Analysis of Covid-19 Daily Results and Information about Patients using SQL and PowerBi

Sulejman Karamehić [a,1,*]

[a] International Burch University, Sarajevo, Bosnia and Herzegovina
[1] sulejman.karamehic@stu.ibu.edu.ba
* corresponding author

ARTICLE INFO

ABSTRACT

COVID-19 is a contagious disease caused by a virus, the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The first known case was identified in Wuhan, China, in December 2019. The disease quickly spread worldwide, resulting in the COVID-19 pandemic. The outburst of COVID-19 pandemic had tremendous effect on the whole world and analysis of the data can be meaningful in many ways to better understand the effect it had in our society. This paper aims in the direction of analysis of COVID-19 daily information based on country and continent level in terms of understanding the number of cases and deaths and their relationship, besides this is aims to better understand the vaccination number by country and effect of cases/death how they have affected these numbers. The solution was created on analysis of a dataset that contains daily information on each country, and using MySQL, SQL and PowerBi to generate the results for this work in way of query results which have been transformed to visuals using PowerBi for better understanding for further research work on this topic.

1.   Introduction

The world is currently facing many difficult challenges. The worst humanitarian disasters of the 20th century were the COVID-19 pandemic and World War II. The COVID-19 outbreak is acute respiratory syndrome and is still ongoing It was declared a pandemic by the World Health Organization (WHO) on March 11, 2020. Additionally, the outbreak first appeared in Wuhan in December 2019 and continues to spread. Rapid outbreak in mainland China and around the world, causing panic and severe casualties for people's lives and the economy. [1] The virus spreads directly from person to person contact and caused many deaths. The COVID-19 pandemic has taken over the world for over a year. Many countries have announced various policies to control the spread, including: B. Work from home; Study at home, lockdowns, travel restrictions, restrictions on the number of people in public places location and other guidelines. It also created a new standard for society. Such as wearing masks frequently, washing hands, and maintaining physical distancing. [2] This condition certainly affects nearly every aspect of life, especially health, social, environmental, economic and business. Accelerating Digital Transformation Programs by various organizations and companies during the pandemic. even online Shopping that avoids contact with people and cashless payments are now essential. Meetings, lectures, promotions, seminars and conference daily activities are also available it is kept online to prevent its spread. The pandemic is also impacting the environment by reducing air pollution. Additionally, lockdowns and working from home are creating guidelines. More people prefer to stay at home, less traffic on the streets and better air quality in

urban areas. In addition, people prefer to ride bicycles rather than use public transportation Close contact between passengers during local travel. The fight against COVID-19 is being fought on the front lines by paramedics and volunteers. In addition, various studies are being conducted to combat this and find solutions. deadly pandemic. [3] Many opportunities are offered to provide technology-based services solution. In his over a year of research on COVID-19 big data. This technology is helping to track cases, monitor epidemics, and spread the virus. and human movement monitoring, precautions, treatment and medication Development. Moreover, advanced technology and architecture make big data to solve and deal with various life problems inevitably. Pandemic. Analysis of social media related to COVID-19 helps solve problems in social life by capturing public opinion, concerns, and policy responses implemented.

2.   Related Work

Since this topic has been in the mainstream for the past few years there has been done a lot of work on this topic in all specters of the scope, following this information many rearchers decided to work on this topic and some of the most meaningfull for this work will be mentioned.

[4] states that R. Wang, G. Hu, C. Jiang, H. Lu, and Y. Zhang compared the prediction of patterns using 3 different methods and comparing their graphs to one another. The Lowest Squares Approach SIR model, the Particle Swarm Optimization SIR model, and the Classical Logistic Regression SIR model are these models. In the end, the chart displays some patients with a novel type of cardiac pneumonia together with Y-axis date. We learn that the data is plotted as a curve by observing the three patterns. "In this study paper [5] proposed by V.Z.Marmarelis, the public figures of daily updated confirmed incidences of Covid-19 from University John Hopkins were analyzed. [5]. The main modeling component for the method is RM as defined by the Riccati Equation. In this study article [5] suggested by V.Z. Mararelis etal. [5], the public statistics of daily updated confirmed incidences of Covid-19 from University John Hopkins were analyzed. The method's primary modeling component is RM, as defined by the Riccati Equation. Applying the equation further, we discover 5 separate factors and their dependency on the daily increase in the number of cases.

