

Project Final Report

Submitted by: Halimathu Sahadiya

Course: Data Science

Organization: Glowlogics Solutions

COVID-19 Data Analysis Using Python

1. Introduction

The COVID-19 pandemic has significantly impacted global health and economies worldwide. This project performs exploratory data analysis (EDA) on COVID-19 datasets to understand trends in confirmed cases, deaths, and recoveries across different countries over time.

2. Objective

The main objectives of this project are:

- To analyze global COVID-19 confirmed, death, and recovery cases
 - To identify the most affected countries
 - To study trends over time
 - To visualize data using graphs and charts
 - To draw meaningful insights from the dataset
-

3. Dataset Information

The dataset was obtained from Kaggle.

Dataset Name: **Corona Virus Report**

Files Used:

- covid_19_clean_complete.csv
- country_wise_latest.csv
- day_wise.csv

The dataset contains the following important columns:

- Date
 - Country/Region
 - Province/State
 - Confirmed
 - Deaths
 - Recovered
-

4. Tools and Technologies Used

- Python
 - Pandas (Data manipulation)
 - Matplotlib (Data visualization)
 - Seaborn (Statistical visualization)
 - Jupyter Notebook
-

5. Data Preprocessing

The following preprocessing steps were performed:

- Loaded dataset using Pandas
 - Converted Date column into datetime format
 - Checked for missing values
 - Grouped data by Date and Country
 - Aggregated global confirmed, deaths, and recovered cases
-

6. Exploratory Data Analysis

6.1 Global Trend Over Time

- Grouped data by Date
- Plotted global confirmed cases over time
- Observed steady increase during peak pandemic period

6.2 Top 10 Most Affected Countries

- Filtered latest available date
- Sorted countries by highest confirmed cases

- Visualized using bar chart

Observation:

Some countries reported significantly higher cases compared to others.

6.3 Death vs Confirmed Analysis

- Compared confirmed cases and deaths
- Observed strong positive correlation

6.4 Correlation Analysis

Correlation heatmap shows:

- Strong correlation between Confirmed and Deaths
 - Strong correlation between Confirmed and Recovered
-

7. Key Insights

- COVID-19 cases increased rapidly during the initial outbreak phase
 - Certain countries experienced significantly higher impact
 - Deaths are directly proportional to confirmed cases
 - Recovery rate improved over time in many regions
-

8. Conclusion

This project successfully analyzed global COVID-19 data using Python. Through data visualization and statistical analysis, meaningful insights were derived regarding the spread and impact of the pandemic.

The analysis demonstrates how data science techniques can be applied to real-world healthcare datasets to understand trends and patterns.

9. Future Scope

- Predict future cases using Machine Learning
- Perform country-specific deep analysis
- Analyze vaccination impact
- Build an interactive dashboard