

```
In [2]: import covid19main
```

```
=== Dataset Structure ===  
Shape (rows, columns): (350085, 67)
```

```
=== Column Check ===
```

```
Columns: ['iso_code', 'continent', 'location', 'date', 'total_cases', 'new_cases', 'new_cases_smoothed', 'total_deaths', 'new_deaths', 'new_deaths_smoothed', 'total_cases_per_million', 'new_cases_per_million', 'new_cases_smoothed_per_million', 'total_deaths_per_million', 'new_deaths_per_million', 'new_deaths_smoothed_per_million', 'reproduction_rate', 'icu_patients', 'icu_patients_per_million', 'hosp_patients', 'hosp_patients_per_million', 'weekly_icu_admissions', 'weekly_icu_admissions_per_million', 'weekly_hosp_admissions', 'weekly_hosp_admissions_per_million', 'total_tests', 'new_tests', 'total_tests_per_thousand', 'new_tests_per_thousand', 'new_tests_smoothed', 'new_tests_smoothed_per_thousand', 'positive_rate', 'tests_per_case', 'tests_units', 'total_vaccinations', 'people_vaccinated', 'people_fully_vaccinated', 'total_boosters', 'new_vaccinations', 'new_vaccinations_smoothed', 'total_vaccinations_per_hundred', 'people_vaccinated_per_hundred', 'people_fully_vaccinated_per_hundred', 'total_boosters_per_hundred', 'new_vaccinations_smoothed_per_million', 'new_people_vaccinated_smoothed', 'new_people_vaccinated_smoothed_per_hundred', 'stringency_index', 'population_density', 'median_age', 'aged_65 Older', 'aged_70 Older', 'gdp_per_capita', 'extreme_poverty', 'cardiovasc_death_rate', 'diabetes_prevalence', 'female_smokers', 'male_smokers', 'handwashing_facilities', 'hospital_beds_per_thousand', 'life_expectancy', 'human_development_index', 'population', 'excess_mortality_cumulative_absolute', 'excess_mortality_cumulative', 'excess_mortality', 'excess_mortality_cumulative_per_million']
```

```
=== Data Preview ===
```

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed
0	AFG	Asia	Afghanistan	2020-01-03	NaN	0.0	NaN	NaN	0.0	
1	AFG	Asia	Afghanistan	2020-01-04	NaN	0.0	NaN	NaN	0.0	
2	AFG	Asia	Afghanistan	2020-01-05	NaN	0.0	NaN	NaN	0.0	

3 rows × 67 columns

```
=== Missing Values Analysis ===
```

```
Null values per column:
```

```
iso_code          0  
continent        16665  
location          0  
date              0  
total_cases      37997  
...  
population        0  
excess_mortality_cumulative_absolute  337901  
excess_mortality_cumulative          337901  
excess_mortality          337901  
excess_mortality_cumulative_per_million  337901  
Length: 67, dtype: int64
```

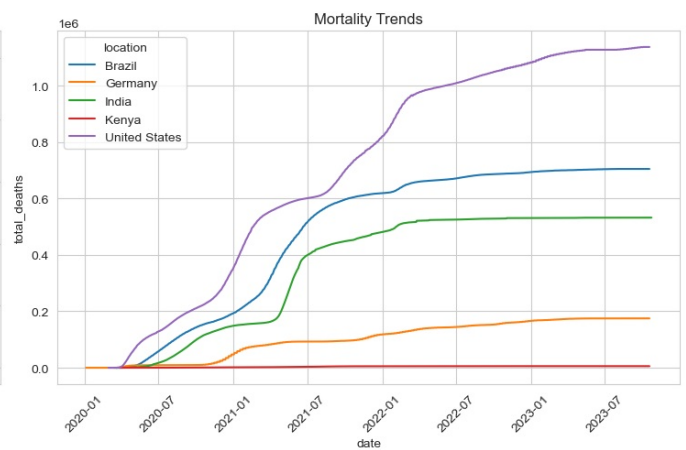
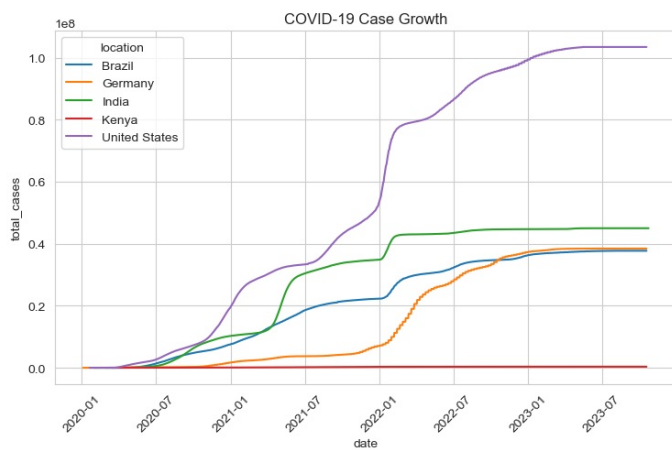
```
=== Handling Missing Values ===
```

