## Phase 3: Development Part 1

**requirements technology:**

## Step 1: Data Loading and Preprocessing

1. Load the Dataset:
   * Open IBM Cognos and create a new project or workspace.
   * Navigate to the data section and import your COVID-19 dataset.
2. Initial Data Check:
   * Verify that the data includes necessary columns like date, location, cases, deaths, etc.
3. Data Cleaning:
   * Handle missing values, remove duplicates, validate data formats, and address any outliers.
4. Data Transformation:
   * Perform any necessary calculations or transformations (e.g., mortality rate, rolling averages).

### Step 2: Data Analysis and Visualization

1. Create Data Model:
   * Define relationships between different data elements if applicable.
2. Create Calculated Fields:
   * Add calculated fields for metrics you want to track (e.g., mortality rate, recovery rate).
3. Create Visualizations:
   * Go to the "Author" or "Report" section.
   * Select the appropriate visualization type (e.g., line chart, bar chart, map).
4. Drag and Drop Data Fields:
   * Drag relevant data fields (e.g., date, cases, deaths) to appropriate areas (e.g., x-axis, y-axis).
5. Apply Filters:
   * Apply filters to focus on specific regions, time periods, or other criteria.
6. Customize Visualizations:
   * Customize labels, colors, and other aspects of the visualization to enhance clarity.

### Step 3: Advanced Analysis (Optional)

1. Time-Series Data Handling:
   * Apply techniques like smoothing or rolling averages to highlight trends.
2. Correlation Analysis (if applicable):
   * Explore potential correlations with external factors (e.g., lockdown measures, vaccination rates).

### Step 4: Create Dashboard

1. Combine Visualizations:
   * Create a dashboard to bring multiple visualizations together for a comprehensive view.
2. Arrange and Format:
   * Arrange visualizations in a logical order and format the dashboard for clarity.

### Step 5: Interactivity and Drill-Downs (Optional)

1. Add Interactivity:
   * Enable interactions between different visualizations for dynamic exploration.
2. Set Up Drill-Downs:
   * Define drill-down paths for more detailed views of the data.

### Step 6: Test and Validate

1. Review and Test:
   * Review the visualizations and test interactivity to ensure everything works as expected.
2. Validate Against Objectives:
   * Confirm that the analysis meets the defined objectives.

### Step 7: Document the Project

1. Create Documentation:
   * Develop a document that outlines the project's goals, technologies used, data preprocessing steps, analysis methods, and visualization outcomes.
2. Include Screenshots:
   * Capture screenshots of relevant visualizations and dashboard views for visual reference.
3. Write a Summary and Conclusion:
   * Provide a summary of the project, highlighting key insights, observations, and any recommendations.
4. Share for Assessment:
   * Share the document with the relevant stakeholders or assessors for evaluation.

Remember to follow best practices for documentation, including clear and concise explanations, organized formatting, and proper citations for any external resources or tools used.

## Objectives:

1. Trend Analysis: Track and visualize the trends of COVID-19 cases and deaths over time to identify peaks, troughs, and potential patterns.
2. Geographical Comparison: Compare the data across different regions or countries to understand how the pandemic has affected different areas.
3. Mortality Rate Calculation: Calculate and analyze the mortality rate to assess the severity of the outbreak in different locations.
4. Vaccination Impact Assessment (if data available): If vaccination data is available, analyze its impact on reducing the number of cases and deaths.
5. Demographic Analysis (if data available): Explore how different age groups, genders, or ethnicities have been affected by the virus.
6. Healthcare System Strain: Assess the strain on healthcare systems by examining hospitalization rates and availability of critical resources.
7. Epidemiological Modeling (if expertise available): If expertise in epidemiology is available, apply modeling techniques to predict future trends based on current data.
8. Public Health Interventions Assessment: Analyze the effectiveness of public health measures such as social distancing, mask mandates, and lockdowns.

