

**GOVERNMENT COLLEGE OF ENGINEERING BARGUR**

**( AUTONOMOUS)**

**DATA ANALYTICS WITH COGNOS**

**PROJECT TITLE: COVID – 19 CASES ANALYSIS**

**PHASE V FINAL SUBMISSION (PROJECT)**

**TEAM MEMBERS:**

**SHALOME A**

**BALAJI S**

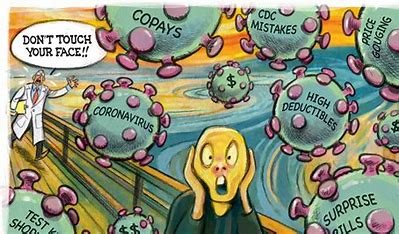
**KAMALAKANNAN V**

**THANGARAJ S**

**VARUN KUMAR M**

**PROBLEM STATEMENT:**

The project aims to analyze and visualize COVID-19 cases using real-time data to gain insights, inform the public, and contribute to the understanding of the pandemic.

****

**PROBLEM SOLUTION:**

**Project Components:**

1. Data Collection:

* Collect real-time COVID-19 data from reliable sources (e.g., World Health Organization, national health agencies, Johns Hopkins University COVID-19 Dashboard, etc.).
* Clean and pre-process the data for analysis.

1. Exploratory Data Analysis (EDA):

* Perform initial data exploration to understand the dataset's structure and characteristics.
* Calculate basic statistics, such as the total number of cases, deaths, recoveries, and active cases.
* Create visualizations (e.g., line graphs, bar charts, heat maps) to represent trends in cases, testing rates, and vaccination rates.

1. Geospatial Analysis:

* Plot COVID-19 cases on maps to visualize regional and global hotspots.
* Analyze the spread of the virus across different regions, countries, or states.

1. Time Series Analysis:

* Create time series plots to show the progression of COVID-19 cases over time.
* Identify key milestones, such as waves, peaks, and troughs in the pandemic.

1. Vaccine Analysis:

* Analyze vaccination data, including vaccine coverage rates and vaccine effectiveness.
* Compare vaccination rates with changes in case numbers.

1. Impact Analysis:

* Assess the impact of COVID-19 on various sectors (e.g., healthcare, economy, education, mental health).
* Discuss policy responses and their effectiveness.

1. Public Awareness:

* Create informative infographics, reports, or articles to share your findings with the public.
* Use social media platforms or a dedicated website to disseminate information and raise awareness.

1. Challenges and Ethical Considerations:

* Discuss the challenges faced during data collection and analysis.
* Address ethical considerations related to privacy, data accuracy, and responsible data sharing.

