

# Project Report

## Introduction

This is the Data Science BootCamp project, which is an analysis of a data set using python and its libraries, Mathplotlib, Numby, and pandas. Here, I used a data set on **Covid-19** that includes cases and deaths from the beginning of the Corona Virus outbreak.

In this project, what we want is to analyze the data and figure out the number of cases and deaths over the last days and months and the time series since the onset of the virus outbreak by answering questions such as:

- What is the daily rate of infections and deaths worldwide since the outbreak of the virus?
- What is the average daily number of cases and deaths by country?
- What is the total monthly cases and deaths in the world?
- what is the total number of cases and deaths per month according to each country?
- Who are the top 5 countries in the number of deaths and cases since the beginning of the virus outbreak?
- What is the time series for the number of deaths and cases?

## Data Description :

dateRep	day	month	year	cases	deaths	countriesAndTerritories	geold	countryterritoryCode	popData2019	continentExp	Cumulative_number_for_14_days_o 19_cases_pe
12/14/2020	14	12	2020	746	6	Afghanistan	AF	AFG	38041757.0	Asia	
12/13/2020	13	12	2020	298	9	Afghanistan	AF	AFG	38041757.0	Asia	
12/12/2020	12	12	2020	113	11	Afghanistan	AF	AFG	38041757.0	Asia	
12/11/2020	11	12	2020	63	10	Afghanistan	AF	AFG	38041757.0	Asia	
12/10/2020	10	12	2020	202	16	Afghanistan	AF	AFG	38041757.0	Asia	
...	...	...	...	...	...	...	...	...	...	...	...
3/25/2020	25	3	2020	0	0	Zimbabwe	ZW	ZWE	14645473.0	Africa	
3/24/2020	24	3	2020	0	1	Zimbabwe	ZW	ZWE	14645473.0	Africa	
3/23/2020	23	3	2020	0	0	Zimbabwe	ZW	ZWE	14645473.0	Africa	
3/22/2020	22	3	2020	1	0	Zimbabwe	ZW	ZWE	14645473.0	Africa	
3/21/2020	21	3	2020	1	0	Zimbabwe	ZW	ZWE	14645473.0	Africa	

## Tools :

- Python in Jupyter notebook
- Numpy and Pandas for data manipulation
- Matplotlib, Seaborn and plotly for plotting

## **Data Pre-processing**

- Check if there are any missing values
- Check if there are any duplicate values
- Columns format
- Show the unique values

## **Data work & Result**

It is shown in the jupyter file on the qithub

## **Conclusion**

The above results have shown us the records of cases and deaths by country and chronology since the onset of the virus outbreak. The data set will be used to predict Corona's future impacts on working life, business and so forth.