## Obtaining COVID-19 Data:

1. Official Health Organizations: Obtain data from trusted sources like the World Health Organization (WHO), Centers for Disease Control and Prevention (CDC), or national health agencies.
2. Research Institutions: Data repositories from institutions like Johns Hopkins University or academic research centers often provide reliable COVID-19 datasets.
3. Government Open Data Portals: Many governments provide open access to COVID-19 data through dedicated portals.
4. APIs: Some organizations provide APIs for accessing real-time COVID-19 data. Examples include the COVID Data API by The New York Times.
5. Data Aggregator Websites: Websites like Our World in Data aggregate and provide access to COVID-19 data from various sources.
6. Data Science Platforms: Data science platforms like Kaggle often host COVID-19 datasets contributed by the community.

Ensure that the data you obtain is well-documented, up-to-date, and comes from reputable sources. It's important to adhere to ethical guidelines and privacy regulations when handling health-related data. Always cite your sources and consider any licensing or usage restrictions associated with the data.

**Program for preprocessing:**

import pandas as pd

df = pd.read\_csv("Covid\_19\_cases4.csv")

print(df.head())

df['dateRep'] = pd.to\_datetime(df['dateRep'], format='%d / %m / %Y')

df['date'] = df['day'].astype(str) + '/' + df['month'].astype(str) + '/' + df['year'].astype(str)

df = df.drop(['day', 'month', 'year'], axis=1)

print(df.head())

df.to\_csv("preprocessed\_covid\_data.csv", index=False)

Create Visualizations:

* + Go to the "Author" or "Report" section.
  + Select the appropriate visualization type (e.g., line chart, bar chart, map).

1. Drag and Drop Data Fields:
   * Drag relevant data fields (e.g., date, cases, deaths) to appropriate areas (e.g., x-axis, y-axis).
2. Apply Filters:
   * Apply filters to focus on specific regions, time periods, or other criteria.
3. Customize Visualizations:
   * Customize labels, colors, and other aspects of the visualization to enhance clarity.

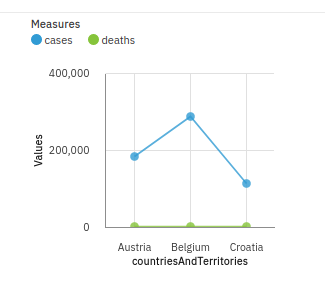
**Cleaning and processing:**

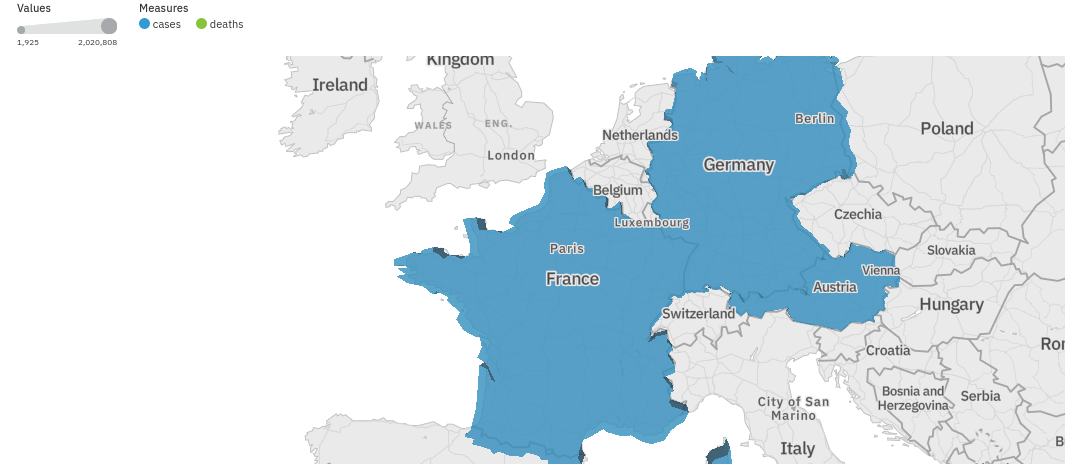
Handle Missing Values:

