DATA VISUALIZATION PROJECT REPORT

on

(COVID ANALYSIS DASHBOARD)

Submitted by

Sakshi

Program Name: MBA(Business Analytics)

TABLE OF CONTENTS

INTRODUCTION

In this course of Introduction to Data Visualization which is part of Data Analysis where we do manipulation of data and make meaningful data from the raw dataset.

Data Analysis is the process of systematically applying statistical and/or logical techniques to describe and illustrate, condense, and recap, and evaluate data.

This project report consists of raw data on Covid. The dataset is in raw form, and I have used my Tableau skill to make it meaningful in a good manner.

In this project I have taken five main problem statements for the analysis which gives meaning to the dataset so that anyone can analyze what short of problems going through world in covid scenario.

This Project, I have worked on Covid Dataset in which I have used different Charts and Maps .

OBJECTIVES/SCOPE OF THE ANALYSIS

The Six main objectives that I've worked on are as follows:

1. Comparison between Total Cases and Total Death happened country wise.
2. Comparison between Total Cases and Total Tests happened country wise.
3. Comparison between Population and Total Cases happened country wise.
4. Comparison between Serious/Critical and Total Cases happened country wise.
5. New Death happened after Total Tests are done country wise.
6. Red Zone area defined by rise of cases >= 70Lakh.

SOURCE OF DATASET

Kaggle is an online community platform for data scientists and machine learning enthusiasts. Kaggle allows users to collaborate with other users, find and publish datasets, use GPU integrated notebooks, and compete with other data scientists to solve data science challenges.

The term data set refers to a file that contains one or more records. The record is the basic unit of information used by a program running on OS. Any named group of records is called a data set.

Dataset Link: https://www.kaggle.com/datasets/imdevskp/corona-virus-report

ETL PROCESS

The 5 steps of the ETL process are: extract, clean, transform, load, and analyze. Of the 5, extract, transform, and load are the most important process steps.

- Extract: Retrieves raw data from an unstructured data pool and migrates it into a temporary, staging data repository. For this I have used kaggle.com, many people over there are posting unlimited number of datasets. We can get any raw dataset from anywhere.

- Clean: Cleans data extracted from an unstructured data pool, ensuring the quality of the data prior to transformation. For that I have used power pivot to clean the data which I got from kaggle.com that is Student performance in exam. By the term cleaning I means checking the Duplicate, Blank space, and unorganized form of data. Cleaning of dataset is must for every analyst because after cleaning data will be easier to manipulate.

- Transform: Structures and converts the data to match the correct target source. Its easy to transform data when its cleaned properly. It shows each field which is going to load is proper or not.

- Load: Loads the structured data into a data warehouse so it can be properly analyzed and used. (For transform and load data I use one menu of power pivot i.e., Manage.)

- Analyze: Big data analysis is processed within the warehouse, enabling the business to gain insight from the correctly configured data. For that I use pivot table, power pivot and power query editor.

ANALYSIS ON DATASET

- **General Description**

In this project I have basically analysed that,

1. Comparison between Total Cases and Total Death happened country wise.
2. Comparison between Total Cases and Total Tests happened country wise.
3. Comparison between Population and Total Cases happened country wise.
4. Comparison between Serious/Critical and Total Cases happened country wise.
5. New Death happened after Total Tests are done country wise.
6. Red Zone area defined by rise of cases >= 70Lakh.

- **Analysis Results**

By visualizing all the data, I can totally say that the virus is made by China. It is totally a political matter as India and USA are making their internal and external bonds good with each other and with other countries as well. So, China spread this virus mostly in USA and India, these two countries are most affected and Brazil loses more people. As all the data and numbers are given in the pictures.

.

- **Visualization**
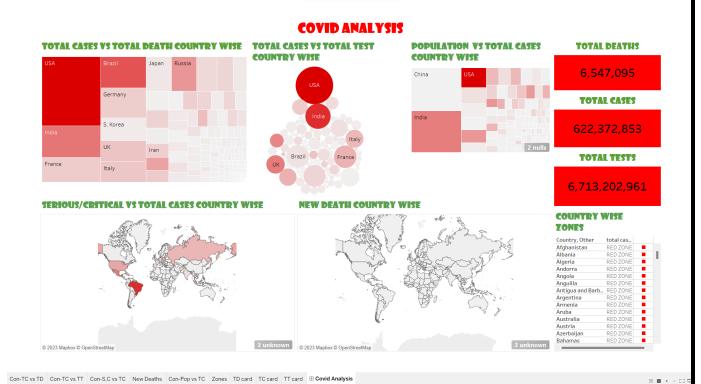
Fig. 5.1 Dashboard

LIST OF ANALYSIS WITH RESULTS

1. Comparison between Total Cases and Total Death happened country wise.

   ANALYSIS- Here we can see by the tree maps that the most affected country is USA.
   Total Cases= 98,166,904
   Death=1,084,282

Fig 5.2 Sheet1 (obj-1)

