World Population datset

In [130…
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
import warnings
warnings.filterwarnings('ignore')
```

In [57]:
```
dataset=pd.read_csv(r"C:\Users\S SHYAMILI\OneDrive\Desktop\data science\prodigy\wor
```

In [59]:
```
dataset
```

Out[59]:

	Rank	CCA3	Country/Territory	Capital	Continent	2022 Population	2020 Population	20 Populati
0	36	AFG	Afghanistan	Kabul	Asia	41128771	38972230	337534
1	138	ALB	Albania	Tirana	Europe	2842321	2866849	28824
2	34	DZA	Algeria	Algiers	Africa	44903225	43451666	395431
3	213	ASM	American Samoa	Pago Pago	Oceania	44273	46189	513
4	203	AND	Andorra	Andorra la Vella	Europe	79824	77700	717
…	…	…	…	…	…	…	…	
229	226	WLF	Wallis and Futuna	Mata-Utu	Oceania	11572	11655	121
230	172	ESH	Western Sahara	El Aaiún	Africa	575986	556048	4918
231	46	YEM	Yemen	Sanaa	Asia	33696614	32284046	285165
232	63	ZMB	Zambia	Lusaka	Africa	20017675	18927715	162482
233	74	ZWE	Zimbabwe	Harare	Africa	16320537	15669666	141549

