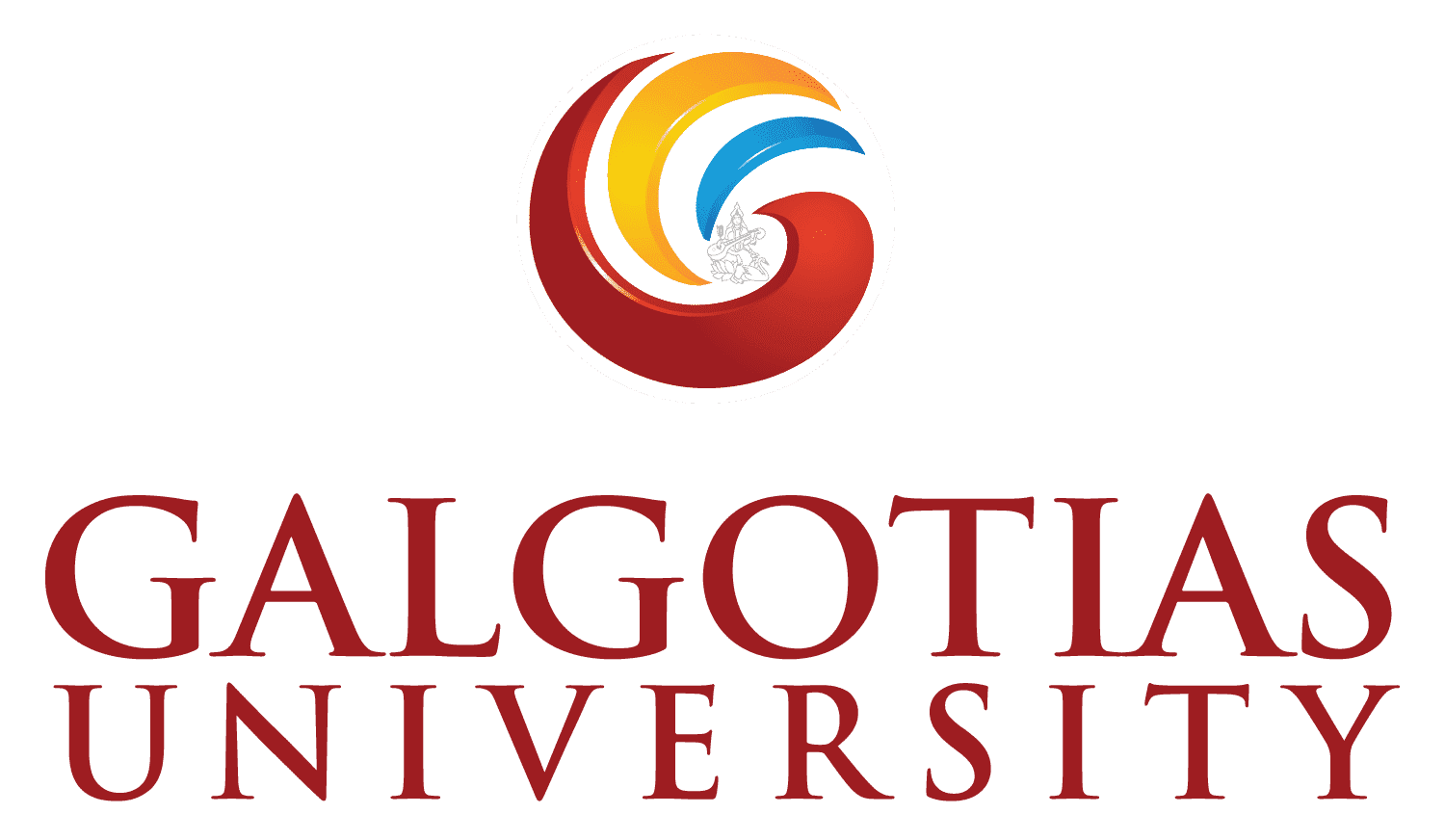
**COVID-19 Data Analysis**

**A Report for the Self –project of Python**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **S.no** | **Enrollment Number** | **Admission Number** | **Student Name** | **Degree / Branch** | **Sem** |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

