Project : Covid-19 Cases Analysis

**Phase 3 : project development part 1**

# Data cleaning overview

## Step 1: Import necessary libraries

This code snippet imports the required libraries for data manipulation and analysis: Pandas for handling data, NumPy for numerical operations, and ‘ zscore ’ function from SciPy for calculating z-scores.

### import pandas as pd import numpy as np

**from scipy.stats import zscore**

## Step 2: Load the dataset

This line reads the CSV file containing COVID-19 data and stores it in a Pandas DataFrame called ‘ data ’.

**data = pd.read\_csv("/content/Covid\_19\_cases4.csv")**

## Step 3: Data Inspection

These lines print the first few rows of the dataset and provide information about the dataset, including the data types and any missing values.

**print(data.head())** # Print the first 5 rows of the dataset

**print(data.info())** # Print the summary of the DataFrame, including the data

types and missing values

## Step 4: Handling Missing Data

This line fills missing values in the dataset with the mean of each respective column. Missing values are replaced with the mean value of their respective columns.

**data.fillna(data.mean(), inplace=True)**

## Step 5: Data Validation and Cleaning

This line removes duplicate rows from the dataset, ensuring that each row is unique.

**data.drop\_duplicates(inplace=True)**

## Step 6: Correct Data Types

This line converts the 'dateRep' column to datetime data type. It assumes that 'dateRep' contains dates and converts them to the datetime format, enabling easier date-based operations and analysis.

**['dateRep'] = pd.to\_datetime(data['dateRep'])**

## Step 7: Further Analysis

After preprocessing, you can perform additional analysis, build models, or proceed with specific tasks based on the cleaned data.

