

```
[1]: import pandas as pd

# Importing the dataset
url = "https://raw.githubusercontent.com/SR1608/Datasets/main/covid-data.csv"
df = pd.read_csv(url)
```

```
[2]: rows, columns = df.shape
print(f"Number of rows: {rows}")
print(f"Number of columns: {columns}")
print(df.dtypes)
print(df.info())
print(df.describe())
```

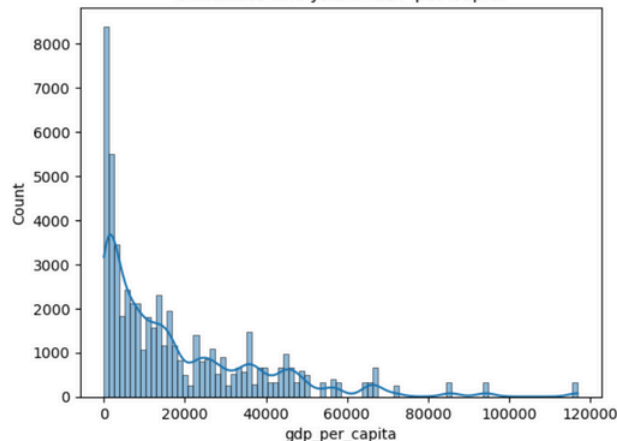
```
std      13.453566      31.645306      2.513193
min       7.700000       1.188000       0.100000
25%      21.400000      21.222000       1.300000
50%      31.400000      52.232000       2.500000
75%      40.900000      83.741000       4.200000
max       78.100000      98.999000      13.800000
```

```
life_expectancy  human_development_index
count      56336.000000      49247.000000
mean         73.937780         0.722223
std           7.397016         0.153261
min          53.280000         0.354000
25%          69.870000         0.601000
```

```
[5]: unique_location:
print(f"Unique \
continent_frequ
max_frequency_cc
print(f"Continer
max_total_cases
mean_total_case:
print(f"Maximum
print(f"Mean tot
```

Continent with minimum GDP per capita: Africa
Missing values in each column:

Univariate Analysis of GDP per Capita



```
[10]: quantiles_total_
print(f"Quantile
max_hdi_continer
print(f"Continer
min_gdp_continer
print(f"Continer
df = df[['contir
```

```
Quantile values
0.25      13.0
0.50      84.0
0.75     727.0
Name: total_deal
Continent with r.....
```

```
[17]: sns.scatterplot(x='total_cases', y='gdp_per_capita', data=df)
plt.title('Scatter plot of Total Cases vs GDP per Capita')
plt.show()
sns.pairplot(df.groupby())
plt.title('Pairplot of Grouped Data by Continent')
plt.show()
```

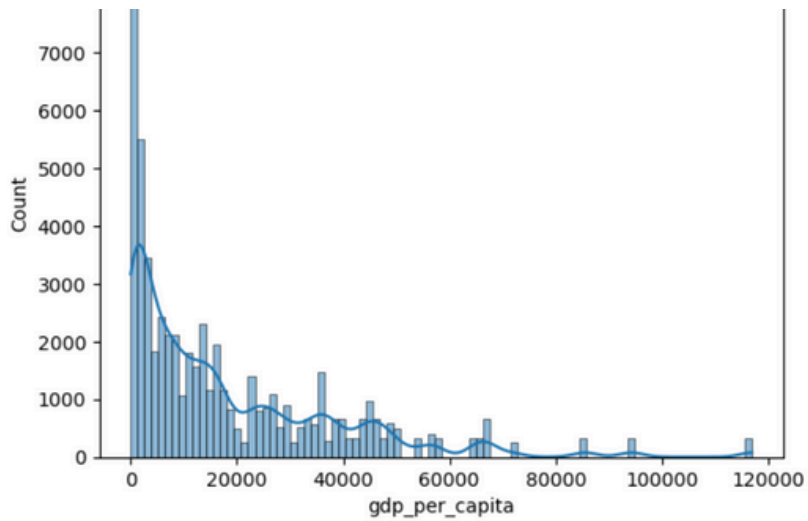
```
date      0
total_cases      3636
total_deaths     13026
gdp_per_capita    7027
human_development_index    8147
dtype: int64
```

```
[13]: df = df.dropna(subset=['continent'])
df = df.fillna(0)
df['date'] = pd.to_datetime(df['date'])
df['month'] = df['date'].dt.month
df_groupby = df.groupby('continent').max().reset_index()
df_groupby['total_deaths_to_total_cases'] = df_groupby['total_deaths'] / df_groupby['total_cases']
```

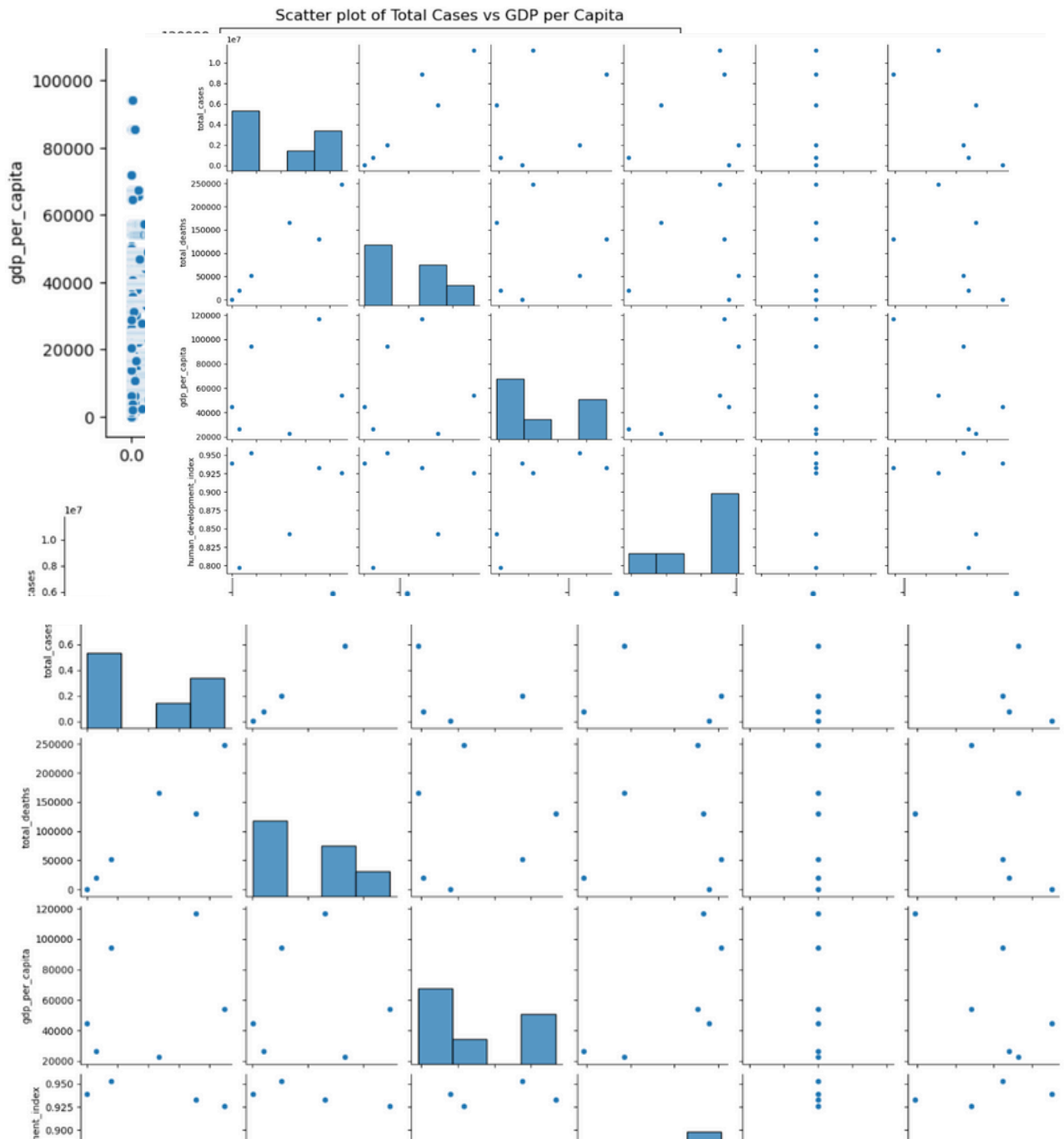
C:\Users\shrey\AppData\Local\Temp\ipykernel_19452\919386993.py:3: UserWarning: Could not infer format, so each element will be parsed individually, falling back to 'dateutil'. To ensure parsing is consistent and as-expected, please specify a format.
df['date'] = pd.to_datetime(df['date'])

```
[15]: import seaborn as sns
import matplotlib.pyplot as plt

sns.histplot(df['gdp_per_capita'], kde=True)
plt.title('Univariate Analysis of GDP per Capita')
```



```
[17]: sns.scatterplot(x='total_cases', y='gdp_per_capita', data=df)
plt.title('Scatter plot of Total Cases vs GDP per Capita')
plt.show()
```



```
df_groupby.to_csv('df_groupby.csv', index=False)  
print("DataFrame saved to df_groupby.csv")
```