**Group No. 13**

**School of Computing Science and Engineering**

**Greater Noida, Uttar Pradesh**

**Winter 2023 – 2024**

Reviewer name & Signature:

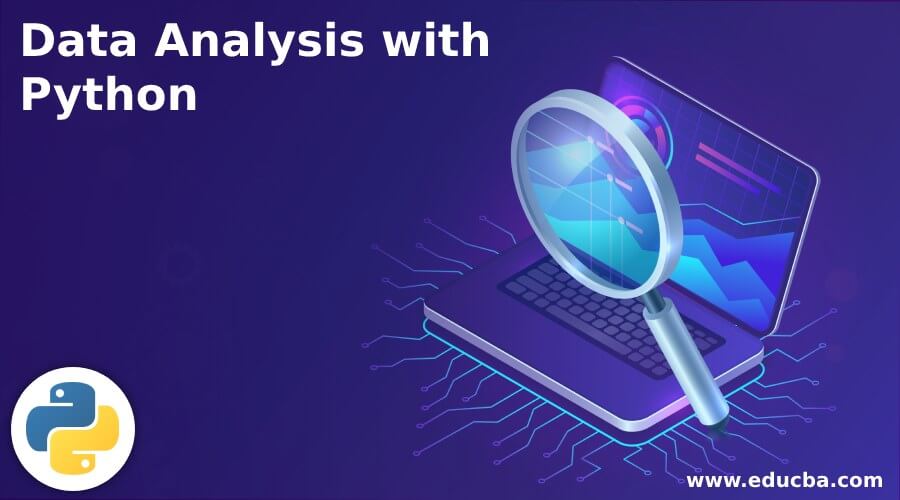
|  |  |  |
| --- | --- | --- |
| **S No.** | **Particulars** | **Page No** |
| **1** | **Title**  **COVID-19 Data Analysis** |  |
| **2** | **Architecture Diagram for Proposed method** |  |
| **3** | **Code for Implementation** |  |
| **4** | **Description of Project Modules** |  |
| **5** | **Output (Screenshot) -- Result** |  |
| **6** | **Conclusion** |  |
|  |  |  |
|  |  |  |
|  |  |  |

**TABLE OF CONTENTS**

**COVID-19 Data Analysis**

**Introduction**

**This project involves analysing a dataset containing COVID-19 statistics such as confirmed cases, deaths, and recoveries. The dataset is explored to extract meaningful insights and trends using Python's pandas, seaborn, and matplotlib libraries.**

****

**Architecture Diagram for Proposed method**

**Code for Implementation**

import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

# Load the dataset

data = pd.read\_csv(r"C:\Users\HP\Downloads\Covid\_19\_data.csv")

# Display the first few rows of the dataset

data.head()

# Check for missing values

data.isnull().sum()

# Visualize missing values

sns.heatmap(data.isnull(), cbar=False)

plt.title("Heatmap of Missing Values")

plt.show()

# Q1: Show the number of Confirmed, Deaths, and Recovered cases in each Region

region\_summary = data.groupby('Region').sum().head(20)

confirmed\_cases\_by\_region = data.groupby('Region')['Confirmed'].sum().sort\_values(ascending=False).head(10)

confirmed\_recovered\_by\_region = data.groupby('Region')[['Confirmed', 'Recovered']].sum()

# Q2: Remove all the records where Confirmed Cases are Less Than 10

data = data[~(data.Confirmed < 10)]

data.head(20)

# Q3: In which Region, maximum number of Confirmed cases were recorded?

max\_confirmed\_region = data.groupby('Region').Confirmed.sum().sort\_values(ascending=False).head(20)

# Q4: In which Region, minimum number of Deaths cases were recorded?

min\_deaths\_region = data.groupby('Region').Deaths.sum().sort\_values(ascending=True).head(50)

