Logo, company name

Description automatically generated

**Advanced Database Topics**

**Group Name:** DATA ANALYZERS(ADT PROJECT GROUP 5)

**Student Names:** Aaditya Pradipbhai Parekh (110084734)

Devanshi Shah (110083753)

Mayank Chauhan (110076457)

Kshitiz Choudhary (110075501)

**Project Title:** Real-World Data Analysis Visualization of COVID-19 Dataset

**Project Proposal**

1. **Introduction:-**

With the wide Spread pandemic COVID-19, the current situation was little better than the past year which made almost the entire world go on lockdown. The death counts and the spread of the virus are alarming and the situation is improved in some countries but still, but in some countries, the situation is still questionable and if it does not improves the world can again slip into the pandemic which can cause disastrous economic depression affecting every single individual. Data science as a science discipline is influenced by informatics and computer science. Data Visualization is the subset of data science and it involves communicating the finding of data scientists and data analysts effectively through graphical representation.

It is very important to understand the domain COVID-19 which is an infectious disease caused by a new virus of SARS-MERS. The tool adopted to carry out this task is the Python programming language which is having many versatile data science libraries to handle data visualization. Python is currently used by the active scientific computing community which allows for greater flexibility because of its highly simplistic and flexible way of writing and extending the code.

1. **Problem Statement:-**

COVID-19 is far more than just a data science issue—it’s a massive public health problem that has resulted in many deaths and is throwing a harsh light onto how we structure our society when it comes to important things like the availability and affordability of healthcare, worker’s rights, and even freedom of movement.

But  I think it’s important to look at the situation from a data science perspective. We’ve all seen curves on Twitter—exponential, flattened, and otherwise—plotted in Excel and been reassured by them or scared by them, or wondered whether we could trust them. That’s a data science question, and there are many similar ones that I want to address here, in the hopes that what I write will inspire others to think about the data and feel more empowered about what to do in this situation.

So by analysis of the data reports being generated as per the records, we can analyze the Covid Situation, its ups and downs, and its impacts. This report can help the society to aware every individuals and help the society to take necessary actions if any similar kind of situation occurs in the future.

1. **Motivation:-**

The problem is interesting because of the dataset as it is highly structured and yet somewhere messy in ways that are meaningfully related to real-life problems. The data is related to the cumulative number of confirmed cases per day in each Country. With this dataset, we will be able to explain why the COVID-19 spread around the world that fast. What is the impact in terms of the confirmed cases and death at the country level which makes it more interesting? With this analysis, we will be able to understand and predict if in the future the same kind of pandemic occurs how will the world react. With this analysis, we will be able to understand how important the role of medicines and knowledge in any pandemic is.   
Giving the solution to the problem may be challenging due to the data which is collected should be standardized while collecting. And also the data should be analyzed in a standardized way. The other main challenge will be searching for the relevant information sources for the data.   
From this analysis of the COVID-19 data from the real world, it will be a benefit to understand which country is lacking and was not able to control the pandemic in time. It will benefit those countries which are required to increase investment in public health and emergency planning is essential to overcome the current pandemic, as well as future public health emergencies. By comparing the severity of COVID-19 with other epidemic outbreaks like EBOLA 2014, and MERS 2012 it will be a benefit to understand the strategies and planning required if we face the same kind of epidemic in the future. As the COVID-19 is still surviving in some parts of the world with this analysis we will be able to identify the countries which require more awareness and medical requirements. As we cannot confidently accept the fact that the pandemic is over it can easily start over again so with this analysis it will be helpful to understand which country is needed to increase preparedness, readiness, and response actions for public health emergencies.  
The analysis of the data for the COVID-19 will help to understand and fight the pandemic more effectively in the future countries like Vietnam, South Korea and Singapore have all experienced previous outbreaks of Middle East respiratory syndrome (MERS) or severe acute respiratory syndrome (SARS) epidemics in the past and have displayed remarkably effective responses to COVID-19, with better control of viral spread and significantly fewer confirmed cases and deaths attributed to COVID-19[1,2].

1. **Solution:-**

Coronavirus data visualization's major goal is to communicate information simply and efficiently using various graphical representations. Because it has various conceivable functions in the domain, visualization is a helpful medium for examining, understanding, and transmitting the information. Python is considered one of the best programming languages for handling data visualization because of its vast and active scientific computing community, as well as its numerous libraries that provide greater flexibility. It may also manage the specific parts of the graphs that are formed and make those requirements code reproducible. Python is also excellent at dealing with data and can manage massive volumes of data without crashing. For data visualization in python, we will use different libraries such as Matplotlib which is one of the most widely-used Python libraries for data visualization, and Ploty which is a declarative JavaScript data visualization toolkit that supports a wide range of statistical, scientific, financial, geographic, and 3-dimensional visualizations and is based on D3 and WebGL, Seaborn which helps users to make their visualizations more appealing, as well as meet some of the most typical data visualization requirements (like mapping color to a variable or using faceting). Working with Pandas DataFrames is easier than using Seaborn.

And we will provide the Data Analysis visualization of a real-time dataset which is updated every day by the Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE). This project represents the data visualization related to worldwide cover cases and how it is increasing. It also represents area plots of these cases using heat maps and static maps. It will also represent new cases, death per hundred cases, confirm cases, recovered cases, and active cases using Bar and Line Plot. The plot also represents how these cases are increasing on a day-by-day basis. The growth rate is represented per hundred cases, a thousand cases, ten thousand cases, and a hundred thousand cases. It also represents the time-lapse of the first and last cases.

1. **References:-**

[1] Seven things countries have done right in the fight against COVID-19. GAVI, The Vaccine Alliance website; Published June 22, 2020. <https://www.gavi.org/vaccineswork/seven-things-countries-have-done-right-fight-against-covid-19>. Accessed 27 May 2022.

[2] Vietnam: coronavirus pandemic country profile. Our World in Data website; 2020. <https://ourworldindata.org/coronavirus/country/vietnam?country=~VNM>. Accessed 27 May 2022.

[3] <https://www.medrxiv.org/content/10.1101/2020.06.25.20140004v2>

[4] <https://www.sciencedirect.com/science/article/pii/S2352340920310726>

[5] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7298499/>

[6] <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0249037>

[7] <https://diagnprognres.biomedcentral.com/articles/10.1186/s41512-022-00120-2>

[8] <https://www.frontiersin.org/articles/10.3389/frai.2021.579931/full>

[9] <https://www.nature.com/articles/s41746-020-00372-6>

[10] <https://github.com/topics/machine-learning-projects>

[11] <https://iopscience.iop.org/article/10.1088/1742-6596/1797/1/012011/pdf>

[12] <https://link.springer.com/article/10.1007/s40031-021-00546-0>

[13] <https://ieeexplore.ieee.org/document/9225621>

[14] <https://www.thelancet.com/journals/eclinm/article/PIIS2589-5370(21)00299-6/fulltext>

[15] <https://www.tandfonline.com/doi/abs/10.1080/09720510.2020.1840076>

[16] <https://www.kaggle.com/datasets/allen-institute-for-ai/CORD-19-research-challenge>