234 rows × 17 columns

In [74]:
```
dataset.isnull().any().sum()
```

Out[74]: 0

In [63]:
```
dataset.shape
```

Out[63]: (234, 17)

In [65]:
```
dataset.describe()
```

Out[65]:

	Rank	2022 Population	2020 Population	2015 Population	2010 Population	2000 Population
count	234.000000	2.340000e+02	2.340000e+02	2.340000e+02	2.340000e+02	2.340000e+02
mean	117.500000	3.407441e+07	3.350107e+07	3.172996e+07	2.984524e+07	2.626947e+07
std	67.694165	1.367664e+08	1.355899e+08	1.304050e+08	1.242185e+08	1.116982e+08
min	1.000000	5.100000e+02	5.200000e+02	5.640000e+02	5.960000e+02	6.510000e+02
25%	59.250000	4.197385e+05	4.152845e+05	4.046760e+05	3.931490e+05	3.272420e+05
50%	117.500000	5.559944e+06	5.493074e+06	5.307400e+06	4.942770e+06	4.292907e+06
75%	175.750000	2.247650e+07	2.144798e+07	1.973085e+07	1.915957e+07	1.576230e+07
max	234.000000	1.425887e+09	1.424930e+09	1.393715e+09	1.348191e+09	1.264099e+09

In [67]:
```
