COVID-19 Data Analysis and Visualization

1. Overview

- This project involves the analysis and visualization of COVID-19 data for India, with a specific focus on the state of Delhi.
- The name of the JSON file is: states_daily.json
- The code was written in python language (Python 3.X).
- The IDE used was Jupyter Notebook.
- Used libraries are Pandas, Numpy, matplotlib, sklearn for Linear Regression
- The data is loaded from a JSON file and various analyses, including linear regression, are performed to understand trends in the number of confirmed, recovered, and deceased cases over time.

2. Methodology

2.1 Data Loading

- The data is loaded from the file `states_daily.json` using the `json` library.
- The data is then normalized and converted into a pandas DataFrame for further analysis.

2.2 Data Preprocessing

- The date fields in the dataset are converted to a proper datetime format using `pd.to_datetime()`.
- The data is inspected and cleaned to ensure there are no missing or incorrect values.

2.3 Data Analysis

- Allows analyzer to input specific dates to get the information of the data in that time period. Performed analysis based on the questions asked in the Data Manipulation section.

2.4 Data Visualization

- Plotted the trends asked and understood the trend types.

2.5 Linear Regression

- Implemented a linear regression function to understand the data and predict future data

3. Data Manipulation

3.1 Total Cases Analysis

a) Objective: Calculate the total number of confirmed, recovered, and deceased cases in India from any start and end dates. For an Example : March 14, 2020, to September 5, 2020.

b) Steps:

- The data is filtered by the given date range.
- A pivot table is created to aggregate the data by date and case status (Confirmed, Recovered, Deceased).
- Cumulative sums are calculated to track the progression of cases over time.

3.2 State-Specific Analysis (Delhi)

a) Objective: Calculate the total number of confirmed, recovered, and deceased cases in Delhi for the same period.

b) Steps:

- The data is filtered by state and date range.
- A pivot table is created to aggregate the data by date and case status for Delhi.
- Cumulative sums are calculated for the specific state.

4. Plotting

4.1 Area Trend Line for Total Cases in India

a) Objective: Visualize the trend of total confirmed, recovered, and deceased cases in India.

b) Steps: An area plot is created to show the cumulative totals over time.

4.2 Area Trend Line for Total Cases in Delhi

a) Objective: Visualize the trend of total confirmed, recovered, and deceased cases in Delhi.

b) Steps: An area plot is created to show the cumulative totals for Delhi over time.

4.3 Area Trend Line for Active Cases in India

a) Objective: Visualize the trend of active cases in India (Active = Confirmed - (Recovered + Deceased)).

b) Steps: An area plot is created to show the number of active cases over time.

5. Linear Regression

5.1 Linear Regression on Delhi's COVID-19 Data

a) Objective: Perform linear regression on Delhi's COVID-19 data to identify trends in confirmed, recovered, and deceased cases.

b) Steps:

- The date is converted to a numeric format (e.g., number of days since the start date).
- Linear regression is applied separately to the confirmed, recovered, and deceased case data.
- The intercept and slope coefficients are reported for each case type.

6. Assumptions

- The data provided in the `states_daily.json` file is accurate and complete.
- The date format in the JSON file is consistent.
- The cumulative sums calculated for each case type are correct.

7. Results

COVID-19 Data Analysis and Visualization

- The total counts of confirmed, recovered, and deceased cases for both India and Delhi have been calculated and visualized.
- The linear regression analysis for Delhi provides insights into the trends of COVID-19 cases, with the intercept and slope coefficients indicating the rate of change over time.