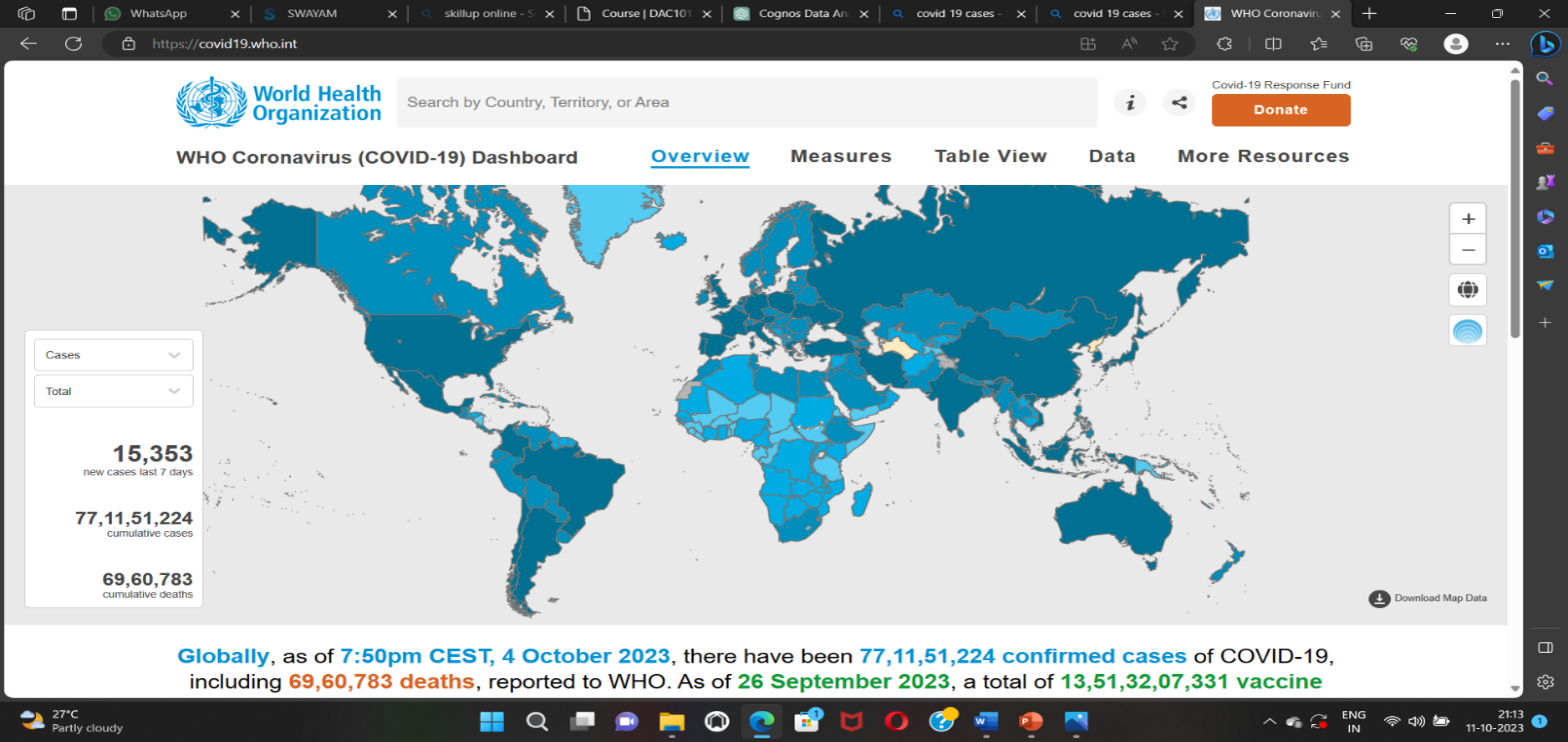


**KEYPOINTS AND ANALYSIS :**

COVID-19, caused by the novel coronavirus SARS-CoV-2, is a global pandemic that has affected countries around the world.

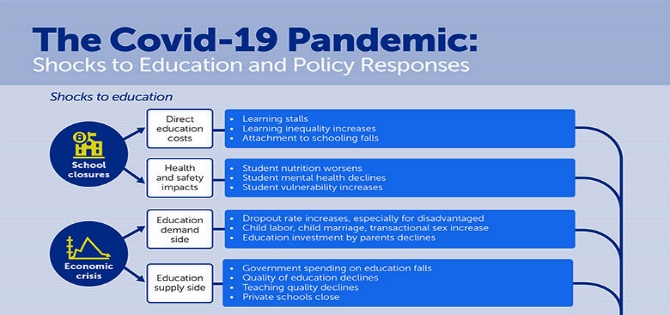
Key points related to covid-19 cases are:

1. Spread and Transmission
2. Symptoms
3. Testing
4. Case Tracking
5. Vaccination
6. Variants
7. Public Health Measures
8. Global Impact
9. Ongoing Research

* **Trend Analysis:** Analyze the data to identify trends, patterns, and hotspots. For example, you can analyze the growth rate of cases in different regions or countries.
* **Statistical Analysis:** Conduct statistical tests to derive insights, such as the correlation between testing rates and infection rates, or the impact of public health measures on the curve.
* **Predictive Modeling:** You can build predictive models to forecast the spread of the virus based on historical data and known variables. Time series analysis and machine learning can be used for this purpose.
* **Interactive Dashboards:** Create interactive dashboards that allow users to explore the data and gain insights. Tools like Tableau, Power BI, or custom web development with JavaScript can help in building these dashboards.
* **Public Communication:** Share your findings and visualizations through public channels like a website, social media, or data visualization platforms. This can help in informing the public and raising awareness.
* **Data Source Attribution:** Always provide proper attribution to the data sources used in your project.
* **Ethical Considerations:** Given the sensitivity of the topic, be mindful of ethical considerations, such as privacy and ensuring that your project's purpose is to inform and not cause panic.
* **Collaboration:** Consider collaborating with epidemiologists, public health experts, and data scientists who have expertise in infectious diseases for a more comprehensive analysis.
* **Documentation:** Document your data sources, methodologies, and findings comprehensively to make your project transparent and reproducible.
* **Feedback Mechanism:** Encourage user feedback and engagement to improve your project and address any concerns or questions from the public

**PROJECT OVERVIEW:**

This project involve a combination of data modeling, report building, and potentially custom scripting depending on your specific requirements. Below is a simplified example of the process. Note that this is a high-level overview, and actual code would depend on your data source and the specific analysis you wish to perform. Also, IBM Cognos uses its own scripting language called "Cognos Expressions" for report development.



**TOOLS AND TECHNOLOGIES:**

* Programming languages (e.g., Python) for data analysis and visualization.
* Data visualization libraries (e.g., Matplotlib, Seaborn, Plotly, Tableau).
* Geospatial analysis tools (e.g., GeoPandas, Leaflet).
* Machine learning libraries (e.g., scikit-learn) for predictive modeling (if applicable).
* Data sources (e.g., APIs, datasets from official health organizations).