In their research paper [6], Yazeed Zoabi, Shira Deri-Rozov, and Noam Shomron agreed that prompt and precise COVID-19 diagnosis is made possible by accurate SARS-CoV-2 screening, which thus lessens the burden on healthcare systems. To determine the likelihood of infection, prediction models have been developed utilizing a variety of parameters. The model anticipated 0.90 auROC in the prospective test set (area under the receiver operating characteristic curve).

The authors of the research paper [7], Enis Karaarslan and DoanAydn, stated that the COVID-19 episode demonstrated that the globe was not willing to spread the virus so quickly. The efficient use of information technology is essential in reducing the negative effects of an epidemic or pandemic. They proposed an epidemic management system (EMS) that depends on the free and prompt exchange of information between nations and organizations. They have been utilizing the MPISA paradigm, which enables the integration of many platforms and provides a fix for the scalability and interoperability problems.

3.   Material and Method
3.1   Dataset

The datasets that we have used in this work is The World Dataset of COVID-19 from kaggle. The author of this dataset is Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU) [8]. The data contains daily records for each country and their information about the covid cases in specific country, also their deaths on a daily level and the total number of deaths and vaccination information for each country. The following table represents the most important columns used for this work which are explained in Table 1.

Table 1. Dataset Attributes

Attribute	Description
iso_code	Country code
continent	Continent of the country
location	Country
date	Day representation in date format
population	Population of the country
total_cases	Total cases in specific country
new_cases	New cases in a day for a country
total_deaths	Total deaths in specific country
new_deaths	New deaths in a day for a country
total_vaccinations	Total vaccinations in specific country
new_vaccinations	New vaccinations in a day for a country

Each row in this dataset represent daily information about country's "battle" with covid and information about new cases, deaths and vaccinations. The main goal of this work is to create meaningfull results and visualizations about covid and it's impact on our society and how to analize large amount of data. In that purpose it has been decided to use SQL for result generation and PowerBi to create the visuals from the SQL query results, which will be talked more indepth in the following chapters of this work.

3.2 MySQL

MySQL is an open source relational database management system (RDBMS). A relational database organizes data into one or more data tables, where the data can be related to each other. These relationships help structure the data. SQL is the language programmers use to create, modify, and extract data from relational databases, and to control user access to databases. In addition to relational databases and SQL, RDBMS like MySQL work with operating systems to implement relational databases in computer storage systems, manage users, allow network access, and perform database integrity testing and backups. make it easier. [9] MySQL is free and open source software under the terms of the GNU General Public License, and is also available under various proprietary licenses. MySQL was owned and sponsored by the Swedish company MySQL AB, which was acquired by Sun Microsystems (now Oracle Corporation). When Oracle acquired Sun in 2010, Widenius sold his open source MySQL project to create his MariaDB. [10]

3.3 SQL

Structured Query Language (SQL) is a standardized programming language used to manage relational databases and perform various operations on the data in them. First developed in the 1970s, SQL is regularly used not only by database administrators, but also by developers writing data integration scripts and data analysts setting up and running analytical queries. [11]

SQL is used for the following:

- modifying database table and index structures
- adding, updating and deleting rows of data
- retrieving subsets of information from within relational database management systems (RDBMSes) - his information can be used for transaction processing, analytical applications, and other applications that require communication with relational databases.

SQL queries and other operations take the form of commands written as statements and packaged into programs that allow users to add, modify, or retrieve data from database tables. A table is the most basic unit of a database and consists of rows and columns of data. A single table contains records, each stored in a row of the table. A table is the most common type of database object or structure that stores or references data in a relational database. Each column in the table corresponds to a category of data (such as customer name or address), and each row contains data values from

intersecting columns. A relational database is relational because it consists of tables that are related to each other. For example, a SQL database used for customer service might contain tables of customer names and addresses, and other tables containing information about specific purchases, product codes, and customer contact information. A table used to keep track of customer contacts is typically placed within another table that stores customer details such as name and contact information, using a unique customer ID, called a key or primary key. Browse customer records. [12]

3.4   PowerBi

Power BI is a collection of software services, apps, and connectors that work together to transform independent data sources into coherent, visually immersive, and interactive insights. Data can be an Excel spreadsheet or a collection of cloud-based and on-premises hybrid data warehouses. Power BI makes it easy to connect to your data sources, visualize and discover the data that matters, and share it with anyone. [13]

Microsoft Power BI is used to find insights into your organization's data. Power BI helps you connect disparate datasets, transform and cleanse your data into data models, and create charts or graphs that visually represent your data. All of these can be shared with other Power BI users in your organization.

