COVID-19 CASES ANALYSIS PROJECT DOCUMENTATION

PHASE-5

CONTENT

1.PROJECT OBJECTIVE

The primary objective of the COVID-19 cases analysis project was to explore and analyze the trends, patterns, and impacts of the COVID-19 pandemic using data-driven methods. This involved examining the progression of cases, identifying hotspots, understanding demographic impacts, and deriving insights that could potentially aid in better understanding and responding to the pandemic.

2. DESIGN THINKING PROCESS AND DEVELOPMENT PHASES:

Understanding the Problem: Initial phase involved gathering information on the spread of COVID-19, available datasets, and defining the scope of analysis. Data Collection and Preprocessing: Retrieval of diverse COVID-19 datasets from reliable sources (e.g., WHO, CDC) and preprocessing this data for analysis. Data Analysis and Visualization: Utilizing IBM Cognos and other data visualization tools to analyze the collected data, creating visual representations to comprehend the trends and patterns. Insights Generation: Deriving insights by comparing data across regions, demographics, and time periods.

3. ANALYSIS OBJECTIVES AND METHODOLOGY:

Data Collection Process: Multiple datasets were gathered, encompassing cases, deaths, recoveries, testing rates, demographics, and geographical information. These datasets were cleaned, integrated, and made ready for analysis. Data Visualization using IBM Cognos was employed to create visualizations, such as line graphs, heat maps, and dashboards, to represent the data comprehensively. Insights Generated: The analysis focused on understanding the trajectory of the pandemic in different regions, identifying the correlation between public health interventions and case rates, and examining demographic disparities in COVID-19 impacts.

4. INSIGHTS AND THEIR IMPLICATIONS:

Trend Analysis: Clear trends were observed in the rise and fall of cases over time, with spikes coinciding with certain events or policy changes. Geographical Hotspots: Identification of areas with high case rates helped in understanding the necessity for targeted interventions. Demographic Disparities: Analysis of the impact on different age groups, ethnicities, and socioeconomic statuses highlighted disparities in susceptibility and outcomes.

5. UNDERSTANDING COVID-19 TRENDS AND IMPACTS:

The insights derived from the analysis can be instrumental in several ways:

1. Public Health Interventions: Understanding what interventions have worked in controlling the spread.

2. Resource Allocation: Directing resources towards the most affected areas and vulnerable demographics.

3. Policy Decisions: Informing policy decisions to better respond to similar future health crises.

6. DATA COLLECTION PROCESS:

The data collection process for the COVID-19 analysis project involved the following steps:

Identification of Reliable Data Sources: The first step was to identify and gather data from credible sources such as the World Health Organization (WHO), Centers for Disease Control and Prevention (CDC), John Hopkins University, government health departments, and other reputable sources providing COVID-19 related datasets.
Data Retrieval: Various datasets were collected, covering a range of COVID-19 metrics including daily cases, deaths, recoveries, testing rates, hospitalizations, demographic information, and geographic data. These datasets were available in different formats such as CSV, JSON, or through APIs.
Data Cleaning and Preprocessing: The collected data required cleaning and preprocessing to ensure consistency, accuracy, and compatibility. This step involved dealing with missing values, standardizing date formats, harmonizing data fields, and ensuring uniformity across different sources.
Data Integration: The cleaned datasets were integrated to create a comprehensive dataset for analysis. This process involved merging multiple datasets based on common fields such as date, location, or unique identifiers to create a unified dataset suitable for analysis.
Data Quality Check: A quality check was conducted to verify the accuracy of the integrated dataset. This involved cross-validating the data, identifying anomalies, and ensuring data integrity.

7. VISUALIZATION USING IBM COGNOS:

The IBM Cognos tool was selected for its robust data visualization capabilities. The tool was configured and prepared for data integration and visualization.

IBM COGNOS SETUP:

The integrated dataset obtained from the data collection process was imported into IBM Cognos

DATA IMPORT:

Using IBM Cognos, various types of visualizations were created to represent the COVID-19 data comprehensively.

VISUALIZATION CREATION:

Showing trends in daily cases, deaths, and recoveries over time.

LINE GRAPHS:

IBM Cognos allowed for interactive analysis, enabling users to drill down into specific data points, filter information, and explore the data dynamically.