**DATA ANALTICS OUTLINE:**

Data Connection Setup:

* In Cognos, we can set up a data connection to your COVID-19 dataset. We would define data source details such as the database or file location, connection credentials, and any required SQL queries to retrieve the data.

Data Modeling:

* In IBM Cognos, we can use IBM Framework Manager to create a data model. This involves defining the structure of our data, such as tables, relationships, calculations, and business rules. This is where we specify how COVID-19 data relates to other data sources if necessary.

Report Development:

* Cognos allows you to design reports and dashboards using a drag-and-drop interface. We can create various types of visualizations like tables, charts, and maps to present COVID-19 data. The code would mostly involve creating and formatting these reports.
* Example:

- Create a new report

- Drag the "Date" dimension to the Rows section

- Drag "New Cases" measure to the Columns section

- Apply necessary formatting and calculations

- Add titles, labels, and a date range filter

Data Analysis:

* Cognos provides various functions and expressions to perform calculations and aggregations on our data. For instance, we can calculate rolling averages, growth rates, or percentage changes in cases. The code for data analysis will depend on the specific calculations we want to perform.

Geospatial Analysis:

* If we wish to include geographic analysis, Cognos can handle geospatial data. We might write code to create a map visualization showing COVID-19 hotspots or trends in different regions.

Automation and Scheduling:

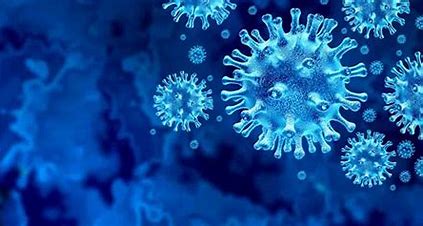
* Cognos allows us to schedule data refreshes and report distribution. We would set up schedules and configurations using the Cognos administration interface.

Custom Scripting:

* Depending on our requirements, we may need to write custom scripts. Cognos uses its own scripting language for expressions, but we can also integrate with external data sources and perform custom scripting in languages like JavaScript if needed.

Security and Compliance:

* In Cognos, we'd configure security settings and data access permissions to ensure that sensitive data is protected.
* Remember that creating a full COVID-19 analysis project in Cognos is a detailed process, and it's important to have a good understanding of Cognos tools and features, as well as a clear project plan to meet our objectives.



**PROCEDURE-COVID-19 ANALYSIS IN PYTHON:**

Step 1: Install Required Libraries

E.g.: pip install pandas matplotlib seaborn scikit-learn

Step 2: Prepare the Data

Obtain COVID-19 data in CSV format.

Save the data file (e.g., covid\_data.csv) in the same directory as your Python script

Step 3: Python Script

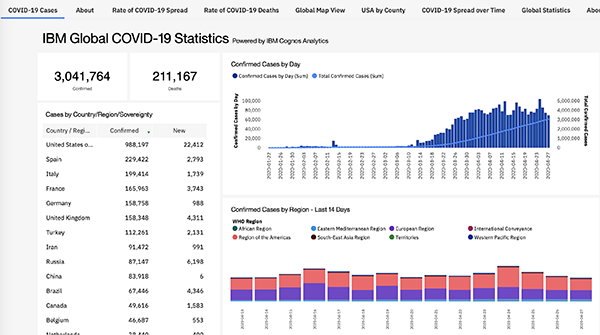
Use any code editor or integrated development environment (IDE) to create and run the Python script.

Step 4: Running the Script.Save and run the script

The script will load the data, perform analysis, and display the visualizations and results as specified in the code.

**PYTHON CODE:**

To perform COVID-19 data analysis in Python, we can use libraries such as pandas for data manipulation, matplotlib and seaborn for data visualization, and scikit-learn for machine learning.



**Covid cases analysis.py**

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

data = [

{

'dateRep': '31-05-2021',

'day': 31,

'month': 5,

'year': 2021,

'cases': 366,

'deaths': 5,

'countriesAndTerritories': 'Austria'

},

Covid cases analysis.py

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

data = [

{

'dateRep': '31-05-2021',

'day': 31,

'month': 5,

'year': 2021,

'cases': 366,

'deaths': 5,

'countriesAndTerritories': 'Austria'

},

{

'dateRep': '28-05-2021',

'day': 28,

'month': 5,

'year': 2021,

'cases': 639,

'deaths': 4,

'countriesAndTerritories': 'Austria'

},

{

'dateRep': '27-05-2021',

'day': 27,

'month': 5,

'year': 2021,

'cases': 405,

'deaths': 19,

'countriesAndTerritories': 'Austria'

},

{

'dateRep': '26-05-2021',

'day': 26,

'month': 5,

'year': 2021,

'cases': 287,

'deaths': 8,

'countriesAndTerritories': 'Austria'

},

{

'dateRep': '25-05-2021',

'day': 25,

'month': 5,

'year': 2021,

'cases': 342,

'deaths': 3,

'countriesAndTerritories': 'Austria'

}

# Convert the data into a DataFrame

df = pd.DataFrame(data)

plt.figure(figsize=(12, 6))

sns.lineplot(data=df, x='dateRep', y='cases')

plt.title('Daily Cases Over Time')

plt.xticks(rotation=45)

plt.show()

plt.figure(figsize=(12, 6))

sns.lineplot(data=df, x='dateRep', y='deaths')

plt.title('Daily Deaths Over Time')

plt.xticks(rotation=45)

plt.show()

plt.figure(figsize=(8, 6))

sns.scatterplot(data=df, x='cases', y='deaths')

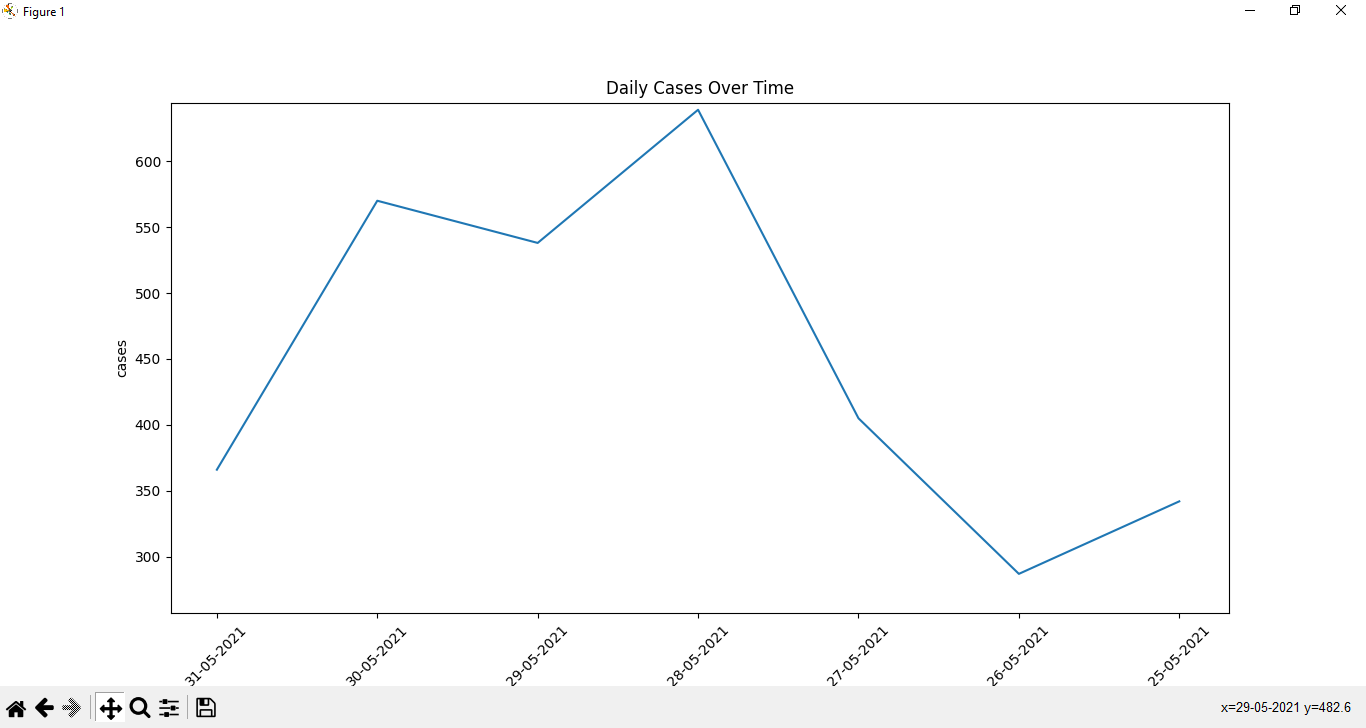
plt.title('Scatter Plot of Cases vs. Deaths')

plt.show()

correlation = df['cases'].corr(df['deaths'])

print(f'Correlation between cases and deaths: {correlation}')

**OUTPUT:**

****

**CONCLUSION:**

1. The coronavirus disease continues to spread across the world following a trajectory that is difficult to predict.
2. The data analysis for COVID-19 cases have been completed using a set of data given as input to which the respective output is generated and that is attached.
3. Hence, the COVID-19 cases analysis is done using the libraries in python and output is represented graphically.
4. Therefore the Data Analysis for the COVID-19 cases is done.

**REFERENCES**:

1. WHO Organization <https://www.who.int/health-topics>/
2. Worlds meters <https://www.worldometers.info/coronavirus/>
3. Wikipedia https://en.m.wikipedia.org/wiki/Contagious\_disease