Data models created from Power BI can be used in your organization in many ways. For example, you can tell stories through charts and data visualizations, or explore "what if" scenarios in your data. Power BI reports can also help answer real-time questions and forecasts to ensure departments are meeting business metrics. Power BI can also provide administrators or managers with executive dashboards, giving them more insight into how their departments are performing. [14]

3.5   Proposed Method

For the experimental part of this work the following has been done. Firstly since this dataset is in csv format it had been decided to process this data into a SQL table for query generation this step had been done by using MYSQL on a local machine and the enormous collection of data had been imported to a tabel called „WorldCOVID", following this step the data was stored in the following table and the amount of data was gigantic since it contained daily information about country's. Next it had been decided to create „scenarios" on how to generate a query using SQL for a specific „scenario" to get the needed information. Following the generation of results , the next step was to create some eye appealing visual for everyone to be able to understand what each of the scenarios represents and how this data can be usefull. The „scenarions" and the results that have been generated will be discused and explained in the following section of this work.
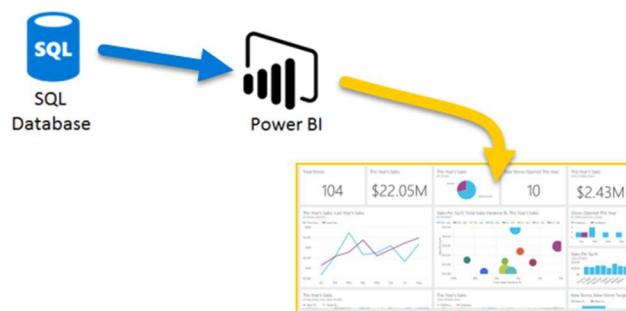

Figure 1. Result process visualization

4.   Results and Discussion

First step prior to geting our nice visual representations of results was creating scenarios for which the querys will be written and thus 3 scenarios have been created for which an SQL querry needed

to be written to generate the data needed for the results shown in the visuals below. For the simplicity of the work each scenario will be explained then the visualization of the results will be explained.

4.1　　Scenario 1

Here, we examine the continents' confirmed instances and fatalities. We are interested in the confirmed total cases, fatalities, and vaccines for each of these continents.
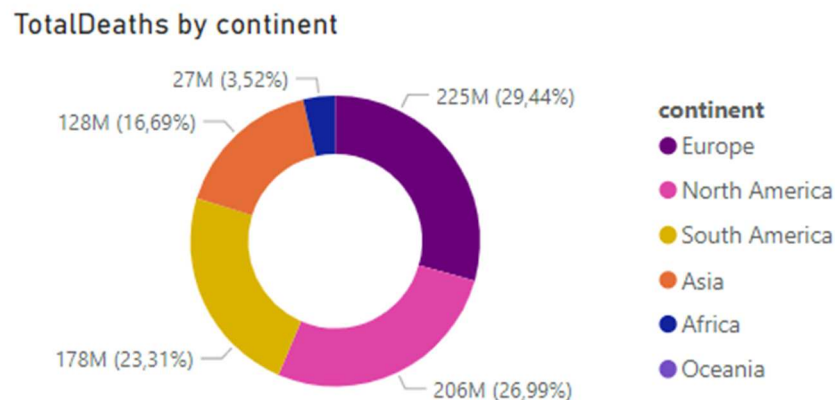