2. Comparison between Total Cases and Total Tests happened country wise.

ANALYSIS- Here we can see that USA & India both have highest no. in rise of Cases and highest no. of Tests are done.
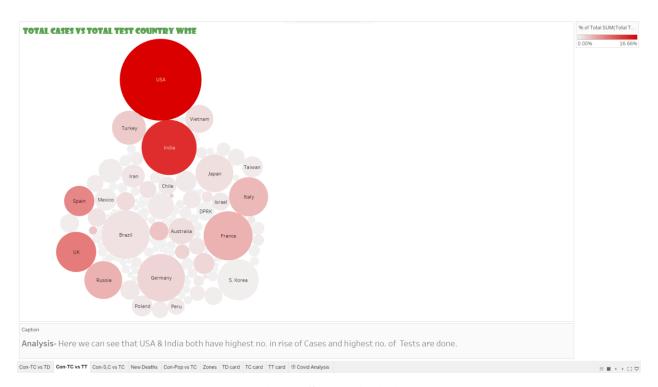
Fig 5.3 Sheet2 (obj-2)

3. Comparison between Serious/Critical and Total Cases happened country wise.

ANALYSIS- Here we can see that Brazil is having most no. Serious/Critical cases.

Fig 5.4 Sheet3 (obj-3)

4. New Death happened after Total Tests are done country wise.

ANALYSIS- Here we can see that Brazil is having most no. Serious/Critical cases.

Fig 5.5 Sheet4 (obj-4)

5. Comparison between Population and Total Cases happened country wise.

Analysis- Here we can see that after the formation of good bond between India and US, China the overpopulated Country is having Less Cases than India & US

Fig 5.6 Sheet5 (obj-5)
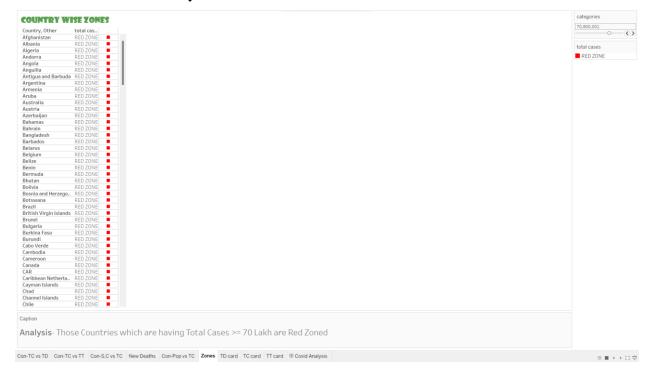
6. Red Zone area defined by rise of cases >= 70Lakh.

COUNTRY WISE ZONES

Country, Other	total cas..	
Afghanistan	RED ZONE	■
Albania	RED ZONE	■
Algeria	RED ZONE	■
Andorra	RED ZONE	■
Angola	RED ZONE	■
Anguilla	RED ZONE	■
Antigua and Barbuda	RED ZONE	■
Argentina	RED ZONE	■
Armenia	RED ZONE	■
Aruba	RED ZONE	■
Australia	RED ZONE	■
Austria	RED ZONE	■
Azerbaijan	RED ZONE	■
Bahamas	RED ZONE	■
Bahrain	RED ZONE	■
Bangladesh	RED ZONE	■
Barbados	RED ZONE	■
Belarus	RED ZONE	■
Belgium	RED ZONE	■
Belize	RED ZONE	■
Benin	RED ZONE	■
Bermuda	RED ZONE	■
Bhutan	RED ZONE	■
Bolivia	RED ZONE	■
Bosnia and Herzego..	RED ZONE	■
Botswana	RED ZONE	■
Brazil	RED ZONE	■
British Virgin Islands	RED ZONE	■
Brunei	RED ZONE	■
Bulgaria	RED ZONE	■
Burkina Faso	RED ZONE	■
Burundi	RED ZONE	■
Cabo Verde	RED ZONE	■
Cambodia	RED ZONE	■
Cameroon	RED ZONE	■
Canada	RED ZONE	■
CAR	RED ZONE	■
Caribbean Netherla..	RED ZONE	■
Cayman Islands	RED ZONE	■
Chad	RED ZONE	■
Channel Islands	RED ZONE	■
Chile	RED ZONE	■

categories

70,800,001

total cases

■ RED ZONE

Caption

Analysis- Those Countries which are having Total Cases >= 70 Lakh are Red Zoned

Con-TC vs TD   Con-TC vs TT   Con-S,C vs TC   New Deaths   Con-Pop vs TC   **Zones**   TD card   TC card   TT card   ⊞ Covid Analysis

Fig 5.7 Sheet6 (obj-6)

REFERENCES

I was able to get all the data I required to create the dashboard, some help was taken from various Websites and YouTube for design inspiration for my project. I came up with my own ideas and tried several methods on my own to complete the project. However, we did use certain concepts from www.javatpoint.com, www.github.com, and www.geeksforgeeks.org, for specific dashboard components.

Geeks for geeks: https://www.geeksforgeeks.org/

Course Hero: https://www.coursehero.com/

Kaggle: https://www.kaggle.com/

_____THE END_____