

Name :- Padmakar Kare

Roll No :- CC-89

PRN :- 202401030021

DataSet Name :- COVID 19 Dataset

*COVID-19 Dataset: 20 Problem Statements

Q1) Find the country with the highest total number of cases.

Q2) Calculate the total number of deaths globally.

Q3) Find the date when the highest number of new cases was recorded globally.

Q4) Compute the death rate (deaths / cases) for each country.

Q5) Find the top 5 countries with the highest recovery rate.

Q6) Find the monthly average number of new cases

Q7) Identify the country with the lowest active cases on the latest date

Q8) Calculate the global positivity rate (total cases / tests performed).

Q9) Find the number of countries that have more than 1 million cases

Q10) Find the total number of active cases across all countries.

Q11) Find the trend of new cases over time globally (plot line graph).

Q12) Find the top 3 countries with the highest number of tests performed.

Q13) Calculate the Case Fatality Rate (CFR) for each country.

Q14) Identify the country with the highest spike in new cases in a single day.

Q15) Calculate moving average (7 days) for global new cases.

Q16) Compare total deaths of two countries: India vs USA.

Q17) Find the date when a country (say Italy) crossed 1,00,000 cases.

Q18) Identify the 5 dates with the maximum global deaths recorded.

Q19) For each country, find how many days it took to reach 50,000 cases from first reported case.

Q20) Calculate the global positivity rate (total cases / tests performed).

Solutions :-

Q1) :-

```
covid_df.groupby('Country')['Total_Cases'].max().idxmax()
```

Q2) :- covid_df['Total_Deaths'].sum()

Q3) :- covid_df.groupby('Date')['New_Cases'].sum().idxmax()

Q4) :- covid_df['Death_Rate'] = covid_df['Total_Deaths'] /
covid_df['Total_Cases']

Q5) :- covid_df['Recovery_Rate'] =
covid_df['Total_Recovered'] / covid_df['Total_Cases']
covid_df.groupby('Country')['Recovery_Rate'].mean().nlargest
(5)

Q6) :- covid_df['Month'] =
pd.to_datetime(covid_df['Date']).dt.month
covid_df.groupby('Month')['New_Cases'].mean()

Q7) :- latest_date = covid_df['Date'].max()
covid_df_latest = covid_df[covid_df['Date'] == latest_date]

```
covid_df_latest.loc[covid_df_latest['Active_Cases'].idxmin()]
```

```
Q8) :- positivity_rate = covid_df['Total_Cases'].sum() /  
covid_df['Tests_Performed'].sum()
```

```
Q9) :- covid_df_latest[covid_df_latest['Total_Cases'] >  
1_000_000]['Country'].nunique()
```

```
Q10) :- covid_df_latest['Active_Cases'].sum()
```

```
Q11) :- covid_df.groupby('Date')['New_Deaths'].sum().mean()
```

```
Q12) :- import matplotlib.pyplot as plt
```

```
global_new_cases =  
covid_df.groupby('Date')['New_Cases'].sum()  
global_new_cases.plot(figsize=(10,6))  
plt.title('Global New COVID-19 Cases Over Time')  
plt.xlabel('Date')  
plt.ylabel('New Cases')  
plt.show()
```

Q13) :-

```
covid_df_latest.groupby('Country')['Tests_Performed'].sum().  
nlargest(3)
```

Q14) :- covid_df['Case_Fatality_Rate'] =

```
(covid_df['Total_Deaths'] / covid_df['Total_Cases']) * 100
```

Q15) :- covid_df.loc[covid_df['New_Cases'].idxmax()]

Q16) :- global_new_cases.rolling(window=7).mean()

Q17) :- india_deaths = covid_df[covid_df['Country'] ==
'India']['Total_Deaths'].max()

usa_deaths = covid_df[covid_df['Country'] ==
'USA']['Total_Deaths'].max()

Q18) :- italy = covid_df[covid_df['Country'] == 'Italy']

```
italy[italy['Total_Cases'] > 100_000]['Date'].iloc[0]
```

Q19) :-

```
covid_df.groupby('Date')['New_Deaths'].sum().nlargest(5)
```

Q20) :- positivity_rate = covid_df['Total_Cases'].sum() /
covid_df['Tests_Performed'].sum()