Figure 2. Total Deaths by Continent

This was sorted by the total cases in descending order, with Europe having the highest number of cases, followed by Asia, NA, SA, Africa, and Oceania. We can also see the continents and their respective total cases and total deaths.
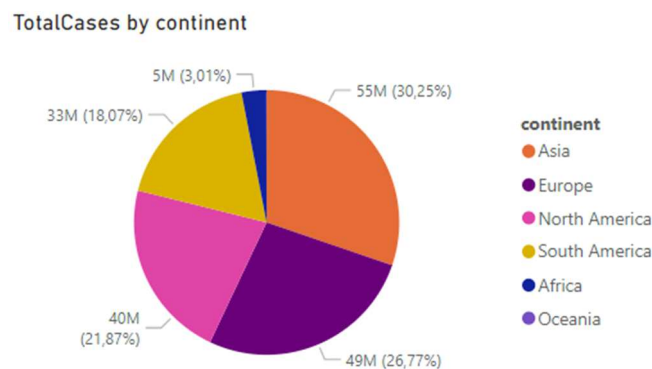

Figure 3. Total cases by Continent

4.2　　Scenario 2

Explores a different way of looking at countries in terms of total cases, total fatalities, percentage of infected population, and percentage of dead population.

Because columns in the dataset capture the cumulative cases and deaths, it is recognized as the nation's highest infection count and highest death count. Together, they indicate same significance as the nation's total cases and total fatalities. As a result, the highest infection count and death count are just their respective columns' maximum values. Below are the query and result.
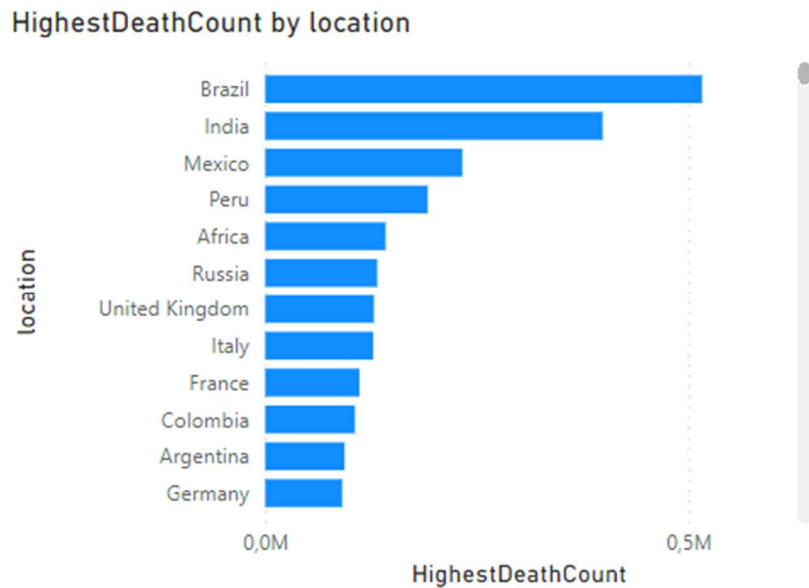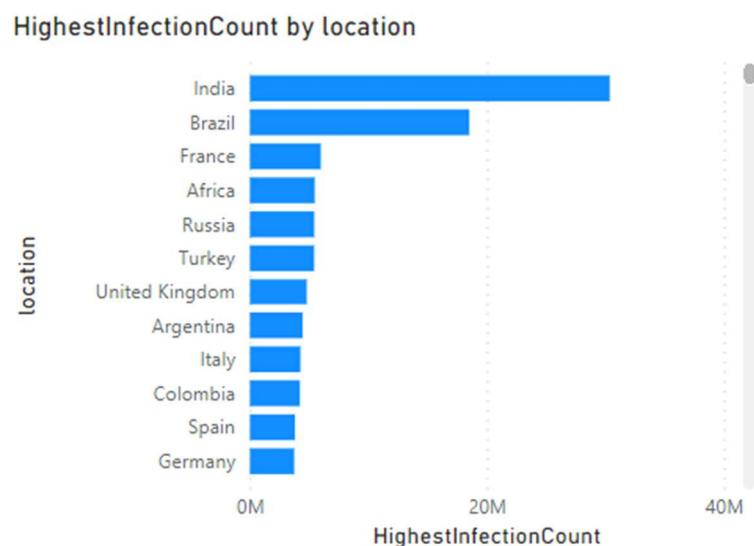
Figure 4. Highest Death Count by Location

Figure 5. Highest Infection Count by Location

4.3   Scenario 3

This examines the daily new cases, new deaths, cumulative cases, cumulative deaths, new vaccinations, and cumulative vaccinations that have been reported for the nation. In order for us to better understand and react to these insights, visualizations were created for this covid-19 case study investigation.