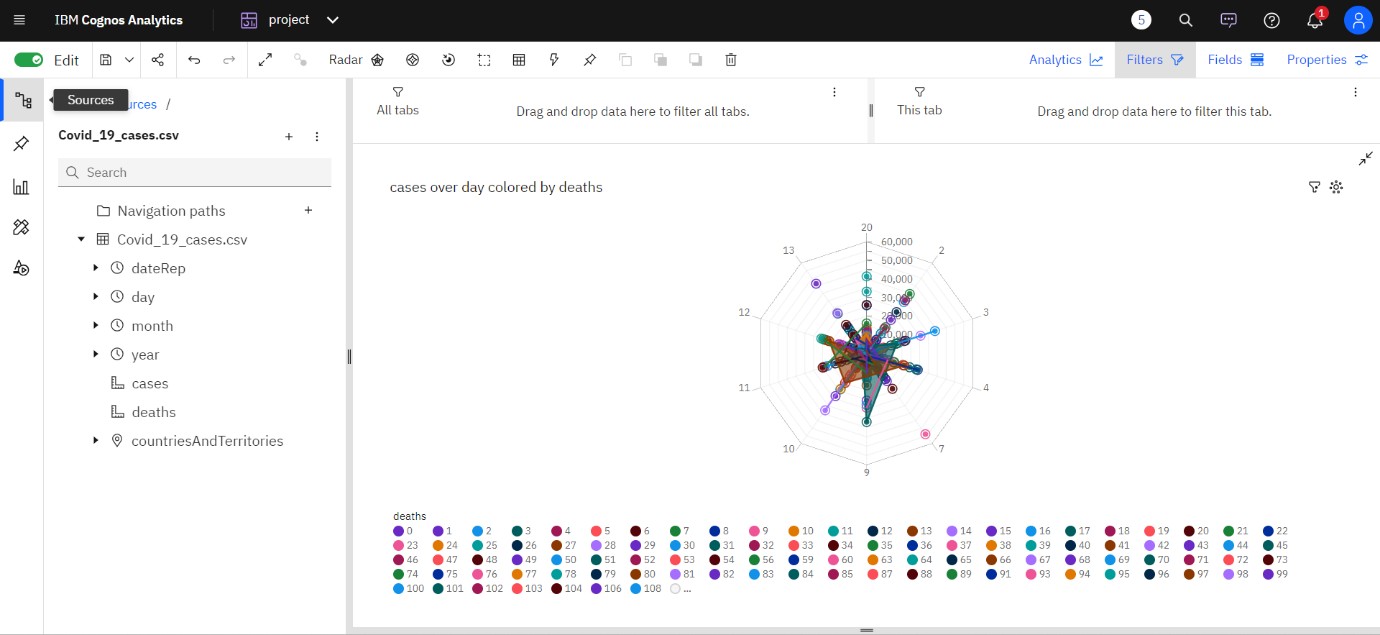
## Step 8: Save the cleaned dataset

Finally, the cleaned dataset is saved to a new CSV file called “Covid\_19\_cases.csv” without including the index column.

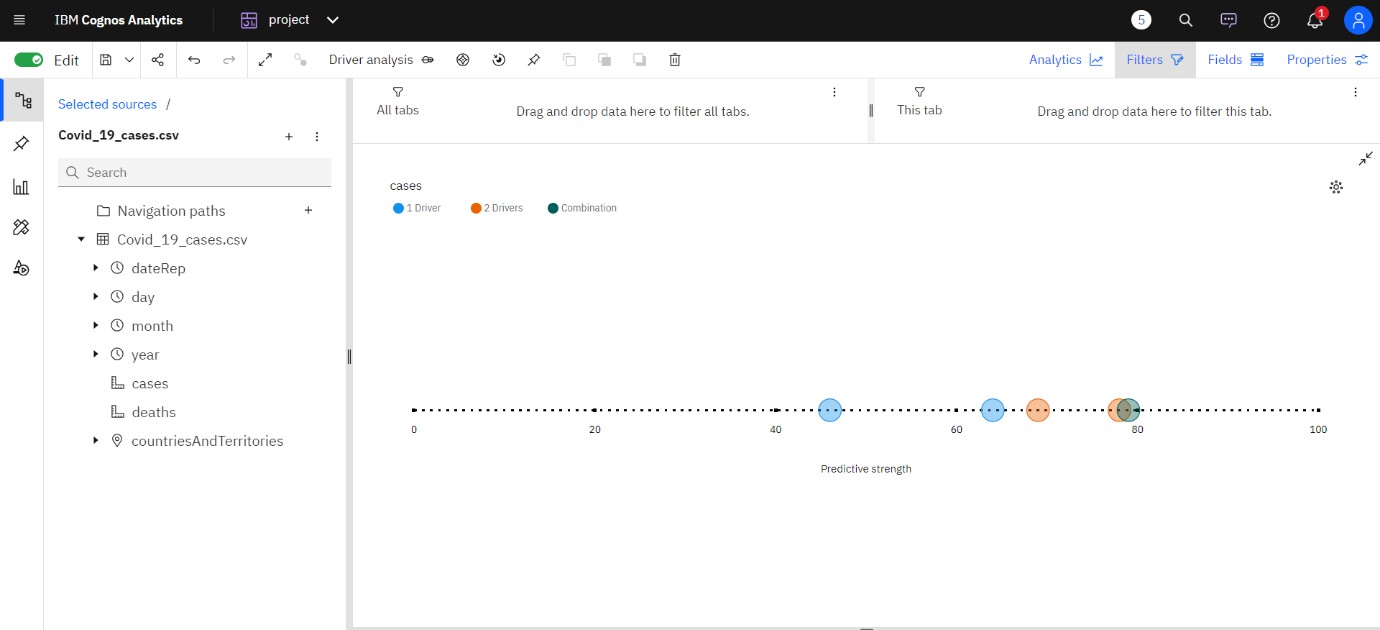
**data.to\_csv(“Covid\_19\_cases.csv”, index=False)**