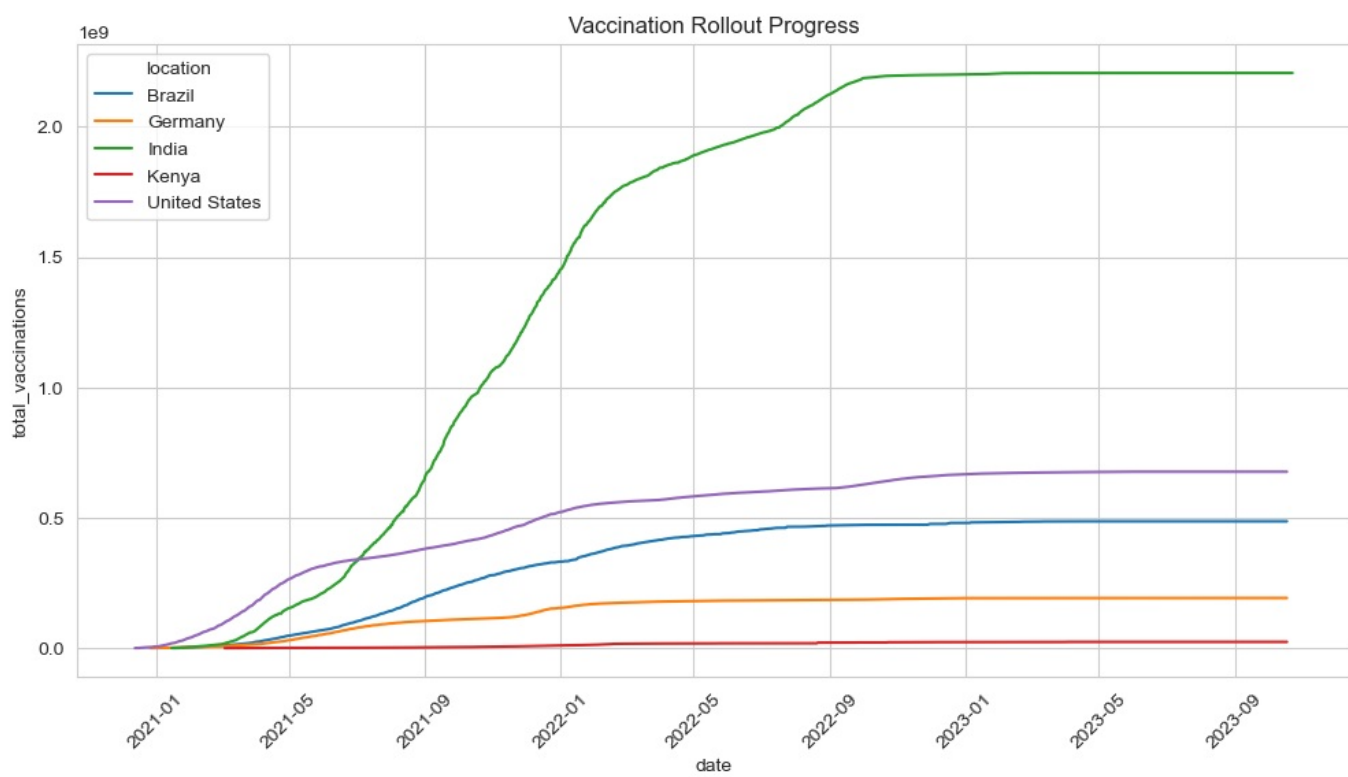
```
Remaining rows after cleaning: 6759
```