The outcome includes data on the nation's daily recorded population, cases, deaths, and vaccines. It also includes data on cases, cases' cumulative sums, deaths' cumulative sum, and vaccinations' cumulative total. The visuals will make it easier to understand the insights from these outcomes.
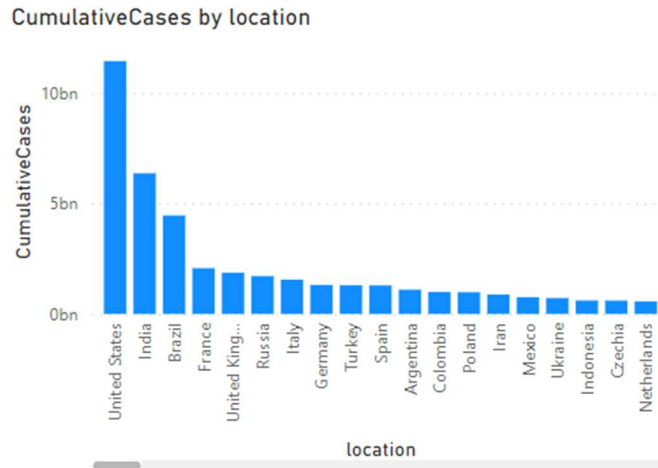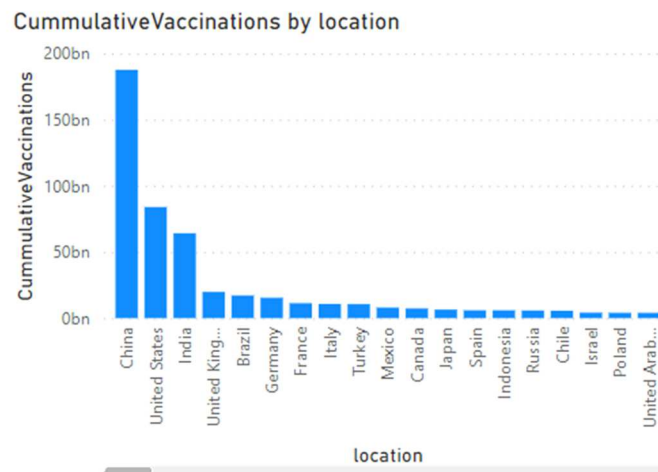
Figure 6. Cumulative Cases by Location

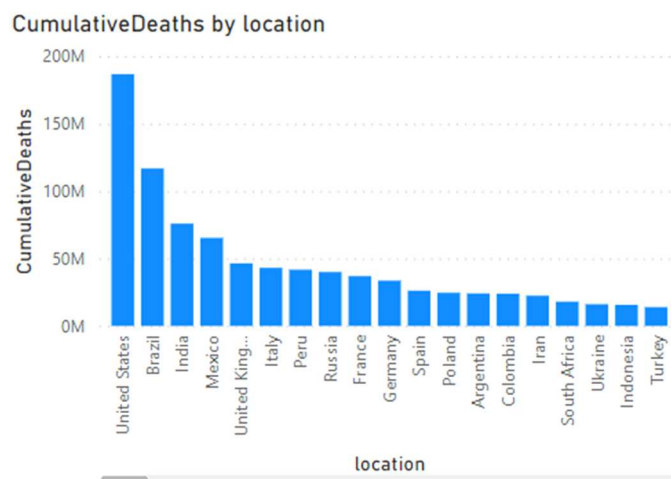
Figure 7. Cumulative Vaccinations by Location