dataset.head(10)
```

Out[67]:

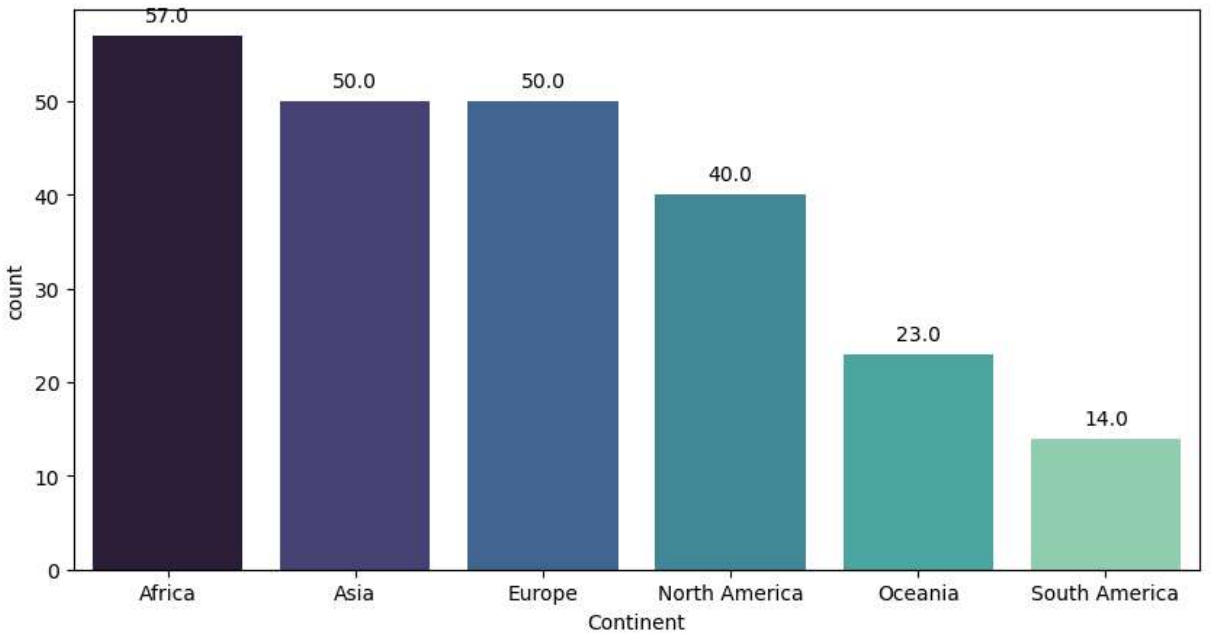
	Rank	CCA3	Country/Territory	Capital	Continent	2022 Population	2020 Population	2015 Population
0	36	AFG	Afghanistan	Kabul	Asia	41128771	38972230	33753499
1	138	ALB	Albania	Tirana	Europe	2842321	2866849	2882481
2	34	DZA	Algeria	Algiers	Africa	44903225	43451666	39543154
3	213	ASM	American Samoa	Pago Pago	Oceania	44273	46189	51368
4	203	AND	Andorra	Andorra la Vella	Europe	79824	77700	71746
5	42	AGO	Angola	Luanda	Africa	35588987	33428485	28127721
6	224	AIA	Anguilla	The Valley	North America	15857	15585	14525
7	201	ATG	Antigua and Barbuda	Saint John's	North America	93763	92664	89941
8	33	ARG	Argentina	Buenos Aires	South America	45510318	45036032	43257065
9	140	ARM	Armenia	Yerevan	Asia	2780469	2805608	2878595

In [69]:
```
dataset.nunique()
```

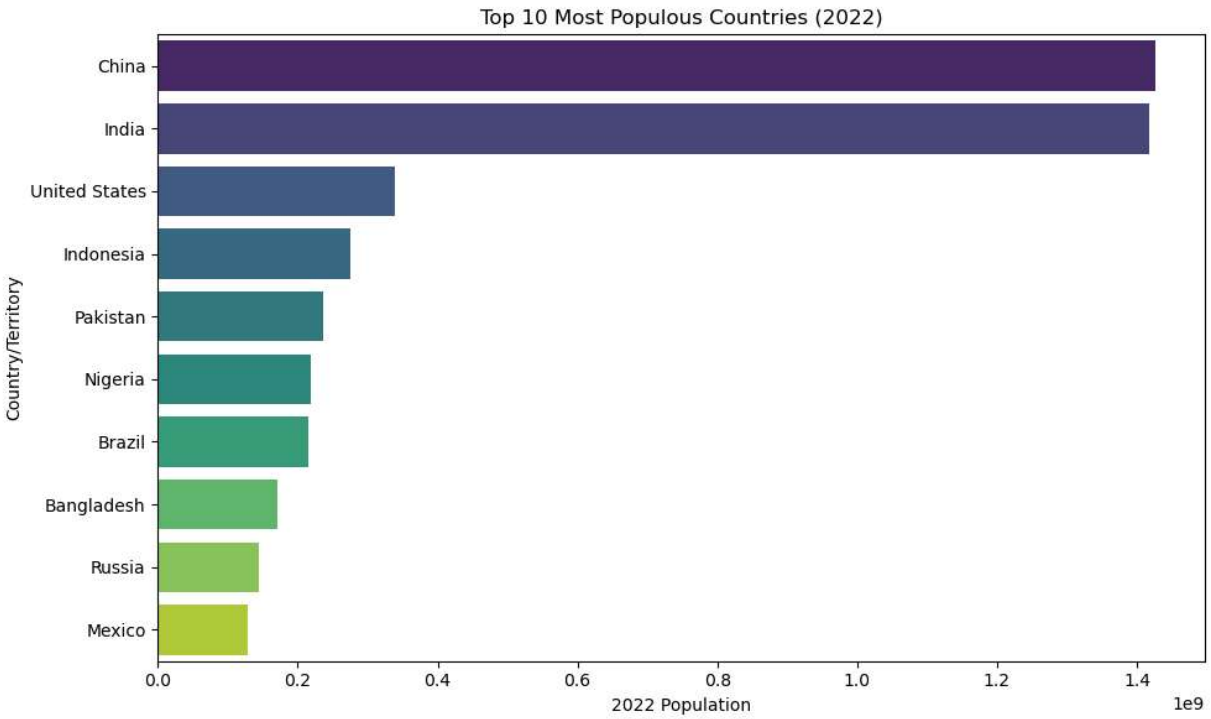
Out[69]:
```
Rank                            234
CCA3                            234