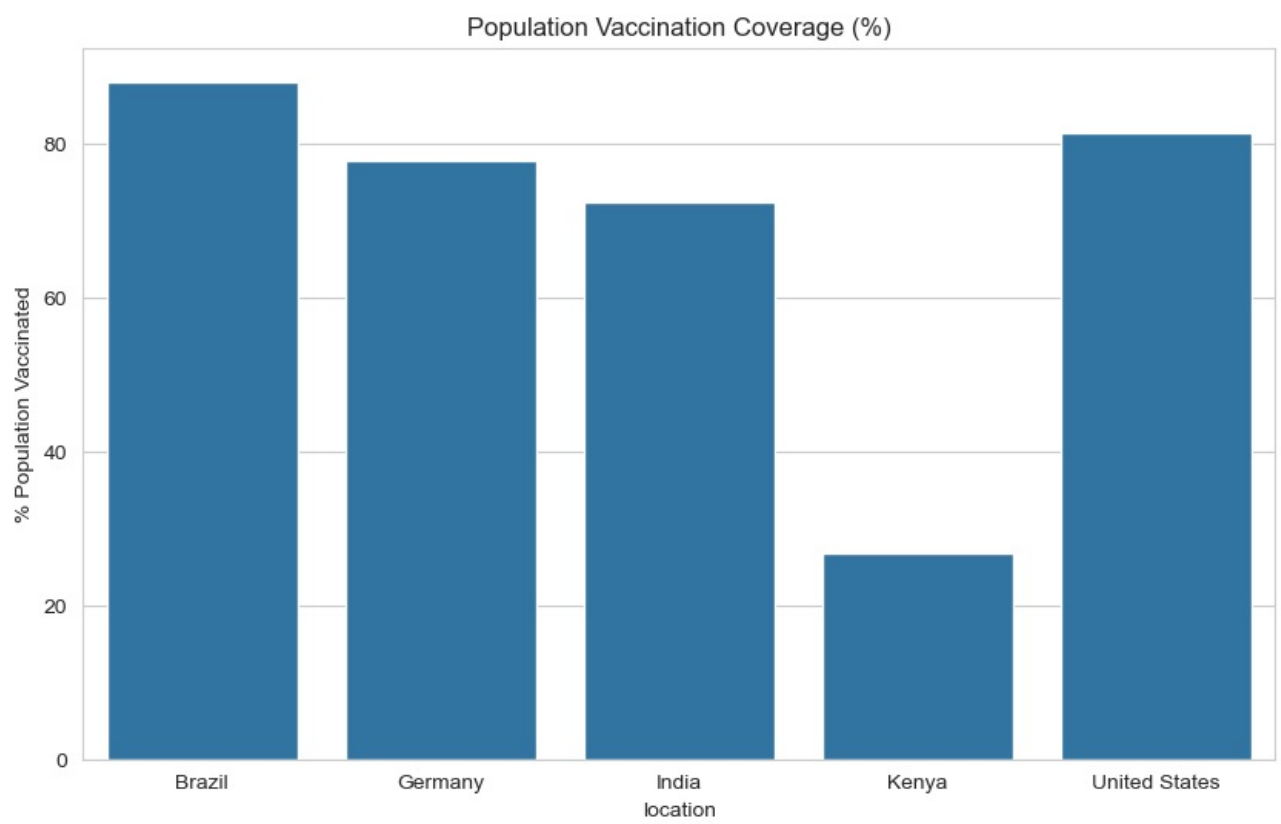
```
C:\Users\hp\covid19main.py:69: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame.  
Try using .loc[row_indexer,col_indexer] = value instead
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy
```

```
df_clean['death_rate'] = (df_clean['total_deaths'] / df_clean['total_cases']) * 100
```







=== Analysis Complete ===

In []:

In []:

In []:

In []:

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js