Figure 8. Cumulative Deaths by Location

In this paper, the main focus was on creating some valuable information about covid and the impact it had. Besides that how some more evolved parts of our society how implemented vaccination to reduce the number of cases and to try to lower the number of deaths globally, we can see from the results above that Europe had the most casualties in respect to this dataset which doesn't mean that this data is finite since these numbers change on a daily basis since this problem is even now present an finite solution for it has not been found, but nonetheless this work has provided some meaningful information as can be seen for example China has stopped in some number the cases and deaths as they can be seen the leading country in vaccinations numbers from our data and we can see that America has the most cumulative cases and deaths, also we can clearly see that not so well developed country like Brazil are having enormous numbers of deaths and cases and their vaccination numbers and not so good. As it had been said this work can be used as timestamp for later research as how it was a certain point in time in our world, as later can be used for comparison and better analysis how the world has dealt with this problem and how the numbers are daily ever so changing in our society. Lots of research has been on this topic but it is mostly in the way of prediction of cases and how it will affect our society, whilst the research for this project was done not many works have been found strictly that have been focused on the analysis in depth of this problem it is almost as a side part of all of the work they go through analysis briefly and generate the results they need, and that is where it is believed that the value of this work is solely on the analysis as it can be referenced as valuable work on top of which even more meaningful research can be done and can conclude more important results or just for comparison as said earlier for future works that will reference this problem as it become a thing of the past in the future.

5.   Conclusion

The cases, deaths, and immunizations that are the topic of this SQL COVID-19 exploration project. Given that the goal of this project was to use fundamental SQL knowledge in researching a real-life problem, I think more studies can be done (COVID-19 Outbreak). The queries include window functions, CTEs, and fundamental SQL functions. The main goal was acomplished which was to create some logical cases/scenarios to generate valuabe and logical information about covid and how it has afftected each country and continet, due to sheer size of this dataset there can be many more valuable researches done on this dataset which could be a more indepth project that could be even ofa bigger value. Furthermore more reaserch is needed to better understand the effects the Covid-19 virus has done and casulties it has done to our society, but nonetheless this project is a good introduction to see the consequences it has done globally.

References

[1]   W. Sirinaovakul, T. Eiamyingsakul, N. Tubtimtoe, S. Prom-on i Taetragool, "The Relations Between Implementation Date of Policies and The Spreading of COVID-19," International Conference on Big Data Analytics and Practices (IBDAP), Bangkok.

[2]   Y. Zheng, "Estimation of Disease Transmission in Multimodal Transportation Networks," Asian J. Public Opin.

[3]   W. He, J. Zhang i W. Li, "Information Technology Solutions, Challenges, and Suggestions for Tackling the COVID-19 Pandemic," Int. J. Inf. Manag, 2020.

[4]   G. H. C. J. H. L. a. Y. R. Wang, "Data Analytics for the COVID-19," IEEE 44th Annual Computers, Software, and Applications, 2020.

[5]   Z. Marmarelis, "Predictive Modeling of Covid-19 Data in the US: Adaptive Phase-Space Approach," IEEE Open Journal of Engineering in Medicine and Biology.

[6]   Y. D.-R. S. &. S. Zoabi, "Machine learning-based prediction of COVID-19 diagnosis based on symptoms," npj Digit. Med. 4, 3 .

[7]   D. A. Enis Kararslan, "A COVID 19 pandemic support for artificial intelligence and resource management systems," e,ISBN 9780128245361, Academic Press.

[8]   C. f. S. S. a. E. (. a. J. H. U. (JHU), "Kaggle," [Na mreži]. Available: https://www.kaggle.com/datasets/aditeloo/the-world-dataset-of-covid19.

[9]   Oracle, "mysql," [Na mreži]. Available: https://www.mysql.com/.

[10]   NaN, "Wikipedia," [Na mreži]. Available: https://en.wikipedia.org/wiki/MySQL.

[11]   P.   Loshin,   "techtarget,"   [Na   mreži].   Available: https://www.techtarget.com/searchdatamanagement/definition/SQL.

[12]   C. Academy, "CodeAcademy," [Na mreži]. Available: https://www.codecademy.com/learn/learn-sql.

[13]   J.   Scardina,   "techtarget,"   [Na   mreži].   Available: https://www.techtarget.com/searchcontentmanagement/definition/Microsoft-Power-BI.

[14]   Microsoft,   "learn.microsoft,"   [Na   mreži].   Available: https://learn.microsoft.com/en-us/power-bi/fundamentals/power-bi-overview.