Country/Territory               234
Capital                         234
Continent                         6
2022 Population                 234
2020 Population                 234
2015 Population                 234
2010 Population                 234
2000 Population                 234
1990 Population                 234
1980 Population                 234
1970 Population                 234
Area (km²)                      233
Density (per km²)               234
Growth Rate                     180
World Population Percentage      70
dtype: int64
```

In [132…
```
plt.figure(figsize=(10, 5))
ax = sns.countplot(x='Continent', data=dataset, order=dataset['Continent'].value_co

for p in ax.patches:
    ax.annotate(f'{p.get_height()}', (p.get_x() + p.get_width() / 2., p.get_height(
```

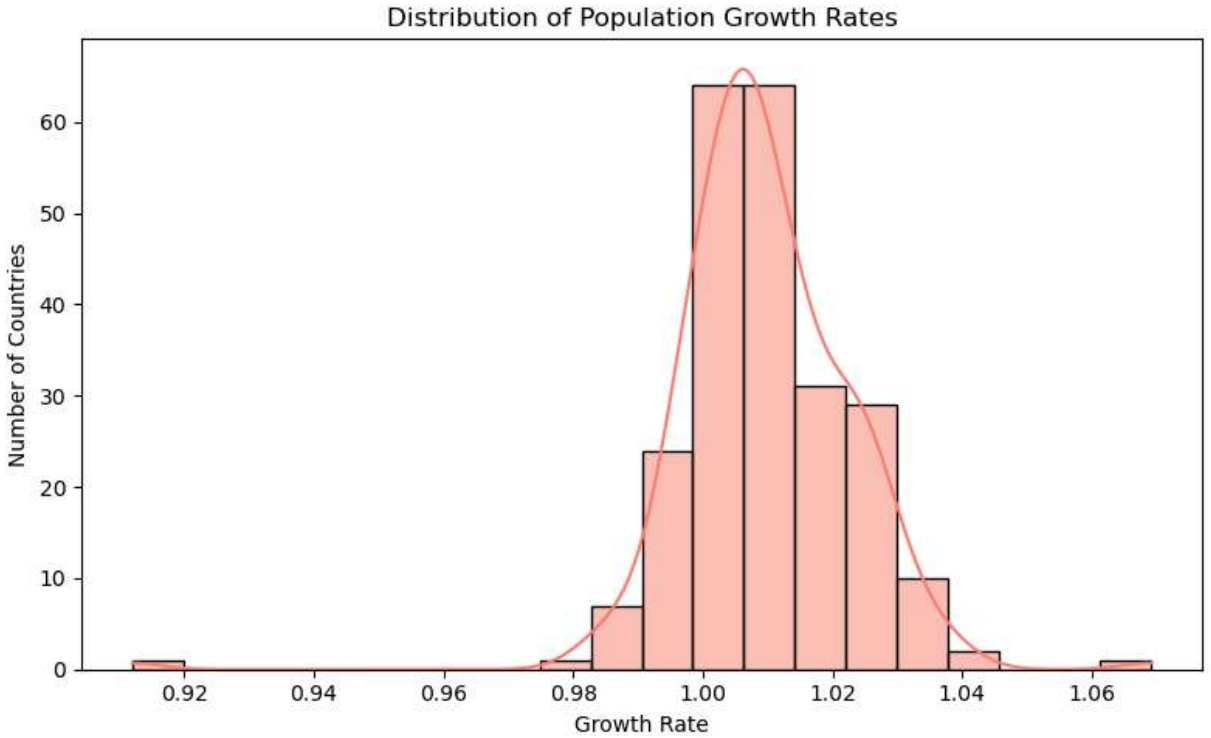


In [134…
```
# 1. Top 10 countries by 2022 Population
top10_pop = dataset.nlargest(10, '2022 Population')