INTERACTIVE ANALYSIS:

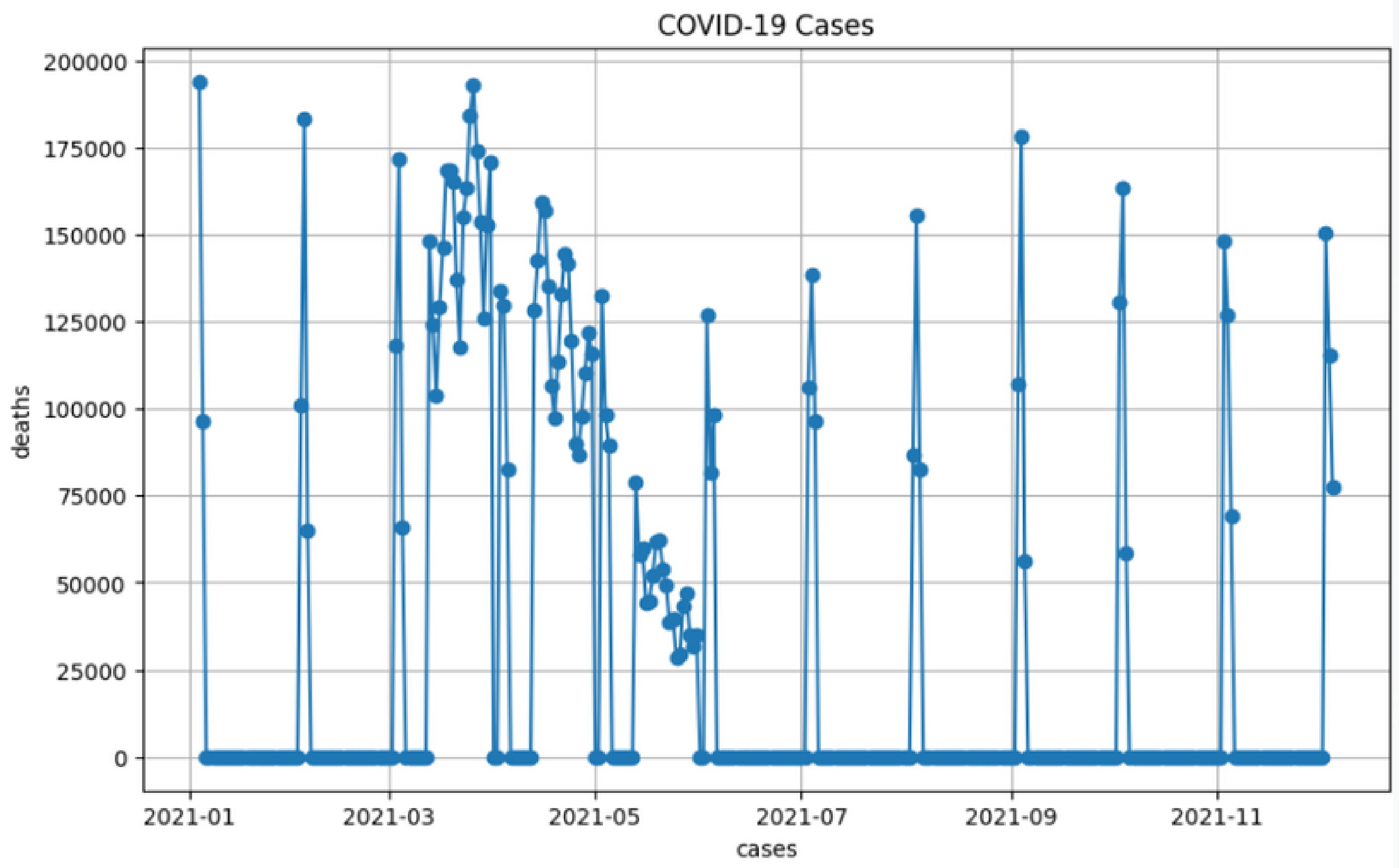
8.CONCLUSION

The project successfully delved into the vast array of COVID-19 data available and drew meaningful insights to aid in understanding the trends, impacts, and potential strategies for managing the pandemic.

CODING

```
import pandas as pd
import matplotlib.pyplot as plt

data = pd.read_csv('Covid_19_cases4.csv')
data['dateRep'] = pd.to_datetime(data['dateRep'])
data.set_index('dateRep', inplace=True)

daily_cases = data['cases'].resample('D').sum()
plt.figure(figsize=(10, 6))
plt.plot(daily_cases, marker='o', linestyle='-')
plt.title('COVID-19 Cases')
plt.xlabel('cases')
plt.ylabel('deaths')
plt.grid(True)
plt.show()
```