# Visualization in ibm cognos data analysis

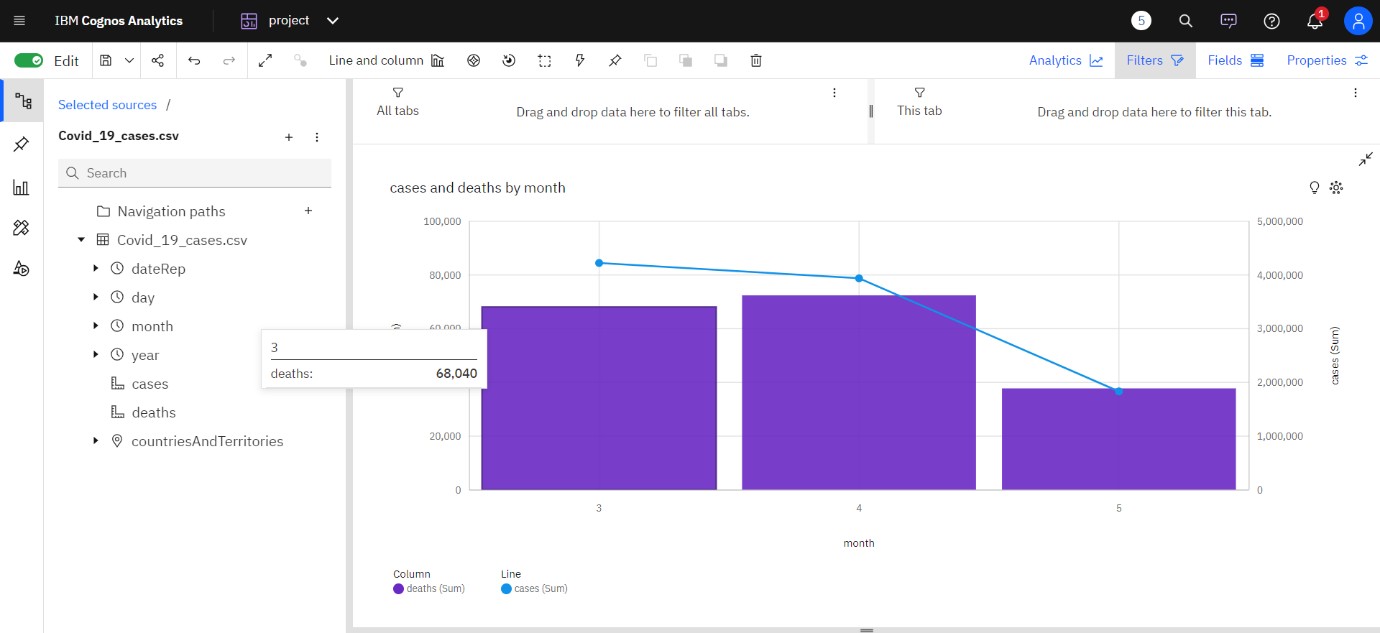
## Visualization of cases over day colored by deaths



