# INDEX

| S.NO | SECTION | DESCRIPTION |
|------|---------|-------------|
| 1 | Introduction to the Project | Overview of the project and its objectives |
| 2 | Introduction | A detailed explanation of the project's purpose |
| 3 | Step-by-Step Implementation | Detailed breakdown of the project implementation |
| 4 | Conclusion | Summary of the findings and insights |
| 5 | Appendix: Full Code | Complete code used in the project |
| 6 | Certificate | The certification of completion |

# COVID-19 Data Analysis Using Python

## Introduction

This project focuses on analyzing the global spread of COVID-19 using Python. The datasets used in this analysis include the COVID-19 confirmed cases data and the World Happiness Report. The project explores trends in infection rates and examines the relationship between the pandemic's spread and economic and happiness indicators across countries.

---

## Step-by-Step Implementation

### 1. Importing Necessary Modules

To conduct the analysis, we import key Python libraries such as Pandas, NumPy, Seaborn, and Matplotlib, which help in data manipulation, analysis, and visualization.

---

### 2. Loading and Preparing the COVID-19 Dataset

#### 2.1 Importing the Dataset

We begin by loading the COVID-19 confirmed cases dataset, which contains the number of confirmed cases per country and region.

#### 2.2 Removing Unnecessary Columns

To clean the dataset, we remove columns that are not essential for the analysis, such as latitude and longitude.

#### 2.3 Aggregating Data by Country

We aggregate the confirmed cases by country, combining regional data to get a national-level summary for each country.

#### 2.4 Visualizing Data for Specific Countries

Data visualization helps to understand the trends of confirmed cases in different countries. We plot the cumulative cases over time for countries like China, Italy, India, and Spain.

---

### 3. Calculating Maximum Infection Rate

#### 3.1 First Derivative of the Curve

To measure the spread of the virus, we calculate the first derivative of the confirmed cases curve, which gives us the daily new cases or infection rate.

## 3.2 Maximum Infection Rate for Specific Countries

We then calculate the maximum infection rate for specific countries by finding the peak of the infection rate curve.

## 3.3 Maximum Infection Rate for All Countries

The maximum infection rate is computed for all countries, and we store these values in a new column of the dataset.

## 3.4 Creating a New DataFrame

We create a new DataFrame containing only the maximum infection rates of each country for further analysis.

---

## 4. Analyzing the World Happiness Report

## 4.1 Importing the Dataset

We import the World Happiness Report, which contains various happiness and economic indicators for countries across the world.

## 4.2 Selecting Relevant Columns

To focus on the factors that might impact the spread of COVID-19, we drop irrelevant columns from the dataset, such as overall happiness rank and generosity.

## 4.3 Setting the Country as the Index

We restructure the dataset by setting the country name as the index for easy merging with other datasets.

## 4.4 Merging the COVID-19 Data and Happiness Report

We merge the COVID-19 dataset (containing maximum infection rates) with the World Happiness Report based on the country name, creating a comprehensive dataset for analysis.

## 4.5 Correlation Matrix

We calculate the correlation matrix to identify relationships between the maximum infection rate and other variables, such as GDP per capita and social support.

---

## Conclusion

In this project, we explored the global spread of COVID-19 and examined its relationship with various socio-economic factors. The analysis demonstrated that countries with higher GDP per capita tend to show different infection rate patterns. Visualization played a key role in understanding the data and the trends associated with the pandemic.