plt.figure(figsize=(10, 6))
sns.barplot(data=top10_pop, x='2022 Population', y='Country/Territory', palette='vi
plt.title('Top 10 Most Populous Countries (2022)')
plt.tight_layout()
plt.show()
```

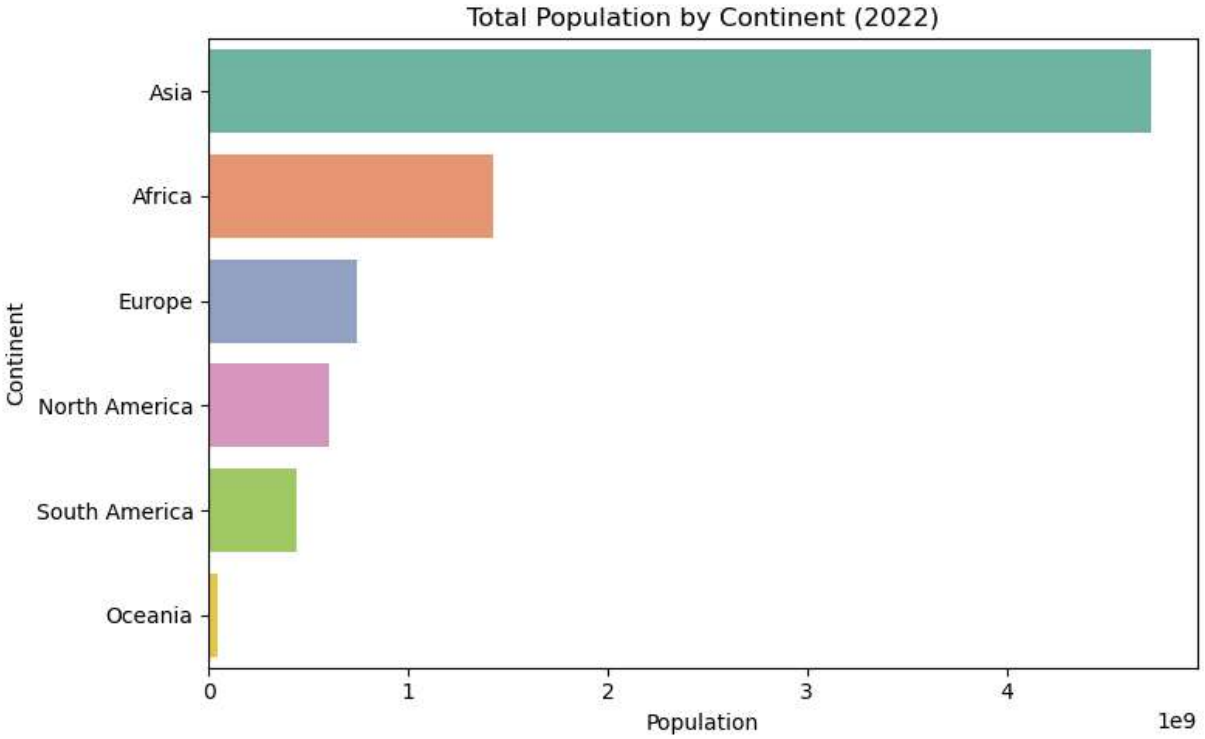


In [137…

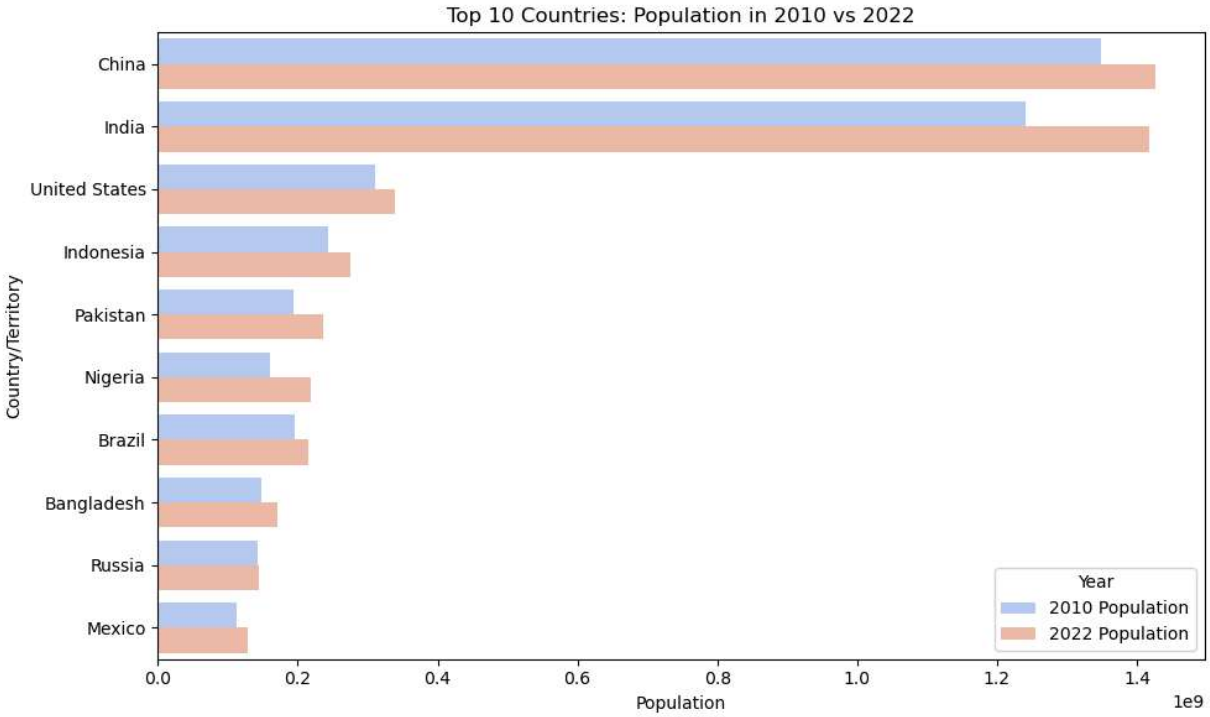
```
#2.Population Growth Rate Distribution (Histogram)
plt.figure(figsize=(8, 5))
sns.histplot(dataset['Growth Rate'], bins=20, kde=True, color='salmon')
plt.title('Distribution of Population Growth Rates')
plt.xlabel('Growth Rate')
plt.ylabel('Number of Countries')
plt.tight_layout()
plt.show()
```



In [139…
```
# 3. Population by Continent (2022)
continent_pop = dataset.groupby('Continent')['2022 Population'].sum().sort_values(a
plt.figure(figsize=(8, 5))
sns.barplot(x=continent_pop.values, y=continent_pop.index, palette='Set2')
plt.title('Total Population by Continent (2022)')
plt.xlabel('Population')
plt.tight_layout()
plt.show()
```