output

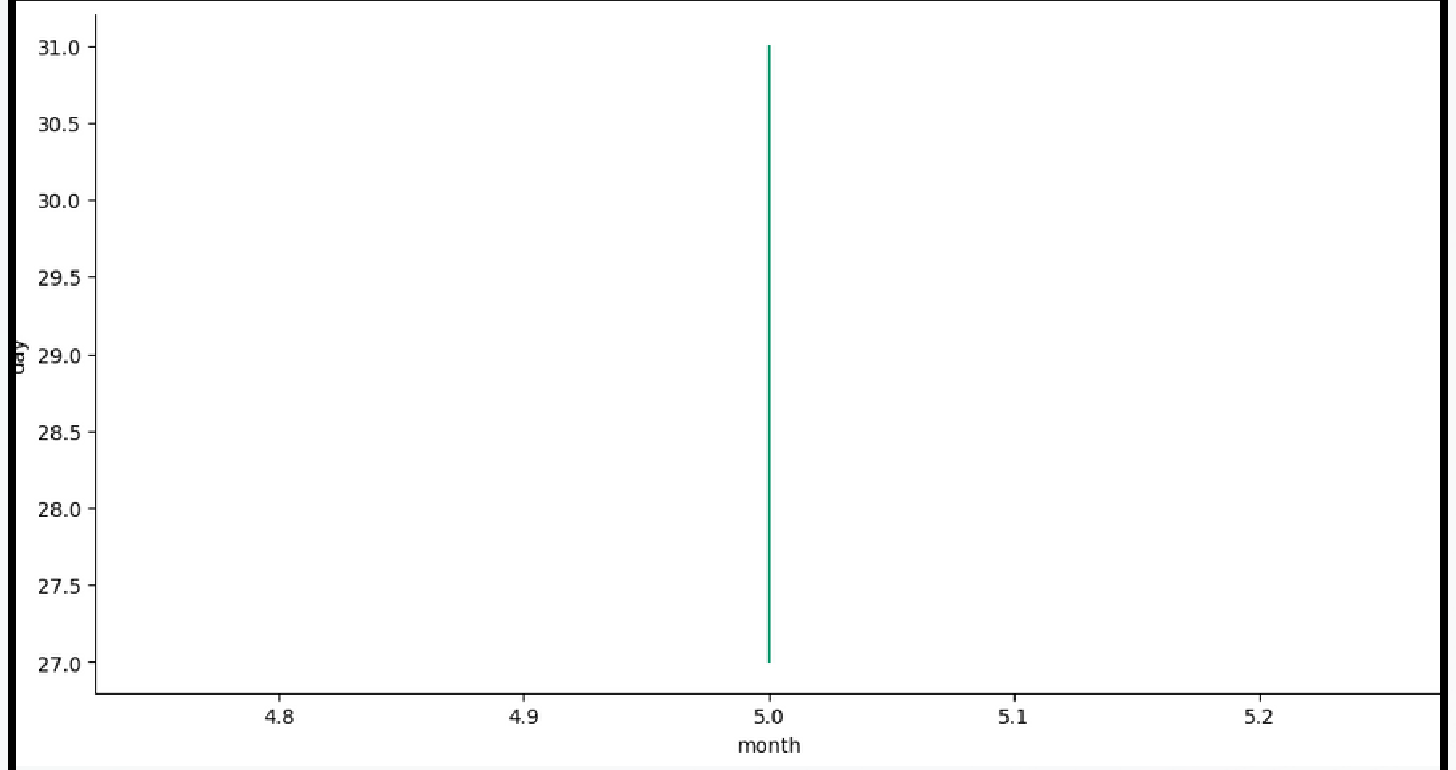


dataset:https://www.kaggle.com/datasets/chakradharmattapalli/covid-19-cases

CODING

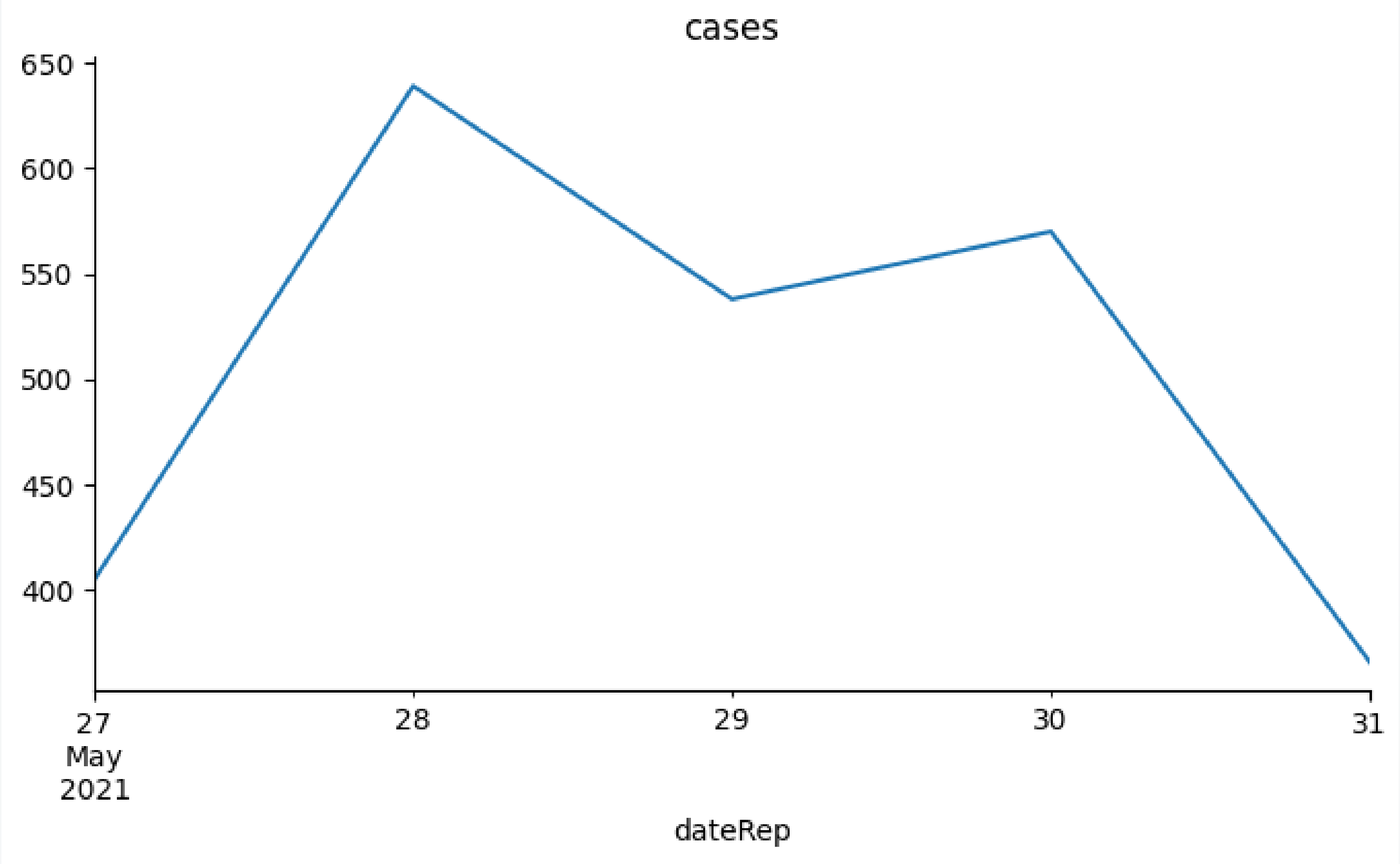
```
from matplotlib import pyplot as plt
import seaborn as sns
def _plot_series(series, series_name, series_index=0):
 from matplotlib import pyplot as plt
 import seaborn as sns
 palette = list(sns.palettes.mpl_palette('Dark2'))
 xs = series['month']
 ys = series['day']

 plt.plot(xs, ys, label=series_name, color=palette[series_index
% len(palette)])

fig, ax = plt.subplots(figsize=(10, 5.2), layout='constrained')
df_sorted = _df_5.sort_values('month', ascending=True)
_plot_series(df_sorted, '')
sns.despine(fig=fig, ax=ax)
plt.xlabel('month')
_ = plt.ylabel('day')
```

