Title: Problem for Covid - 19 Data Analysis Project using Python

**Dataset link :**

**Url =**<https://raw.githubusercontent.com/SR1608/Datasets/main/covid-data.csv>

**Perform following analysis on above dataset :**

**1. Import the dataset using Pandas from above mentioned url.**

**2. High Level Data Understanding:**

**a. Find no. of rows & columns in the dataset**

**b. Data types of columns.**

**c. Info & describe of data in dataframe.**

**3. Low Level Data Understanding :**

**a. Find count of unique values in location column.**

**b. Find which continent has maximum frequency using values counts.**

**c. Find maximum & mean value in 'total\_cases'.**

**d. Find 25%,50% & 75% quartile value in 'total\_deaths'.**

**e. Find which continent has maximum 'human\_development\_index'.**

**f. Find which continent has minimum 'gdp\_per\_capita'.**

**4. Filter the dataframe with only this columns**

**['continent','location','date','total\_cases','total\_deaths','gdp\_per\_capita','**

**human\_development\_index'] and update the data frame.**

**5. Data Cleaning**

**a. Remove all duplicates observations**

**b. Find missing values in all columns**

**c. Remove all observations where continent column value is missing**

**Tip : using subset parameter in dropna**

**d. Fill all missing values with 0**

**6. Date time format :**

**a. Convert date column in datetime format using pandas.to\_datetime**

**b. Create new column month after extracting month data from date**

**column.**

**7. Data Aggregation:**

**a. Find max value in all columns using groupby function on 'continent'**

**column**

        Tip: use reset\_index() after applying groupby

**b. Store the result in a new dataframe named 'df\_groupby'.**

**(Use df\_groupby dataframe for all further analysis)**

**8. Feature Engineering :**

**a. Create a new feature 'total\_deaths\_to\_total\_cases' by ratio of**

**'total\_deaths' column to 'total\_cases'**

**9. Data Visualization :**

**a. Perform Univariate analysis on 'gdp\_per\_capita' column by plotting**

**histogram using seaborn dist plot.**

**b. Plot a scatter plot of 'total\_cases' & 'gdp\_per\_capita'**

**c. Plot Pairplot on df\_groupby dataset.**

**d. Plot a bar plot of 'continent' column with 'total\_cases' .**

**Tip : using kind='bar' in seaborn catplot**

**10.Save the df\_groupby dataframe in your local drive using pandas.to\_csv**

**function .**