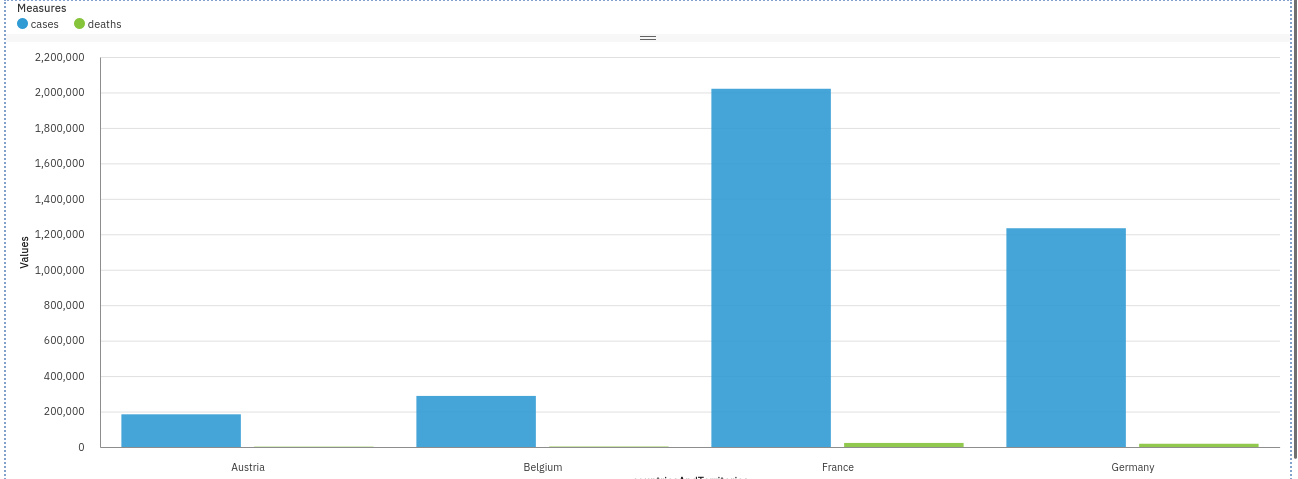
* + Identify and address missing values in the dataset. Depending on the nature of the missing data, you might choose to impute values, remove rows with missing values, or use other techniques.

1. Remove Duplicates:
   * Check for and remove any duplicate records in the dataset. Duplicates can skew your analysis.
2. Data Validation:
   * Ensure that data is in the correct format. For example, dates should be in the right format, and numeric fields should contain only numbers.
3. Outlier Detection and Handling:
   * Identify and handle outliers that might be erroneous or skew your analysis. This could involve smoothing techniques or more advanced statistical methods.
4. Standardize Data Units:
   * Ensure that all data points are in the same units. For example, if there are multiple units for measurements, convert them to a consistent unit.
5. Normalize or Standardize Data (if applicable):
   * Depending on your analysis, you may need to normalize or standardize the data to bring it to a common scale.
6. Time-Series Data Handling:
   * If working with time-series data, ensure that it is organized chronologically. You may want to apply techniques like smoothing or calculate rolling averages to highlight trends.
7. Calculate Derived Metrics:
   * Derive additional metrics that might be relevant to your analysis. For example, calculate mortality rate, recovery rate, or any other specific metric you need.
8. Check for Data Integrity:
   * Ensure that there are no logical inconsistencies in the data. For example, verify that the number of cases does not exceed the population of a region.
9. Handle Categorical Data (if applicable):
   * If your data includes categorical variables, make sure they are properly encoded or transformed for analysis.
10. Data Imputation (if applicable):
    * If you've removed missing data, you might need to impute values for some variables. Be cautious in choosing the imputation method, as it can impact the results.
11. Document Data Cleaning Process:
    * Keep a detailed record of all the steps you've taken to clean and process the data. This documentation is crucial for transparency and reproducibility.

Remember, data cleaning is an iterative process, and you may need to revisit and revise these steps as you gain more insights from your analysis. Always ensure that you're working with high-quality, reliable data to produce accurate results.

****

****

****