output



CODING

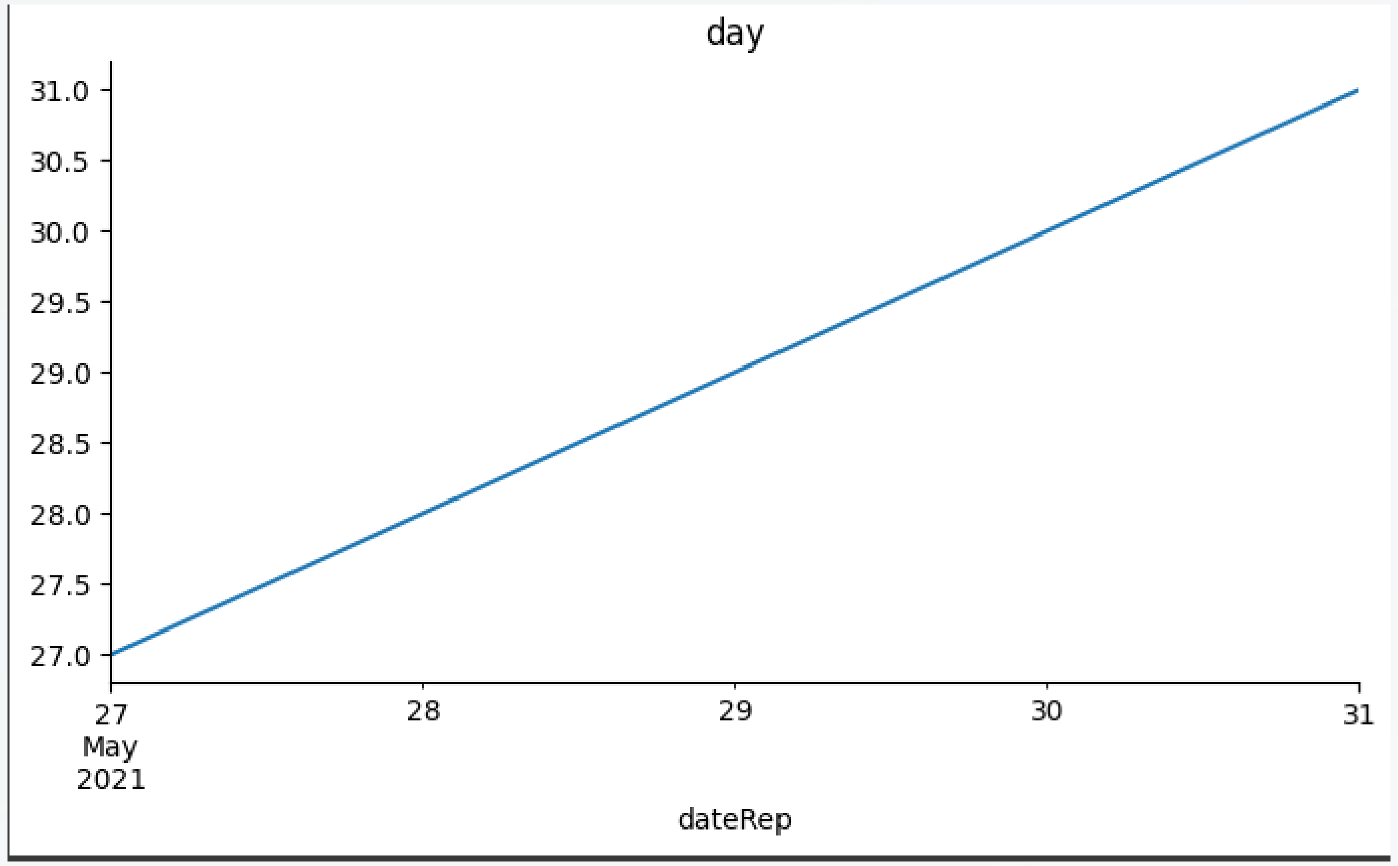
output




```
from matplotlib import pyplot as plt
import seaborn as sns
_df_10['cases'].plot(kind='line', figsize=(8, 4), title='cases')
plt.gca().spines[['top', 'right']].set_visible(False)
```

CODING

output




```
from matplotlib import pyplot as plt
_df_9['day'].plot(kind='line', figsize=(8, 4), title='day')
plt.gca().spines[['top', 'right']].set_visible(False)
```

CODING

```
import pandas as pd
data = pd.read_csv('Covid_19_cases4.csv')
data['dateRep'] = pd.to_datetime(data['dateRep'])
cases_by_country = data.groupby('countriesAndTerritories')
['cases'].sum()
deaths_by_country = data.groupby('countriesAndTerritories')
['deaths'].sum()
total_cases = data['cases'].sum()
total_deaths = data['deaths'].sum()
total_countries = data['countriesAndTerritories'].nunique()
print("Total Cases by Country:")
print(cases_by_country)
print("\nTotal Deaths by Country:")
print(deaths_by_country)
print("\nOverall Summary Statistics:")
print(f"Total Cases: {total_cases}")
print(f"Total Deaths: {total_deaths}")
print(f"Total Countries: {total_countries}")
```

output

```
Total Cases by Country:
countriesAndTerritories
Austria          184416
Belgium          288119
Bulgaria         171236
Croatia          113168
Cyprus            37700
Czechia          421221
Denmark           69188
Estonia           62916
Finland           34760
France          2020808
Germany         1234058
Greece           210201
Hungary          371613
Iceland             527
Ireland           42057
Italy           1290738
Latvia            46912
Liechtenstein       437
Lithuania         77040
Luxembourg        14464
Malta              7586
Netherlands      557983
Norway            53995
...
Overall Summary Statistics:
Total Cases: 9994560
Total Deaths: 178247
Total Countries: 30
```