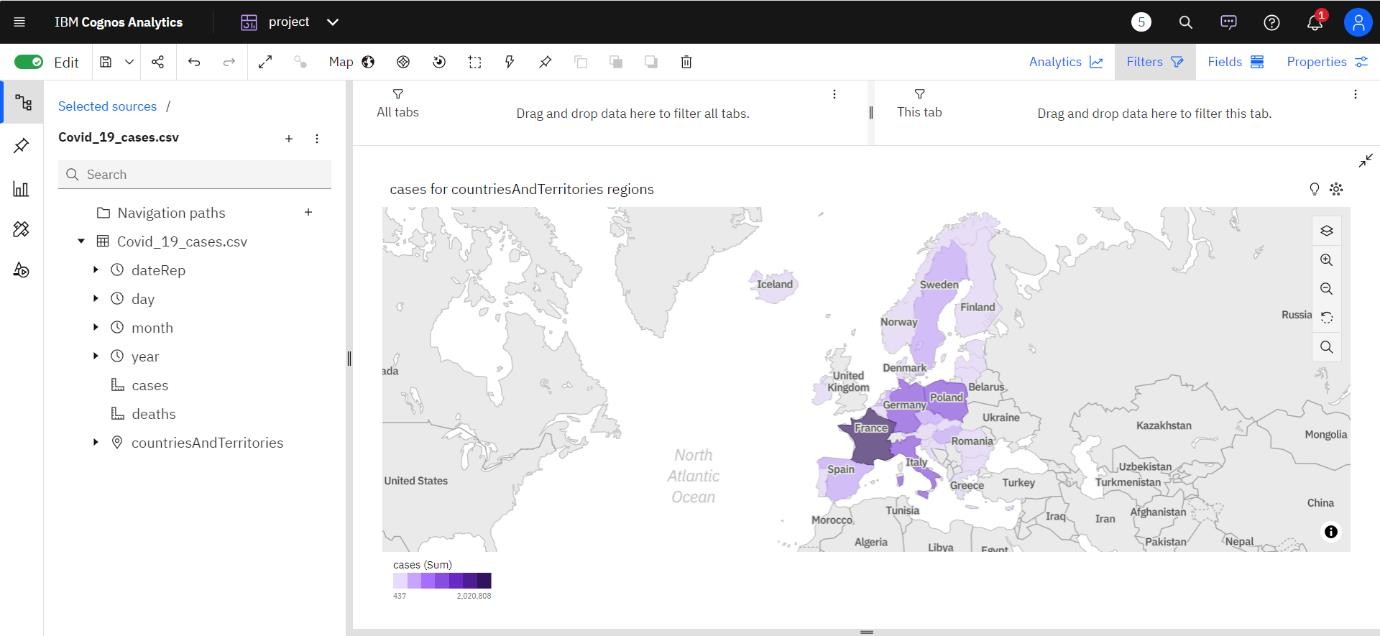
**Visualization by cases**



## Visualization cases and deaths by month



**visualization of cases for countries and territories regions**



# Analysis objectives of covid 19 cases and death

## Temporal Analysis:

### Objective:

Understand how the number of cases and deaths have evolved over time.

### Possible Analyses:

* + Daily, monthly, and yearly trends of COVID-19 cases and deaths.
  + Identify peak periods and analyze factors contributing to the surges.
  + Calculate growth rates to assess the speed of the outbreak.

## Geographical Analysis:

### Objective:

Explore the geographical spread of COVID-19 cases and deaths.

### Possible Analyses:

* + Identify countries with the highest and lowest infection rates.
  + Compare infection and mortality rates across different countries and regions.
  + Visualize the data on maps to understand the global distribution of cases and deaths.

## Mortality Rate Analysis:

### Objective:

Investigate the mortality rate in different countries and regions.

### Possible Analyses:

* + Calculate and compare mortality rates (deaths/cases) for various countries.
  + Analyze factors influencing differences in mortality rates, such as healthcare infrastructure and government responses.

## Seasonal Patterns Analysis:

### Objective:

Determine if there are any seasonal patterns in COVID-19 cases.

### Possible Analyses:

* + Analyze how cases and deaths vary with different seasons.
  + Explore the relationship between climate factors and the spread of the virus.

## Epidemiological Analysis:

### Objective:

Understand the epidemiological characteristics of the virus.

### Possible Analyses:

* + Calculate and analyze the basic reproduction number (R₀) to understand the virus's transmission potential.
  + Study the age distribution of cases and deaths to identify vulnerable populations.
  + Investigate the effect of interventions (such as lockdowns and vaccinations) on the spread of the virus.

## Comparative Analysis:

### Objective:

Compare COVID-19 data with other factors or events to draw correlations.

### Possible Analyses:

* + Correlate COVID-19 data with economic indicators, healthcare spending, or population density.
  + Study the impact of policy measures (such as mask mandates or travel restrictions) on the number of cases and deaths.

## Predictive Analysis:

### Objective:

Develop predictive models to forecast future COVID-19 cases and deaths.

### Possible Analyses:

* + Use time series forecasting techniques to predict future case and death counts.
  + Incorporate external factors (such as vaccination rates) into predictive models for more accurate forecasts.

## Demographic Analysis:

### Objective:

Analyze how COVID-19 affects different demographic groups.

### Possible Analyses:

* + Study the impact on different age groups, genders, and ethnicities.
  + Investigate disparities in healthcare access and outcomes among various demographic groups.