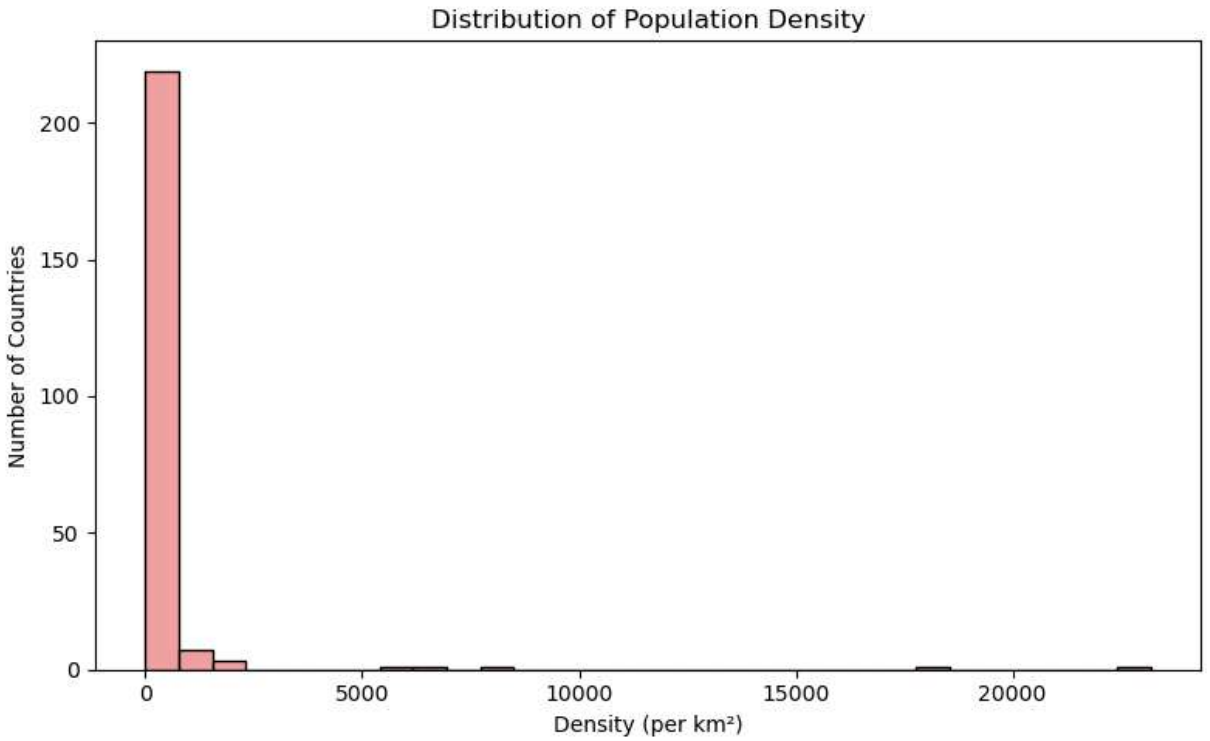


In [145…
```
# 4. Barplot: Population in 2022 vs 2010 for Top 10 Countries
top10 = dataset.nlargest(10, '2022 Population')[['Country/Territory', '2010 Populat
top10_melted = pd.melt(top10, id_vars='Country/Territory', var_name='Year', value_n
plt.figure(figsize=(10, 6))
sns.barplot(data=top10_melted, x='Population', y='Country/Territory', hue='Year', p
plt.title('Top 10 Countries: Population in 2010 vs 2022')
plt.tight_layout()
plt.show()
```



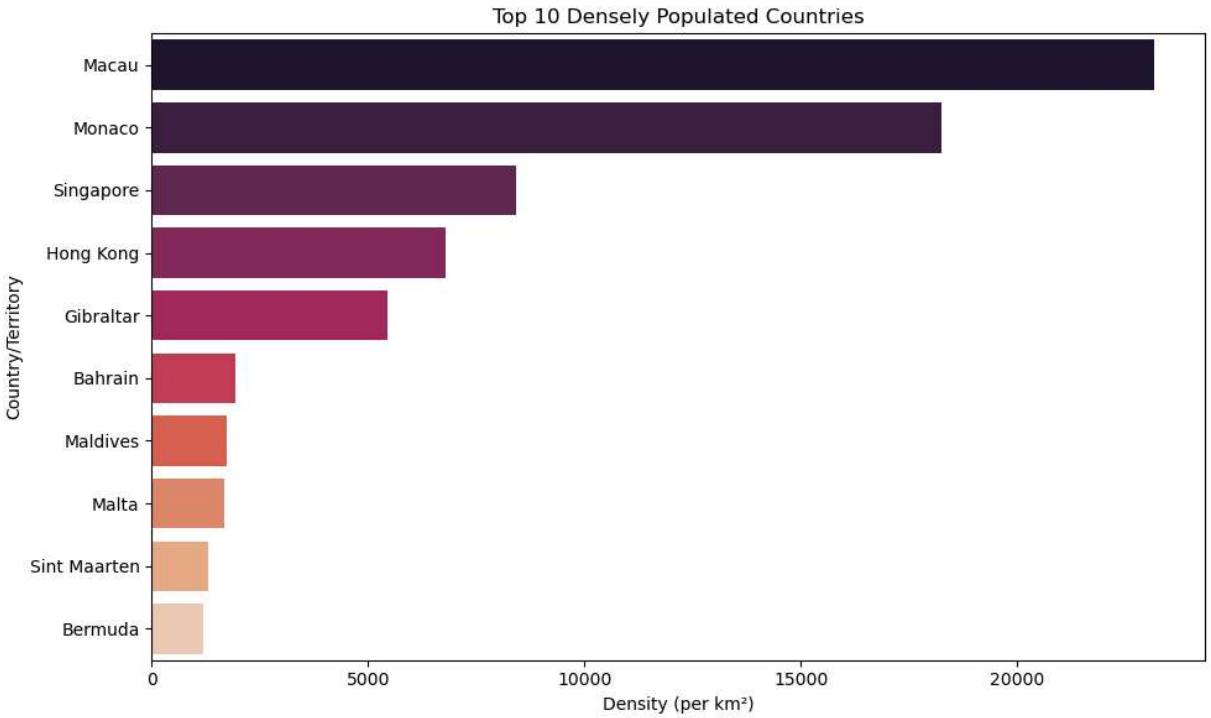
In [147…

```
# 5. Histogram: Density Distribution
plt.figure(figsize=(8, 5))
sns.histplot(dataset['Density (per km²)'], bins=30, kde=False, color='lightcoral')
plt.title('Distribution of Population Density')
plt.xlabel('Density (per km²)')
plt.ylabel('Number of Countries')
plt.tight_layout()
plt.show()
```



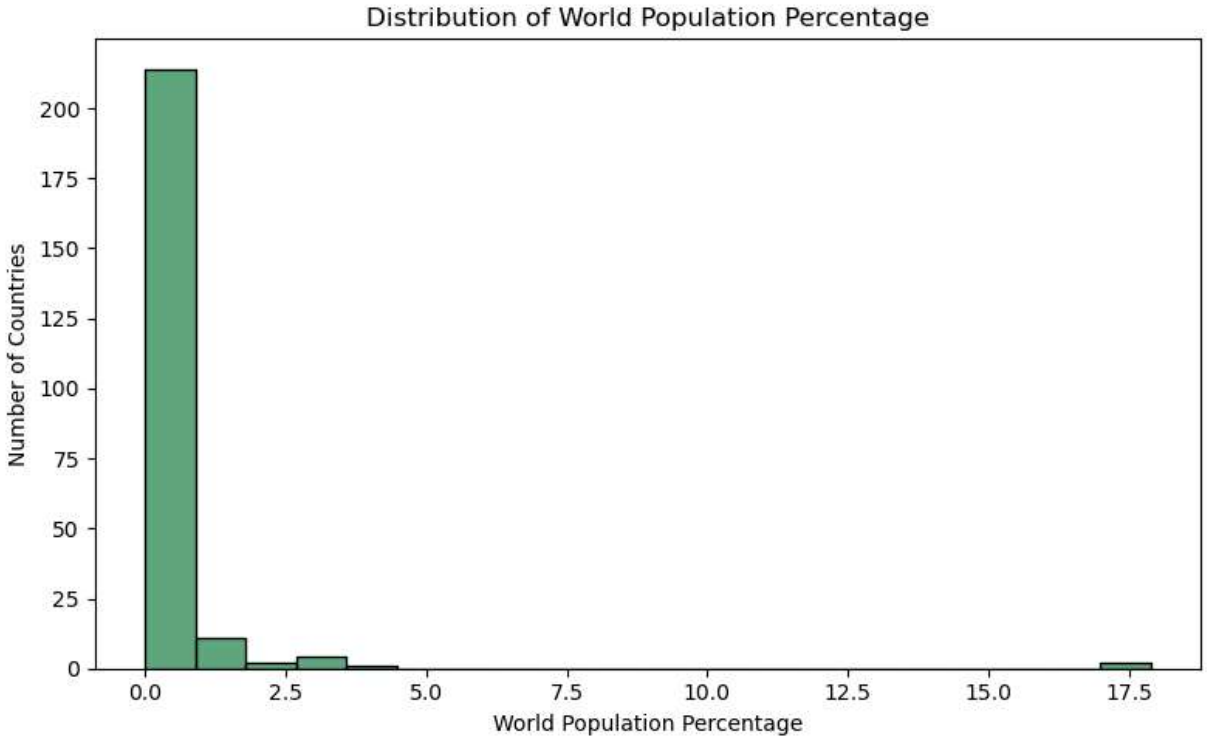
In [149…

```
# 6. Barplot: Top 10 Densely Populated Countries
top10_density = dataset.nlargest(10, 'Density (per km²)')
plt.figure(figsize=(10, 6))
sns.barplot(data=top10_density, x='Density (per km²)', y='Country/Territory', palet
plt.title('Top 10 Densely Populated Countries')
plt.tight_layout()
plt.show()
```



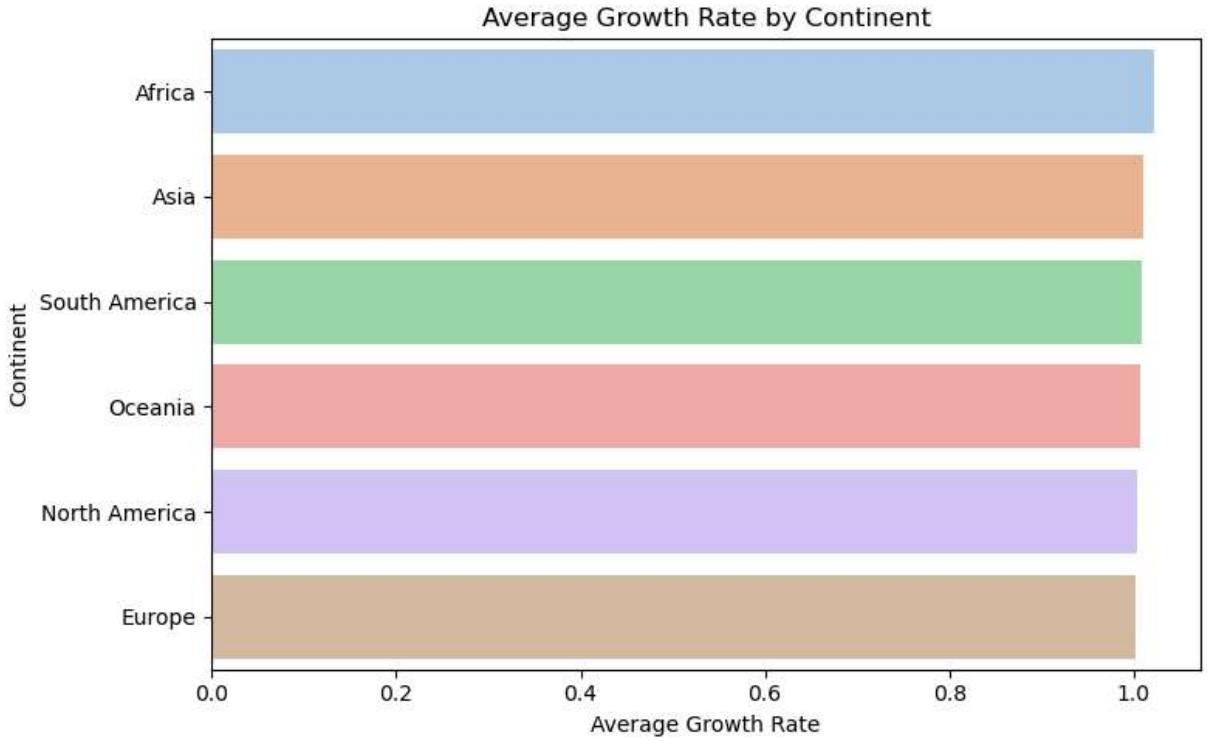
In [151…

```
# 7. Histogram: World Population Percentage Distribution
plt.figure(figsize=(8, 5))
sns.histplot(dataset['World Population Percentage'], bins=20, kde=False, color='sea
plt.title('Distribution of World Population Percentage')
plt.xlabel('World Population Percentage')
plt.ylabel('Number of Countries')
plt.tight_layout()
plt.show()
```

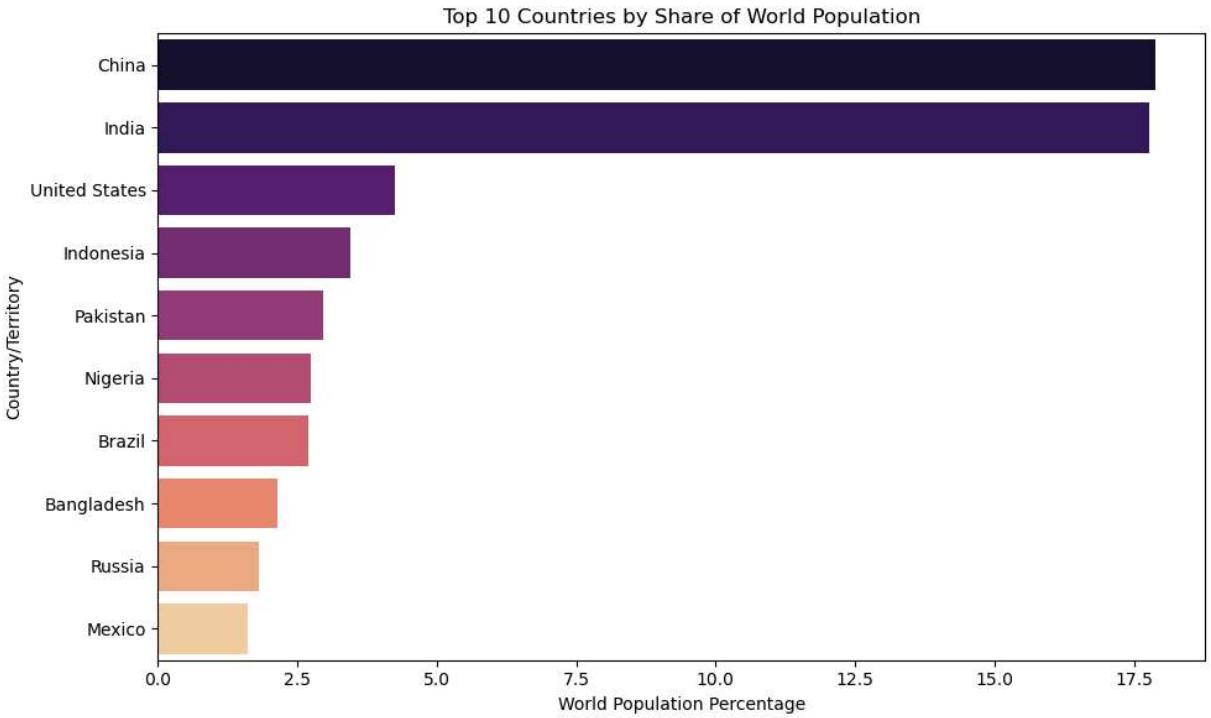


In [153…

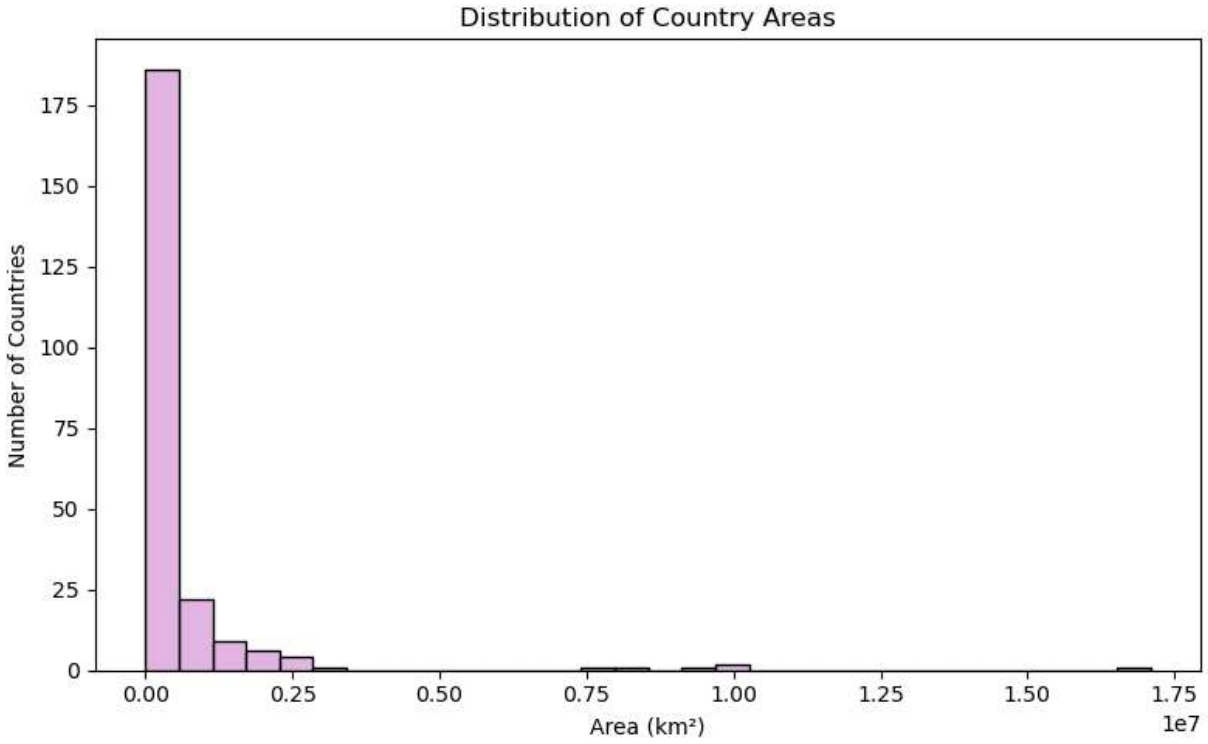
```
# 8. Barplot: Average Growth Rate by Continent
continent_growth = dataset.groupby('Continent')['Growth Rate'].mean().sort_values(a
plt.figure(figsize=(8, 5))
sns.barplot(x=continent_growth.values, y=continent_growth.index, palette='pastel')
plt.title('Average Growth Rate by Continent')
plt.xlabel('Average Growth Rate')
plt.tight_layout()
plt.show()
```



In [155…
```
# 9. Barplot: Top 10 Countries by World Population Share
top10_share = dataset.nlargest(10, 'World Population Percentage')
plt.figure(figsize=(10, 6))
sns.barplot(data=top10_share, x='World Population Percentage', y='Country/Territory
plt.title('Top 10 Countries by Share of World Population')
plt.tight_layout()
plt.show()
```



In [159…
```
# 10. Histogram: Area (km²) Distribution
plt.figure(figsize=(8, 5))
sns.histplot(dataset['Area (km²)'], bins=30, kde=False, color='plum')
plt.title('Distribution of Country Areas')
plt.xlabel('Area (km²)')
plt.ylabel('Number of Countries')
plt.tight_layout()
plt.show()
```


Distribution of Country Areas
Number of Countries
175
150
125
100
75
50
25
0
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
Area (km²)
1e7

In [ ]: