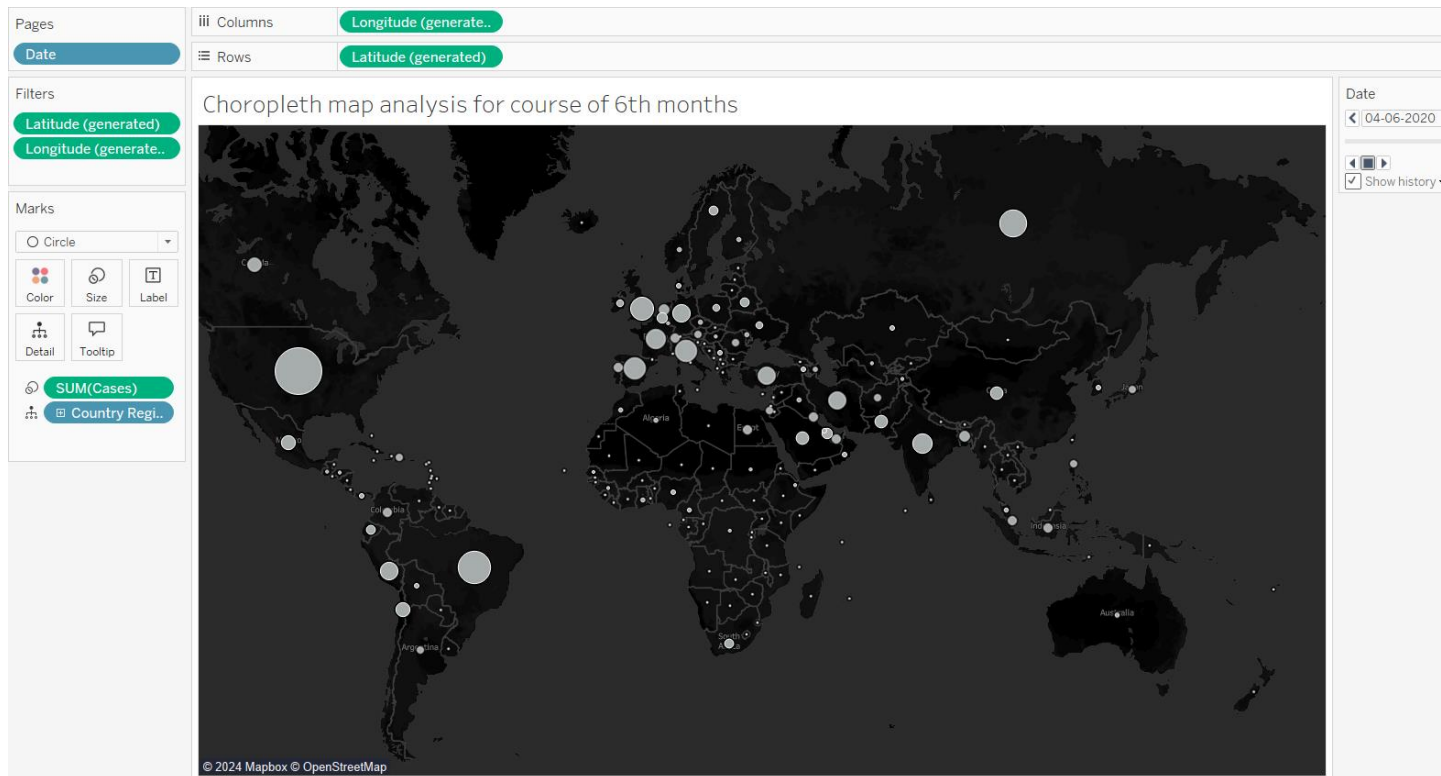
The choropleth map provides a powerful visual representation of how COVID-19 spread across the globe during its initial six months (January to June 2020). By plotting cumulative cases by country, it offers a clear understanding of the pandemic's geographic distribution over time. This visualization highlights the region's most severely affected by the virus, underscoring the rapid escalation in specific countries and continents.

This map leverages size (bubble scale) and location to indicate the intensity of cases per region, making it easy to identify global hotspots. Notably, countries such as the U.S., Italy, and Brazil show significant outbreaks early on.

Map Elements:

Bubbles (Markers): Represent the total number of cases for each country. Larger bubbles indicate higher case counts.