# Q5: How many Confirmed, Deaths & Recovered cases were reported from India till 29 April 2020?

india\_data = data[data.Region == 'India']

india\_confirmed = india\_data['Confirmed'].sum()

india\_deaths = india\_data['Deaths'].sum()

india\_recovered = india\_data['Recovered'].sum()

print(f'Confirmed: {india\_confirmed}, Deaths: {india\_deaths}, Recovered: {india\_recovered}')

# Q6: Sort the entire data wrt No. of Confirmed cases in ascending order

sorted\_data = data.sort\_values(by='Confirmed', ascending=True).head(50)

**Description of Project Modules**

**Data Loading:**

**This module involves loading the COVID-19 dataset into a pandas DataFrame. The dataset contains information on confirmed cases, deaths, and recoveries across various regions.**

**Data Cleaning:**

**This module checks for and visualizes missing values in the dataset using a heatmap. It ensures the data is clean and ready for analysis by removing records with less than 10 confirmed cases.**

**Data Analysis:**

**This module performs various analyses to answer specific questions:**

**Confirmed, Deaths, and Recovered Cases by Region: Summarizes the number of cases in each region.**

**Maximum Confirmed Cases by Region: Identifies the region with the highest number of confirmed cases.**

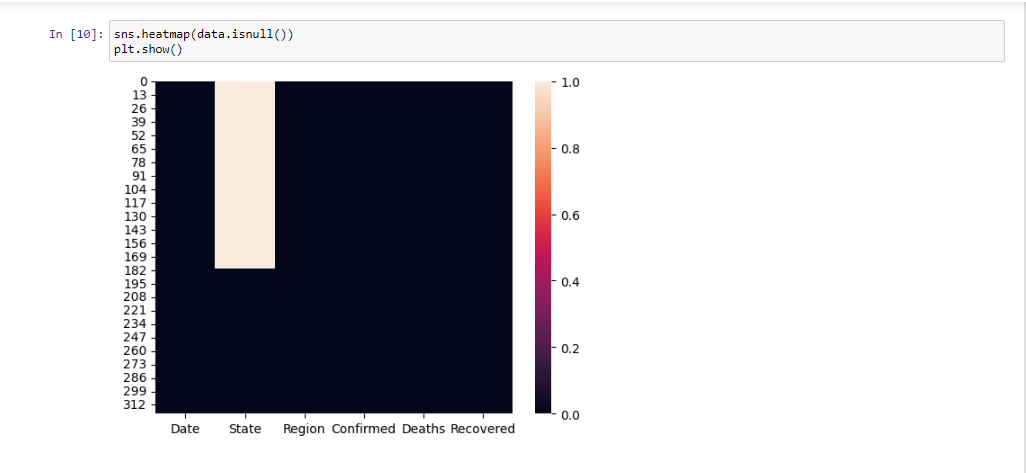
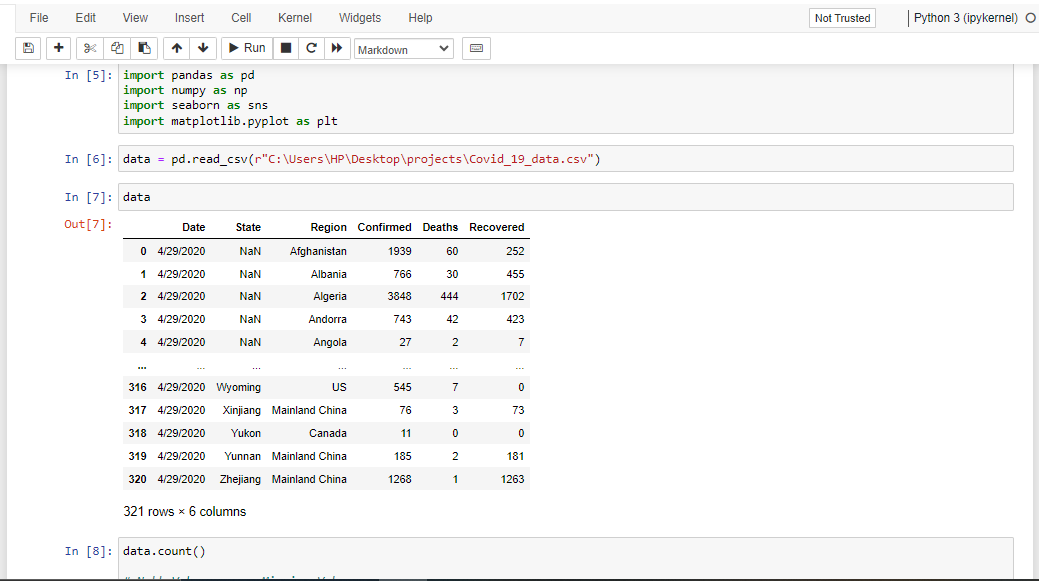
**Minimum Deaths by Region: Identifies the region with the lowest number of deaths.**

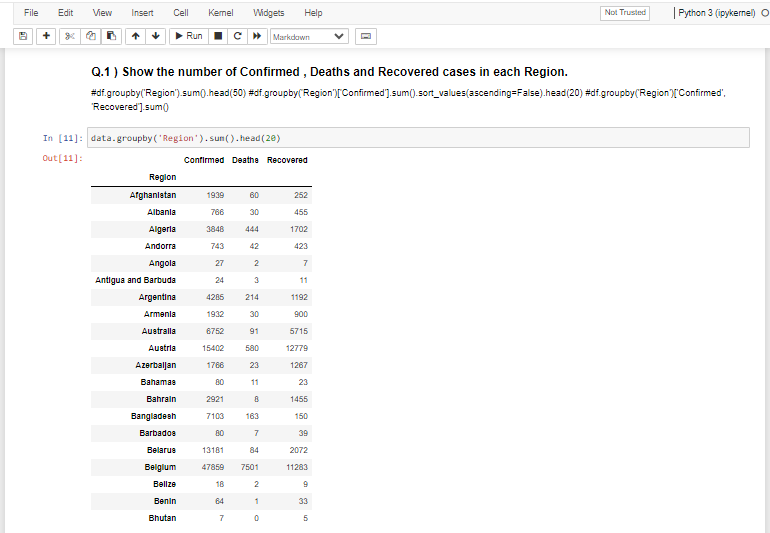
**COVID-19 Cases in India: Provides detailed statistics on COVID-19 cases in India as of April 29, 2020.**

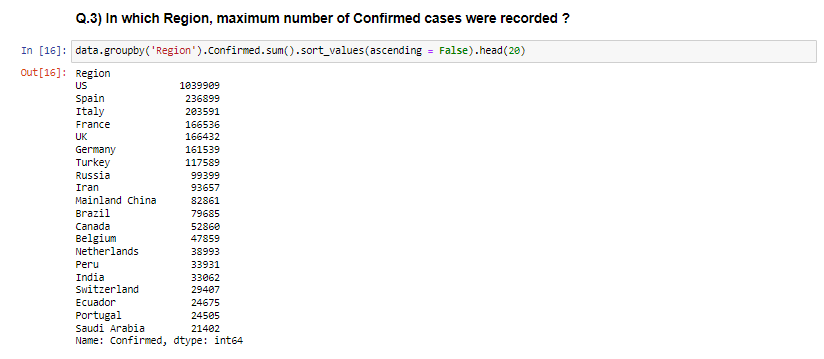
**Sorted Data: Sorts the dataset based on the number of confirmed cases in ascending order.**

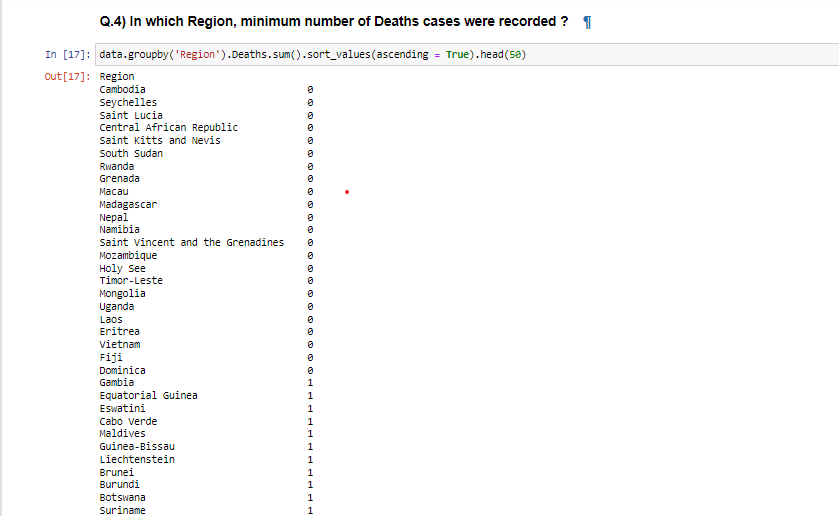
**Data Visualization**

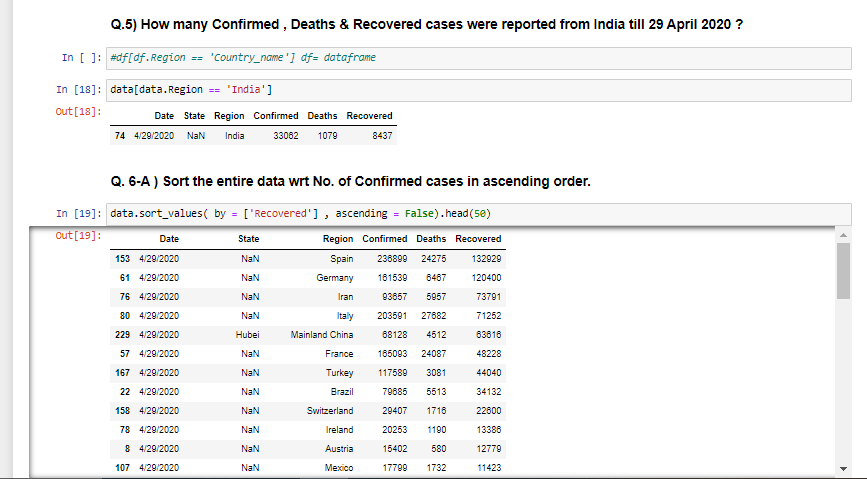
**This module creates visualizations using Seaborn and Matplotlib to aid in understanding the distribution and trends in the data.**

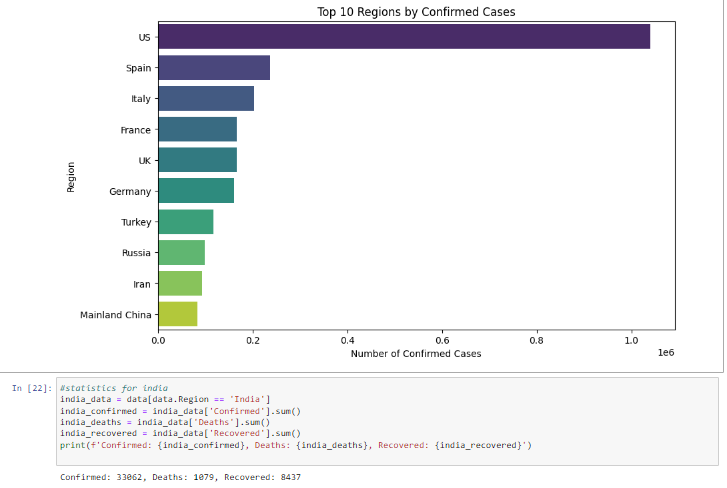
**Output**

****

****

****

****

****

**Conclusion:**

**The analysis provided several key insights into the COVID-19 pandemic:**

**The US had the highest number of confirmed cases.**

**Laos had the lowest number of deaths.**

**India had significant confirmed cases with notable recovery rates.**

**Recommendations**

**Focus efforts on regions with high confirmed cases.**

**Investigate regions with low death rates to replicate successful strategies.**

**Regularly update and monitor data for effective response planning.**

**Thank you**

**Github:** [**https://github.com/Bharatdungriyal/python\_Basic\_projects**](https://github.com/Bharatdungriyal/python_Basic_projects)