Dynamic Date Filtering: Users can explore the spread of COVID-19 by adjusting the date slider, which enables an understanding of how the virus proliferated over time.

Color Scheme: A dark background helps emphasize the size of the bubbles, enhancing readability.

Regions Highlighted:

- North America (particularly the U.S.) displays one of the largest bubbles, indicating a sharp rise in cases.

- Europe, specifically Italy and Spain, also showcases substantial early outbreaks.

- South America (Brazil) starts to see significant growth by May-June.

- Asia, notably China (initial epicenter), and India see notable shifts in cases as time progresses.

Key Observations:

By June 2020, COVID-19 cases were widely distributed, with significant clusters in every continent.

The map also reveals the pandemic's transition from Asia (initial stages) to Europe and subsequently to the Americas

4.2 CASES VS DEATH ANALYSIS
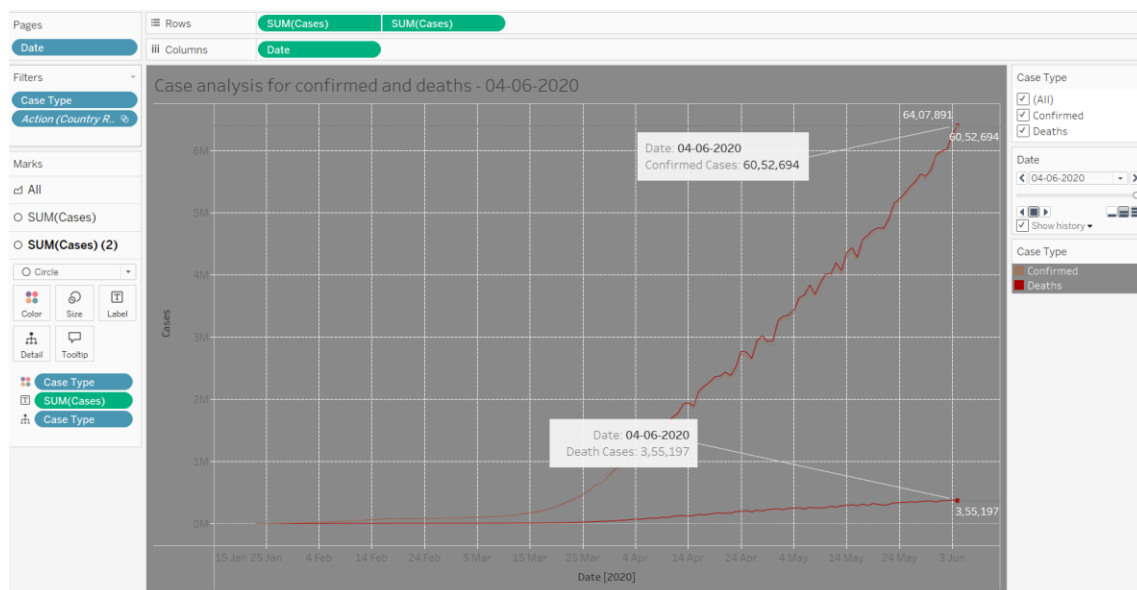
This visualization provides a cumulative line chart comparing **confirmed cases** and **deaths** globally over the first six months of the pandemic. It helps track the progression and severity of COVID-19, offering a clear timeline for how the virus surged worldwide.

The dual-line setup for **confirmed cases** and **deaths** emphasizes the stark difference in the total counts while maintaining a focus on the mortality trend relative to the growing infection base.

Detailed Analysis

- **Data Source**: Tracks global cumulative numbers from January 15 to June 4, 2020.

- **Key Elements**:

  o **X-Axis**: Represents time (dates from January to June 2020).

  o **Y-Axis**: Total number of cases (confirmed and deaths).

  o **Line Details**:

    ▪ **Confirmed Cases (Orange Line)**: Rapid growth after mid-March, culminating in over **6.4 million cases** by June 4, 2020.

    ▪ **Deaths (Brown Line)**: A slower yet significant upward trend, with global deaths reaching approximately **355,000** by the same date.

  o **Data Points & Tooltips**: The interactive points, such as the one on **June 4, 2020**, highlight specific values for both metrics:

    ▪ Confirmed cases: **6,407,891**

    ▪ Deaths: **355,197**

- **Trend Observations**:

  o **Exponential Growth**: A noticeable surge in confirmed cases around **late March to May**, correlating with the spread of the virus across Europe and the Americas.

  o **Death Curve**: Although slower in growth than confirmed cases, it still demonstrates a steep rise, reflecting the pandemic's severity.

- **Dynamic Filtering**: This chart allows users to toggle between confirmed and death cases using the **Case Type** filter, offering flexibility in analysis.

Key Takeaways:

- The sharp divergence between confirmed cases and deaths highlights the scale of global infections while emphasizing the fatal impact.

- Critical time frames like late March through April saw the most significant spikes, correlating with widespread outbreaks in multiple regions.

4.3 TOP 5 COUNTRIES RANKING (on basis of cases)
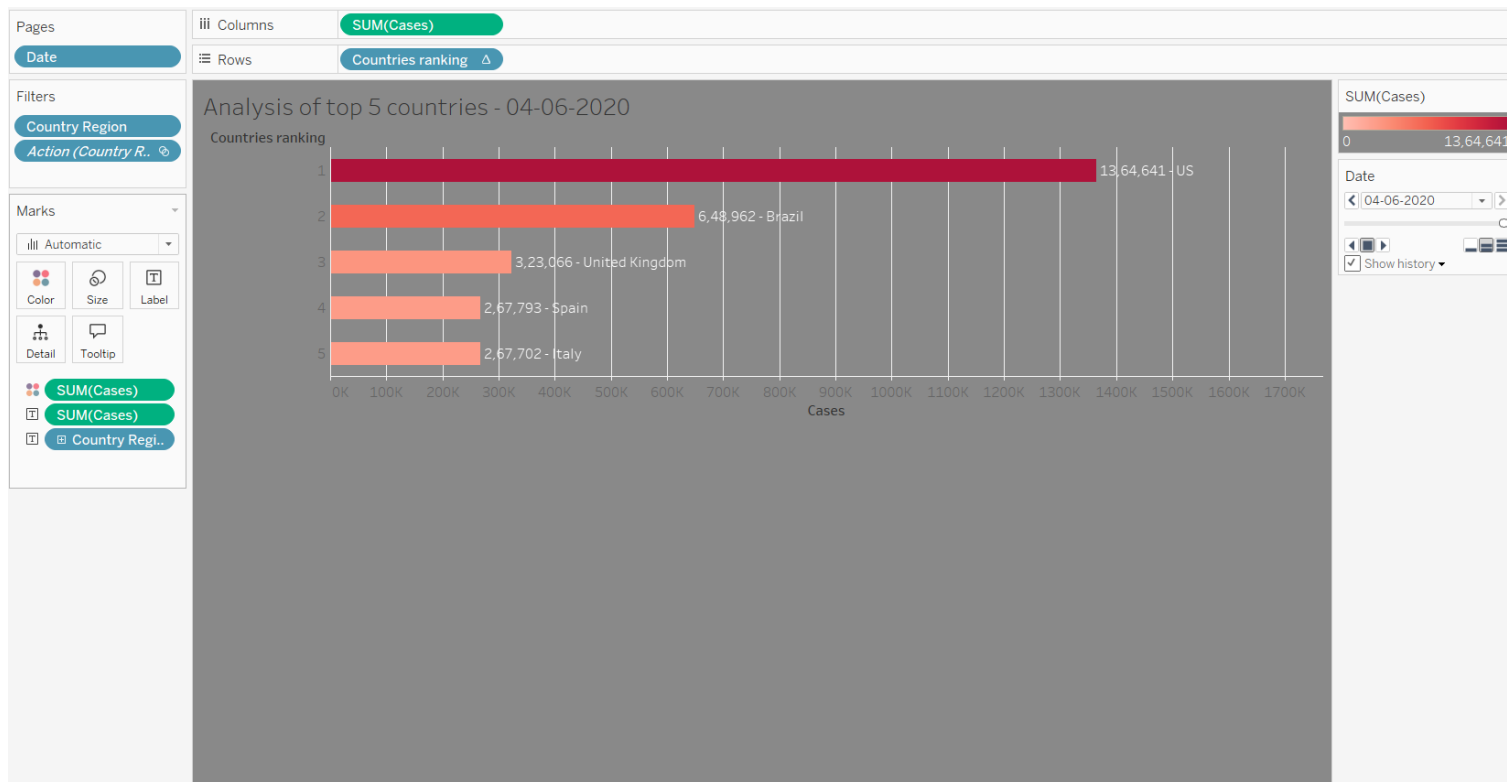
This chart uses the **Rank Unique** function to display a bar chart ranking the top five countries by COVID-19 cases as of the selected date (April 12, 2020, in the provided screenshot). It helps visualize the global distribution of the pandemic's impact, showing the countries most affected at a glance.

Detailed Breakdown:

- **Chart Type**: Horizontal Bar Chart.

- **Ranking Function**:

  Rank Unique assigns each country a unique rank based on the number of cases, ensuring no duplicate rankings. This ensures clear differentiation, even when countries report similar numbers of cases.
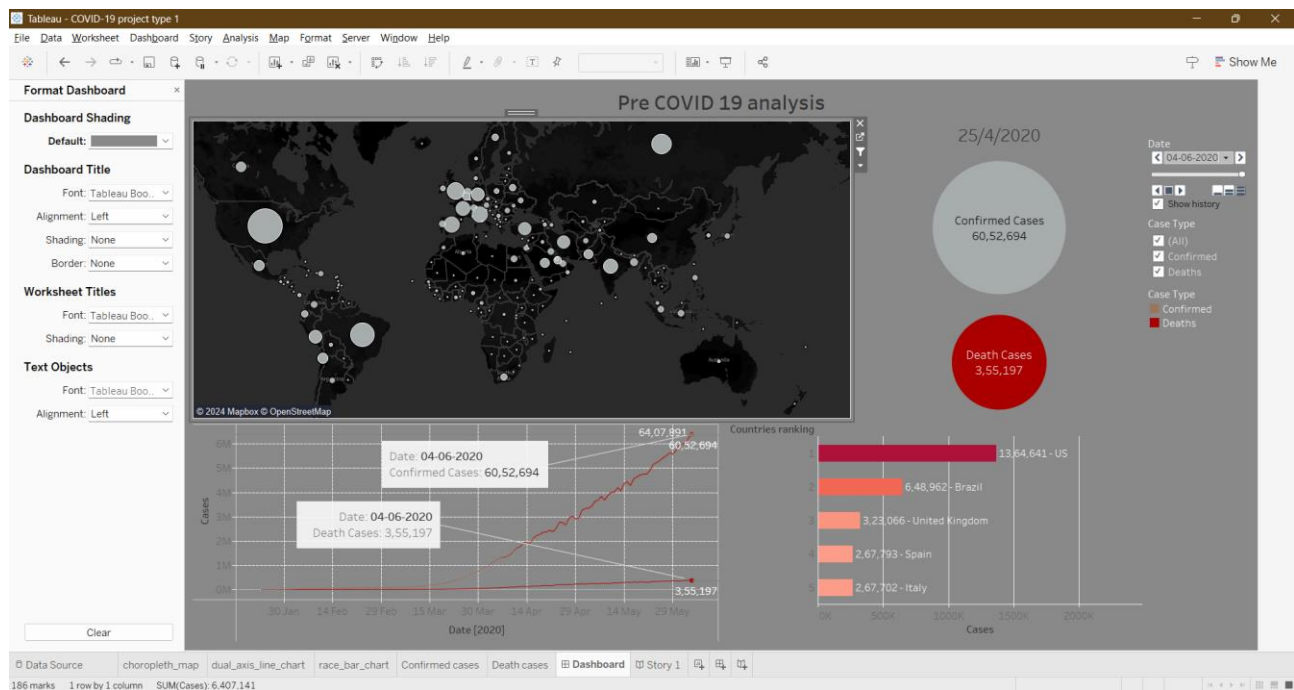
- **Date Filter**:

  The chart is dynamic, allowing users to analyze top-affected countries at any specific date within the six-month period.

- **Case Summary**:

  Each bar is color-coded based on the total number of cases, ranging from lighter to darker shades of red for quick visual impact.

4.4 FINAL DASHBOARD

5. STORIES

The project also consists of an extension of important insights which are shown in the stories section, a few of those feature: