In [33]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import plotly.express as px
import seaborn as sns
```

In [5]:
```
pop=pd.read_csv("world_population.csv")
pop.head(10)
```

Out[5]:

	Rank	CCA3	Country/Territory	Capital	Continent	2022 Population	2020 Population	2015 Population	P
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
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	

DATA EXPLORATION

In [6]:
```
pop.shape
```

Out[6]:
```
(234, 17)
```

In [7]:
```
pop.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 234 entries, 0 to 233
Data columns (total 17 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   Rank                            234 non-null    int64
 1   CCA3                            234 non-null    object
 2   Country/Territory              234 non-null    object
 3   Capital                         234 non-null    object
 4   Continent                       234 non-null    object
 5   2022 Population                 234 non-null    int64
 6   2020 Population                 234 non-null    int64
 7   2015 Population                 234 non-null    int64
 8   2010 Population                 234 non-null    int64
 9   2000 Population                 234 non-null    int64
 10  1990 Population                 234 non-null    int64
 11  1980 Population                 234 non-null    int64
 12  1970 Population                 234 non-null    int64
 13  Area (km²)                      234 non-null    int64
 14  Density (per km²)               234 non-null    float64
 15  Growth Rate                     234 non-null    float64
 16  World Population Percentage     234 non-null    float64
dtypes: float64(3), int64(10), object(4)
memory usage: 31.2+ KB
```

In [8]:
```
pop.describe()
```

Out[8]:

	Rank	2022 Population	2020 Population	2015 Population	2010 Population	200 Populatic
count	234.000000	2.340000e+02	2.340000e+02	2.340000e+02	2.340000e+02	2.340000e+0
mean	117.500000	3.407441e+07	3.350107e+07	3.172996e+07	2.984524e+07	2.626947e+0
std	67.694165	1.367664e+08	1.355899e+08	1.304050e+08	1.242185e+08	1.116982e+0
min	1.000000	5.100000e+02	5.200000e+02	5.640000e+02	5.960000e+02	6.510000e+0
25%	59.250000	4.197385e+05	4.152845e+05	4.046760e+05	3.931490e+05	3.272420e+0
50%	117.500000	5.559944e+06	5.493074e+06	5.307400e+06	4.942770e+06	4.292907e+0
75%	175.750000	2.247650e+07	2.144798e+07	1.973085e+07	1.915957e+07	1.576230e+0
max	234.000000	1.425887e+09	1.424930e+09	1.393715e+09	1.348191e+09	1.264099e+0

In [9]:
```
pop.describe(include=object)
```

Out[9]:

	CCA3	Country/Territory	Capital	Continent
count	234	234	234	234
unique	234	234	234	6
top	AFG	Afghanistan	Kabul	Africa
freq	1	1	1	57

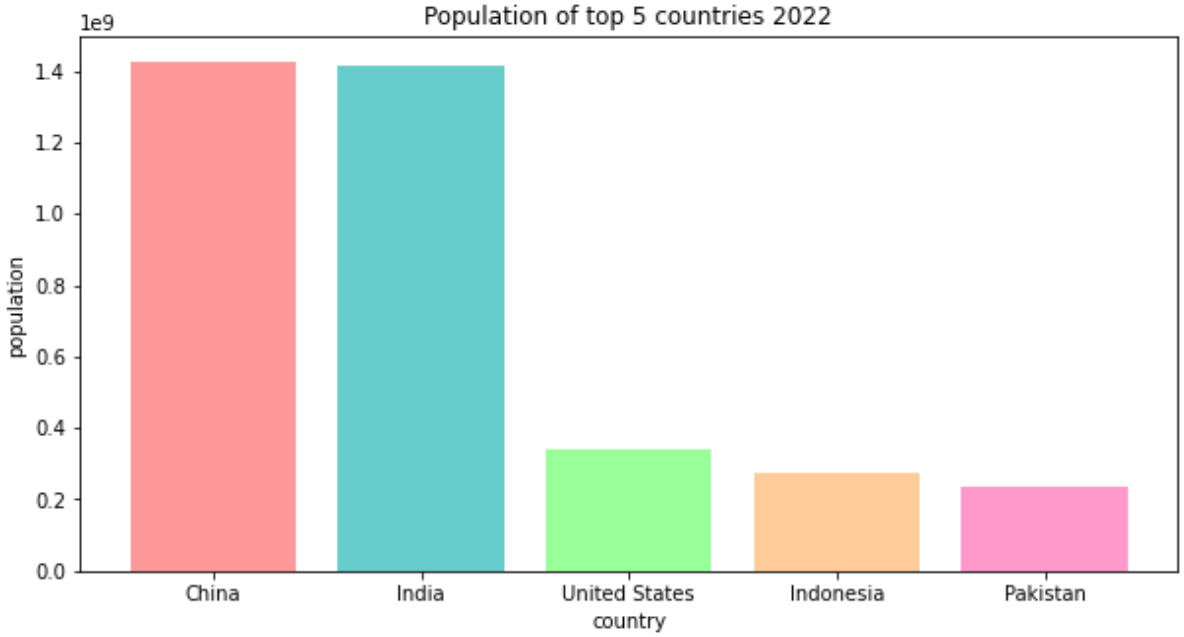
In [10]:
```
pop.isna().sum()
```

Out[10]:
```
Rank                            0
CCA3                            0
Country/Territory               0
Capital                         0
Continent                       0
2022 Population                 0
2020 Population                 0
2015 Population                 0
2010 Population                 0
2000 Population                 0
1990 Population                 0
1980 Population                 0
1970 Population                 0
Area (km²)                      0
Density (per km²)               0
Growth Rate                     0
World Population Percentage     0
dtype: int64
```

In [11]:
```
#defining colours
pastel_colors = ['#FF9999', '#66CCCC', '#99FF99', '#FFCC99', '#FF99CC']
```

Top 5 populated countries in the world in 2022

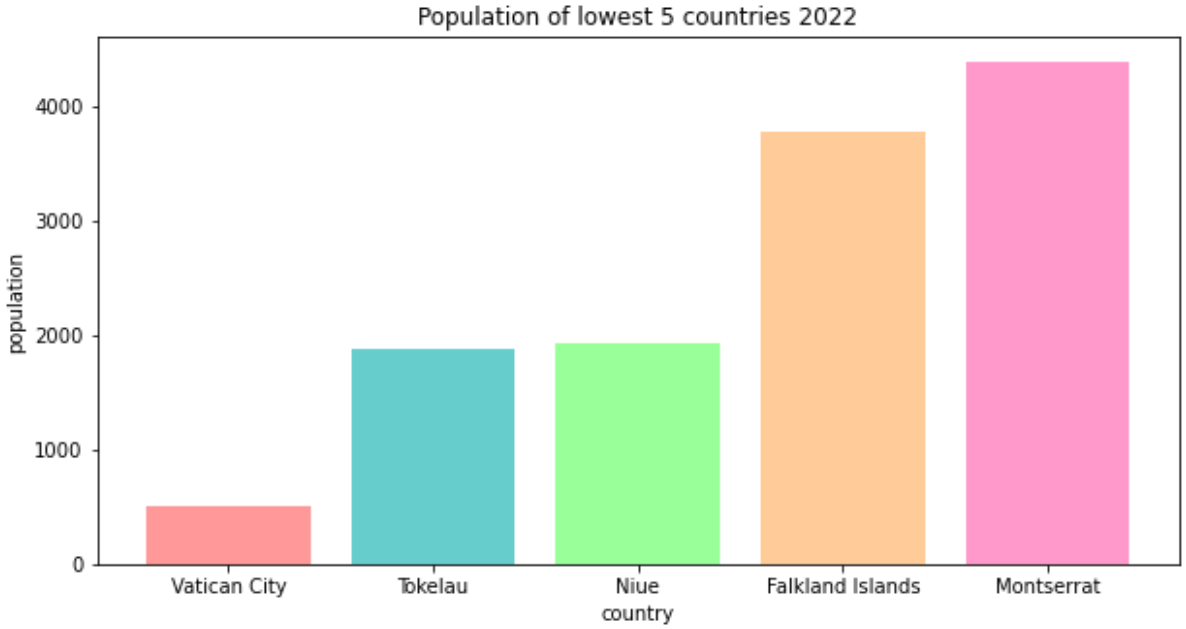
In [12]:
```
#population of top 5 countries 2022

sorted_2022=pop.sort_values(by='2022 Population',ascending=False)
sorted_2022=sorted_2022.head(5)
plt.figure(figsize=(10,5))
plt.bar(sorted_2022['Country/Territory'],sorted_2022['2022 Population'],colo
plt.xlabel('country')
plt.ylabel('population')
plt.title("Population of top 5 countries 2022")
plt.show()
```



We see that china and India are the most populated countries in 2022

5 Least populated countries in 2022

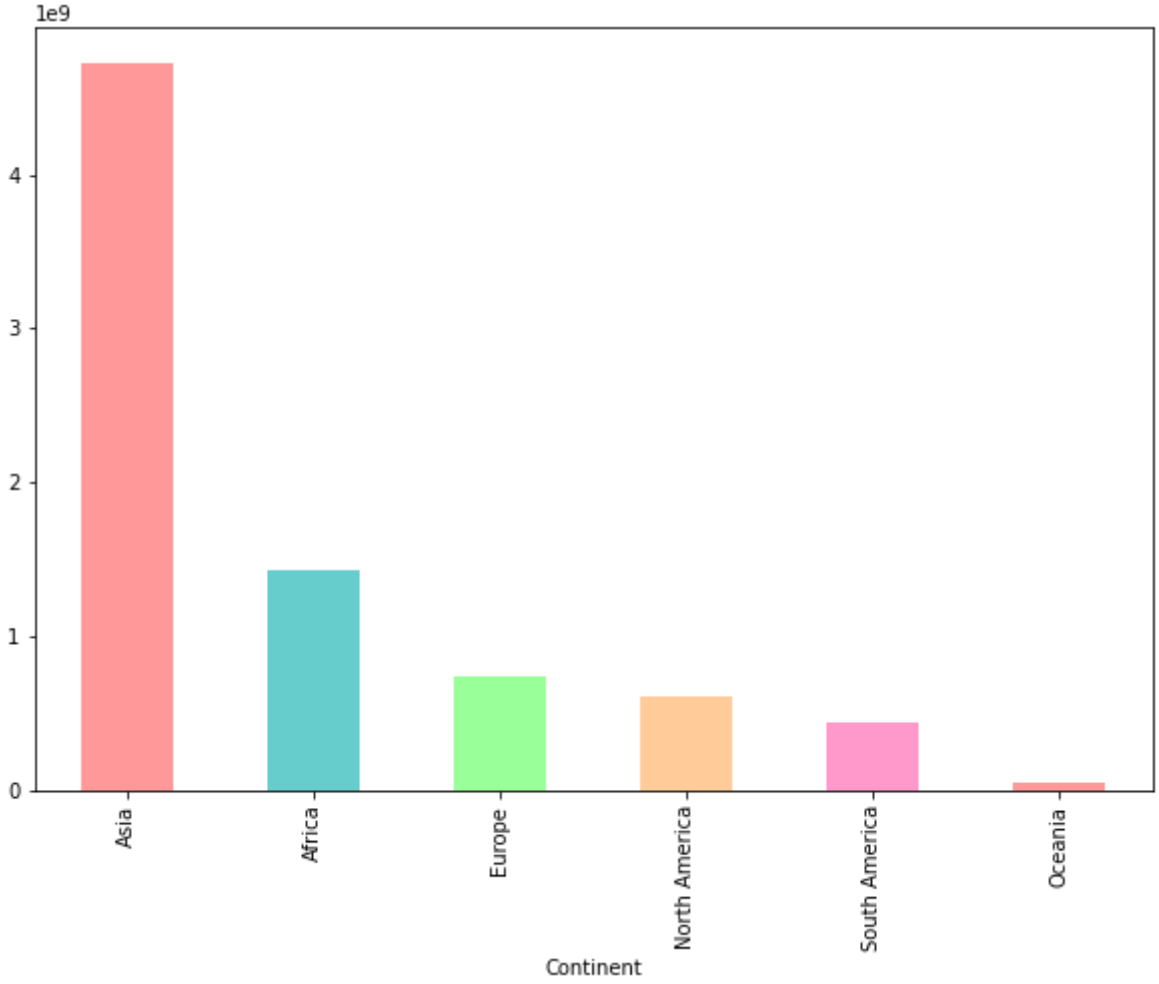
In [13]:
```
#population of least 5 countries 2022
sorted_2022=pop.sort_values(by='2022 Population')
sorted_2022=sorted_2022.head(5)
plt.figure(figsize=(10,5))
plt.bar(sorted_2022['Country/Territory'],sorted_2022['2022 Population'],colo
plt.xlabel('country')
plt.ylabel('population')
plt.title("Population of lowest 5 countries 2022")
plt.show()
```



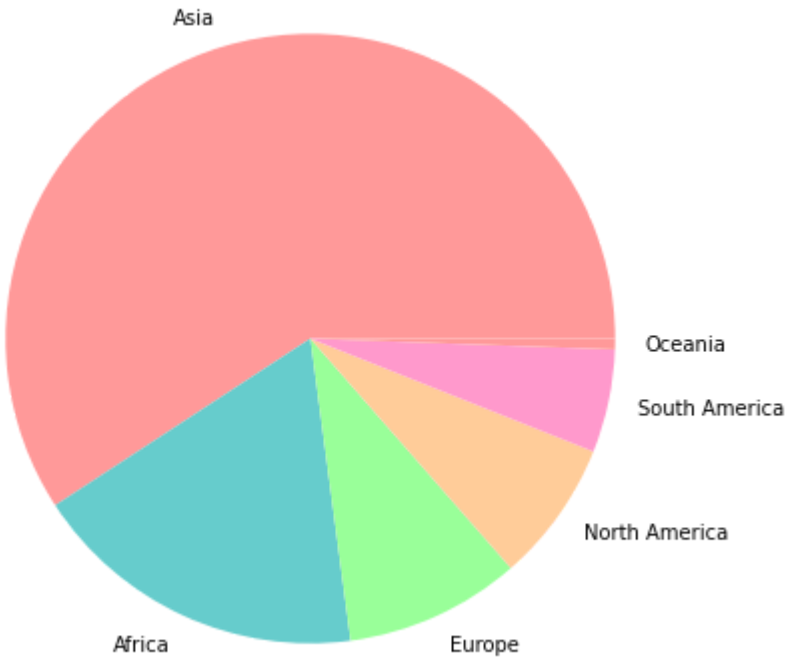
Vatican city has the lowest population in 2022

Continent wise population 2022

In [11]:
```
#continent wise population

# Group the data by continent and calculate the total population for each co
continent_pop =pop.groupby('Continent')['2022 Population'].sum()
continent_pop=continent_pop.sort_values(ascending=False)

plt.figure(figsize=(10,7))
continent_pop.plot(kind='bar',color=pastel_colors)
plt.show()
```



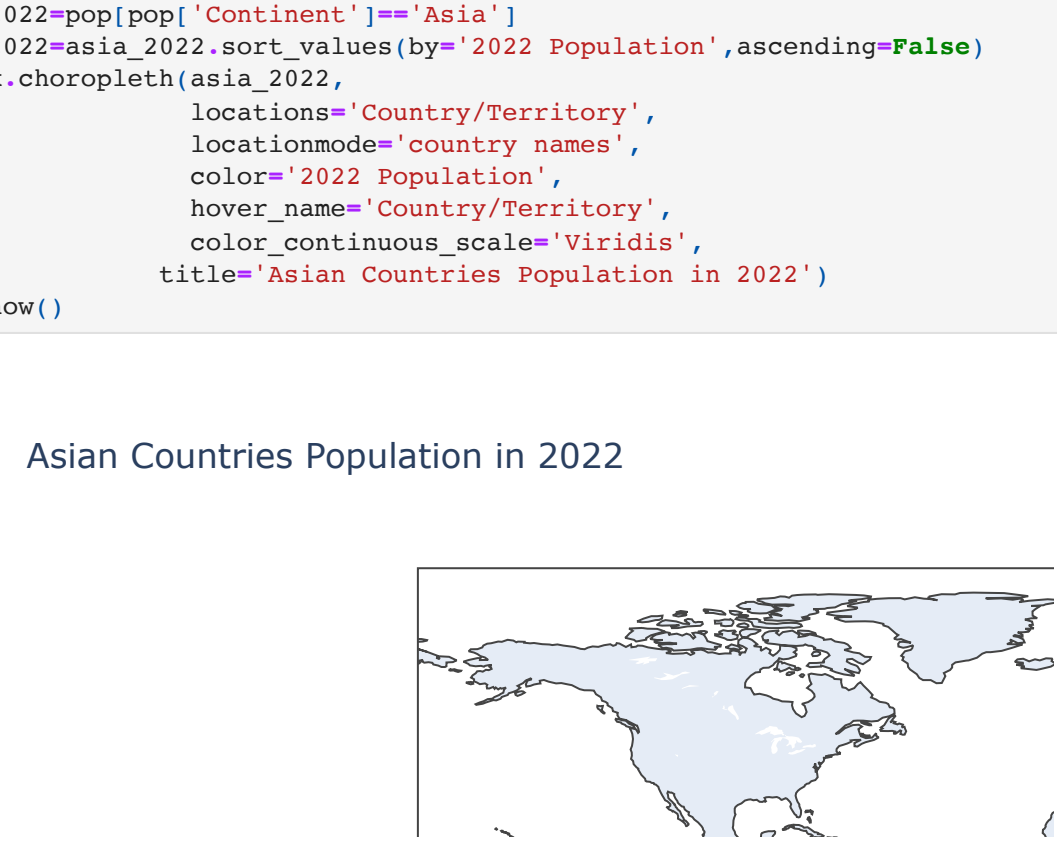
In [12]:
```
#pie chart
plt.figure(figsize=(10,7))
plt.pie(continent_pop,labels=continent_pop.index,colors=pastel_colors)
plt.show()
```



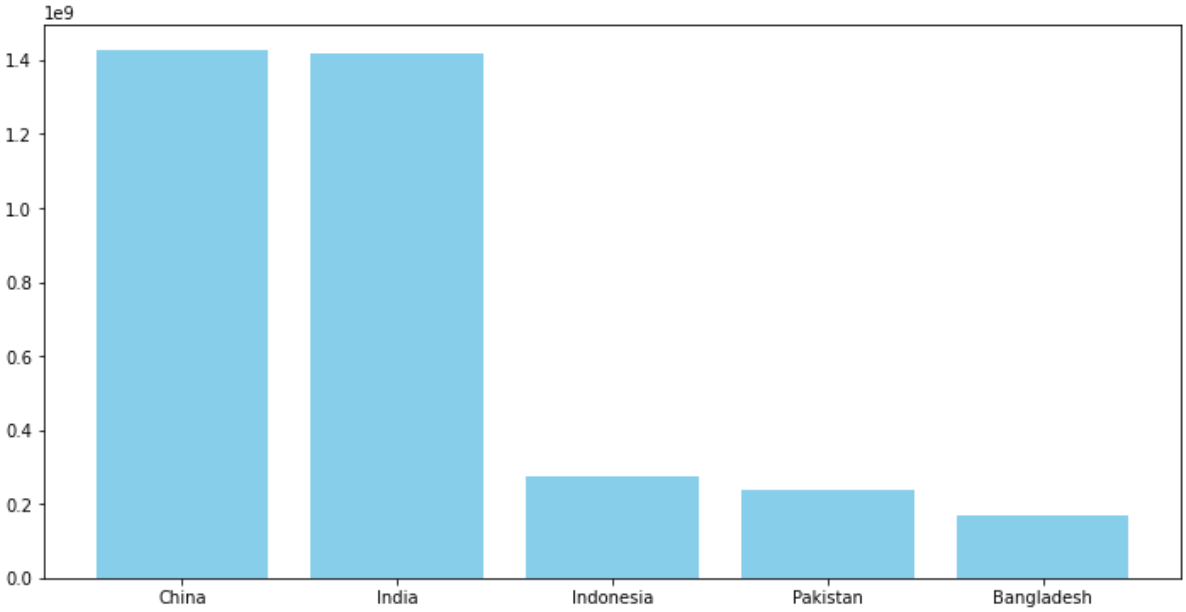
We can see that Asia is the most populated continent and second most populated continent is Africa.

Population in countries of Asia

In [16]:
```
#population in Asia 2022

asia_2022=pop[pop['Continent']=='Asia']
asia_2022=asia_2022.sort_values(by='2022 Population',ascending=False)
fig=px.choropleth(asia_2022,
                  locations='Country/Territory',
                  locationmode='country names',
                  color='2022 Population',
                  hover_name='Country/Territory',
                  color_continuous_scale='Viridis',
                 title='Asian Countries Population in 2022')
fig.show()
```



In [14]:
```
plt.figure(figsize=(12, 6))
plt.bar(asia_2022['Country/Territory'].head(5), asia_2022['2022 Population']
plt.show()
```

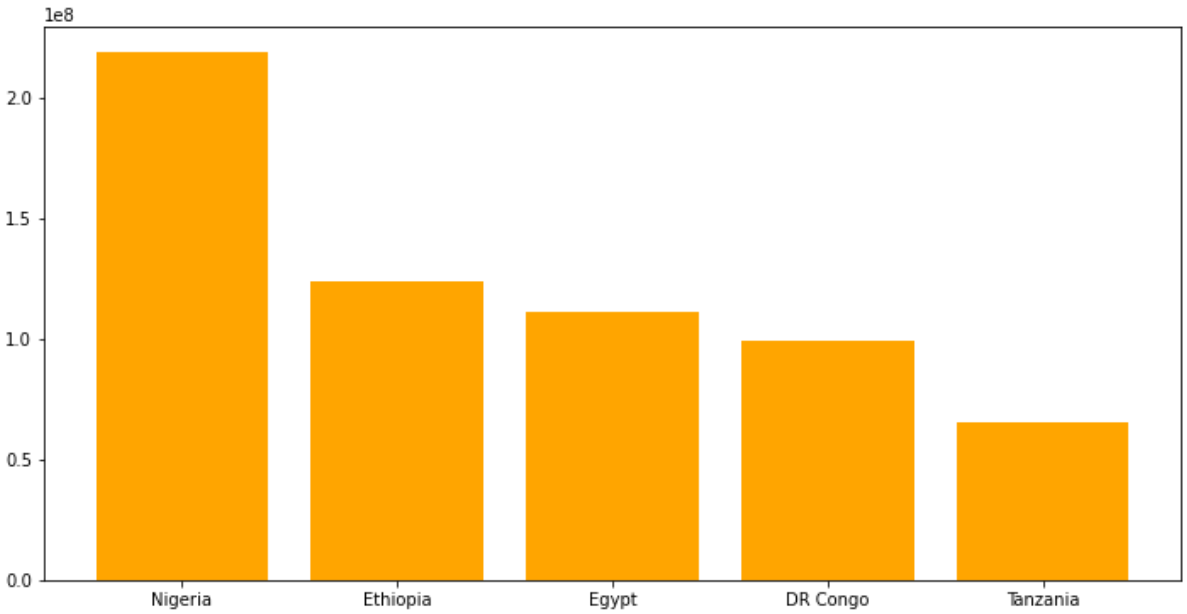


Population in countries of Africa 2022

In [18]:
```
 #population in Africa 2022

africa_2022=pop[pop['Continent']=='Africa']
africa_2022=africa_2022.sort_values(by='2022 Population',ascending=False)
fig=px.choropleth(africa_2022,
                  locations='Country/Territory',
                  locationmode='country names',
                  color='2022 Population',
                  hover_name='Country/Territory',
                  color_continuous_scale='Viridis',
                 title='African Countries Population in 2022')
fig.show()
```

African Countries Population in 2022

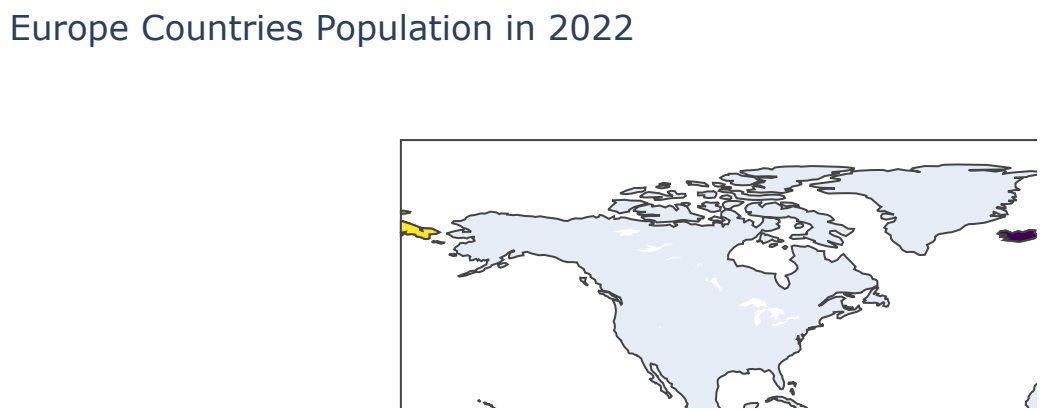
In [19]:
```
plt.figure(figsize=(12, 6))
plt.bar(africa_2022['Country/Territory'].head(5), africa_2022['2022 Populati
plt.show()
```



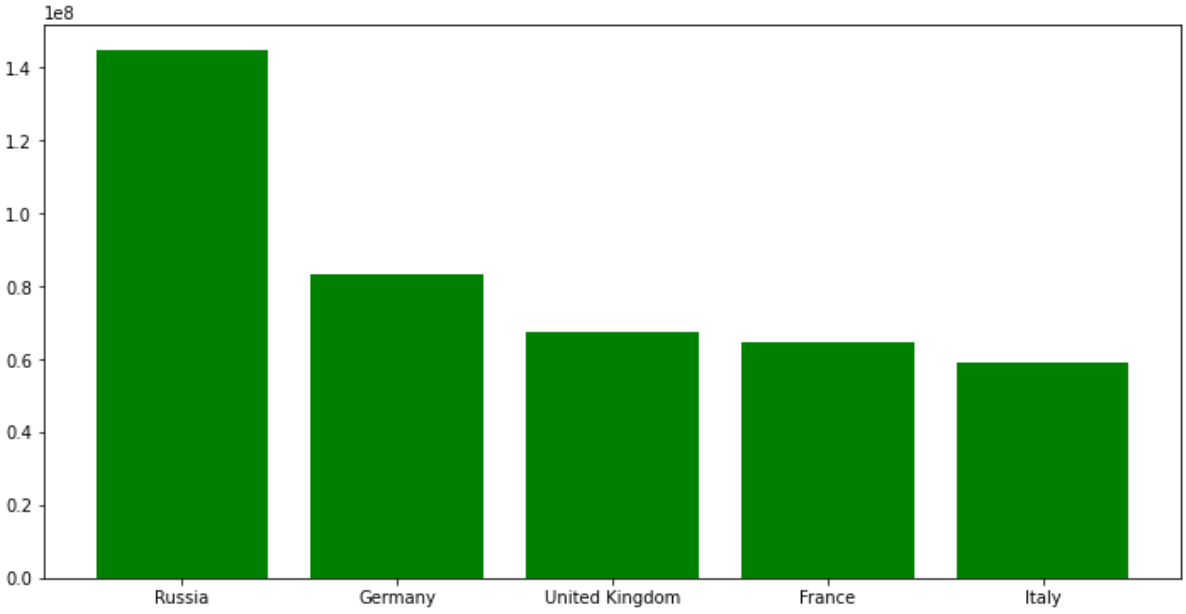
In [20]:
```
### Population in countries of Europe 2022
```

In [21]:
```
#population in Europe 2022

europe_2022=pop[pop['Continent']=='Europe']
europe_2022=europe_2022.sort_values(by='2022 Population',ascending=False)
```

```
fig=px.choropleth(europe_2022,
                  locations='Country/Territory',
                  locationmode='country names',
                  color='2022 Population',
                  hover_name='Country/Territory',
                  color_continuous_scale='Viridis',
                 title='Europe Countries Population in 2022')
fig.show()
```



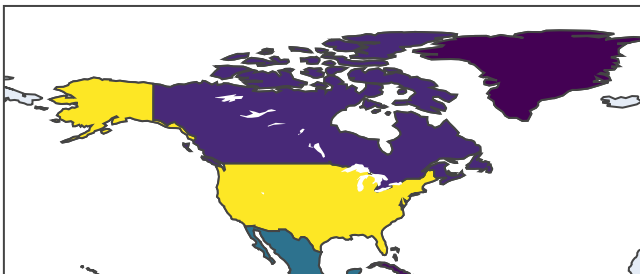
In [22]:
```
plt.figure(figsize=(12, 6))
plt.bar(europe_2022['Country/Territory'].head(5), europe_2022['2022 Populati
plt.show()
```



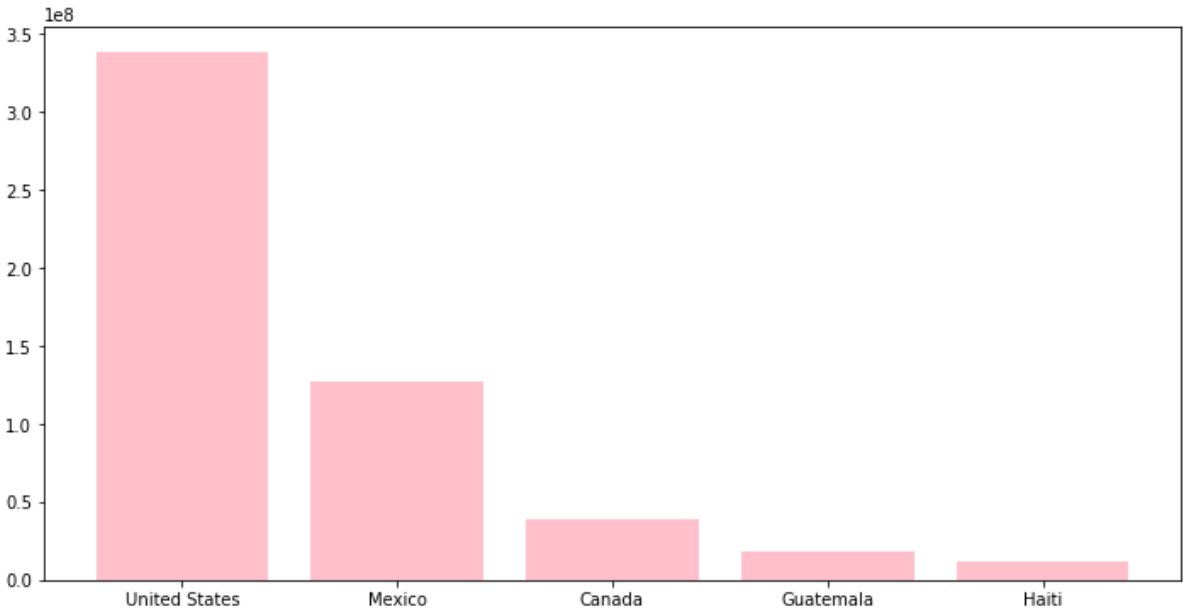
Population in countries of North America 2022

In [23]:
```
#population in North America 2022

na_2022=pop[pop['Continent']=='North America']
na_2022=na_2022.sort_values(by='2022 Population',ascending=False)
fig=px.choropleth(na_2022,
                  locations='Country/Territory',
                  locationmode='country names',
                  color='2022 Population',
                  hover_name='Country/Territory',
                  color_continuous_scale='Viridis',
                 title='North America Countries Population in 2022')
fig.show()
```

North America Countries Population in 2022



In [24]:
```
plt.figure(figsize=(12, 6))
plt.bar(na_2022['Country/Territory'].head(5), na_2022['2022 Population'].hea
plt.show()
```



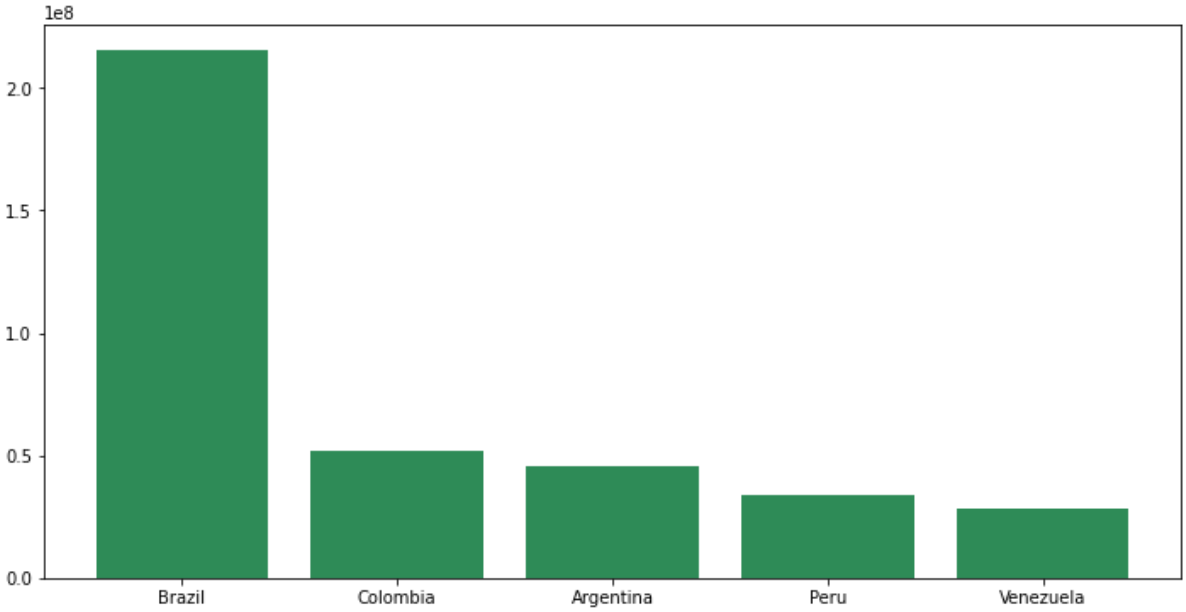
Population in countries of South America 2022

In [27]:
```
#population in South America 2022

sa_2022=pop[pop['Continent']=='South America']
```

```
sa_2022=sa_2022.sort_values(by='2022 Population',ascending=False)
fig=px.choropleth(sa_2022,
                  locations='Country/Territory',
                  locationmode='country names',
                  color='2022 Population',
                  hover_name='Country/Territory',
                  color_continuous_scale='Viridis',
                 title='South America Countries Population in 2022')
fig.show()
```

South America Countries Population in 2022

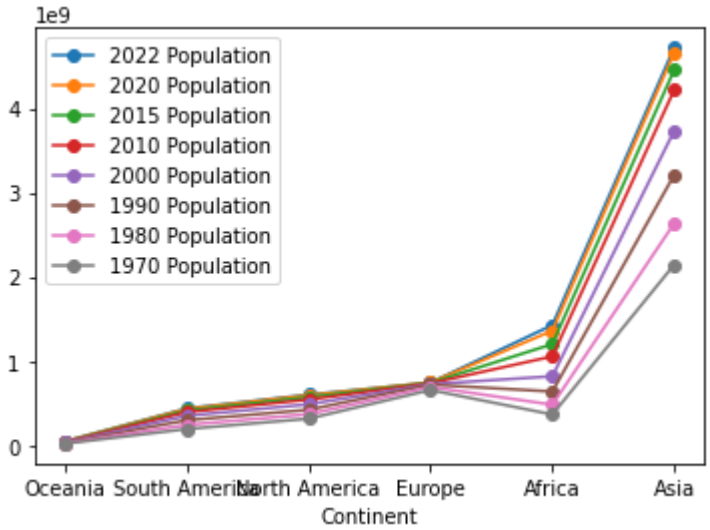
In [28]:
```
plt.figure(figsize=(12, 6))
plt.bar(sa_2022['Country/Territory'].head(5), sa_2022['2022 Population'].hea
plt.show()
```



Population growth in continents over the years (1970-2022)

In [30]:
```
#continent population growth over the years

# Filter data for the desired columns
continent_population = pop.groupby('Continent')[['2022 Population', '2020 Po
continent_population=continent_population.sort_values(by='2022 Population')

plt.figure(figsize=(20,20))

continent_population.plot(marker='o')
plt.show()
```

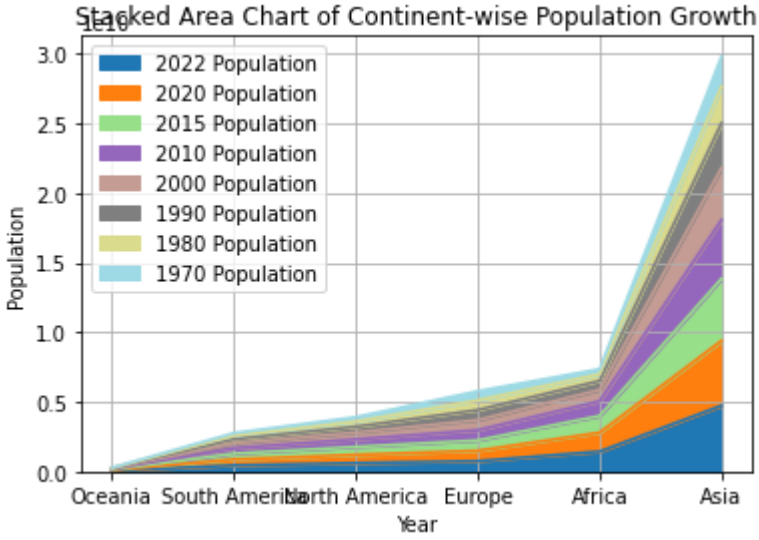
```
<Figure size 1440x1440 with 0 Axes>
```



In [31]:
```
# Set up the figure and axes
plt.figure(figsize=(12, 6))

# Create the stacked area plot
continent_population.plot(kind='area', stacked=True, colormap='tab20')

# Add labels and title
plt.xlabel('Year')
plt.ylabel('Population')
plt.title('Stacked Area Chart of Continent-wise Population Growth')
```

```
# Show the plot
plt.grid(True)
plt.show()
```

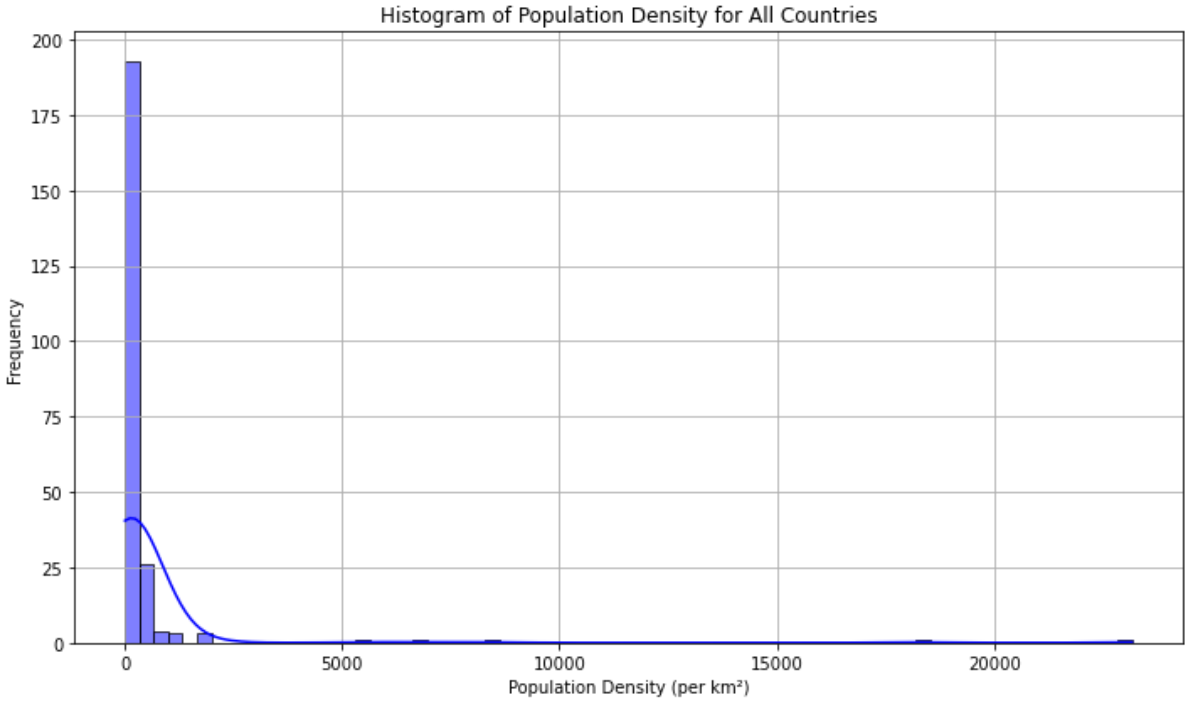
```
<Figure size 864x432 with 0 Axes>
```



We can see population growth in each continent year-wise ,and also that Asia's population is more rapid compared to other continents.

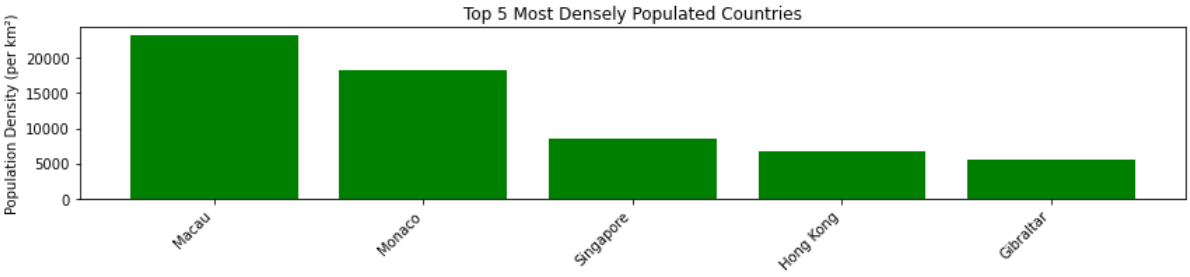
Population Density Distribution

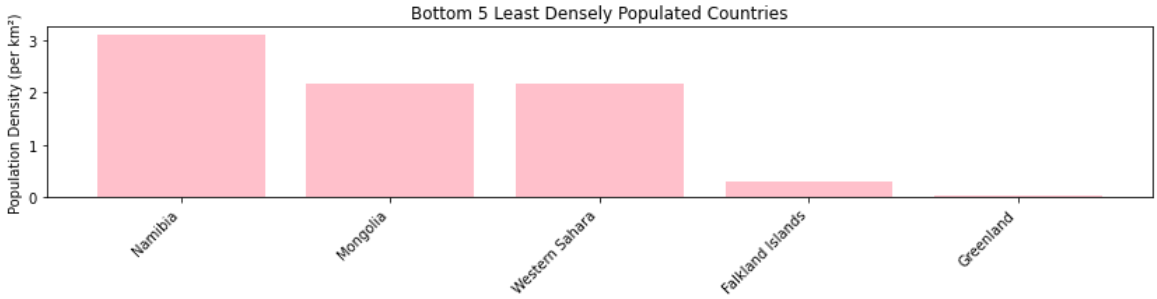
In [34]:
```
#population density

plt.figure(figsize=(10, 6))

# Create the histogram plot
sns.histplot(data=pop, x='Density (per km²)', bins=70,kde=True, color='blue'

# Add labels and title
plt.xlabel('Population Density (per km²)')
plt.ylabel('Frequency')
plt.title('Histogram of Population Density for All Countries')

# Show the plot
plt.grid(True)
plt.tight_layout()
plt.show()
```



population density for almost all the countries is between 0 and 2000 approximately.

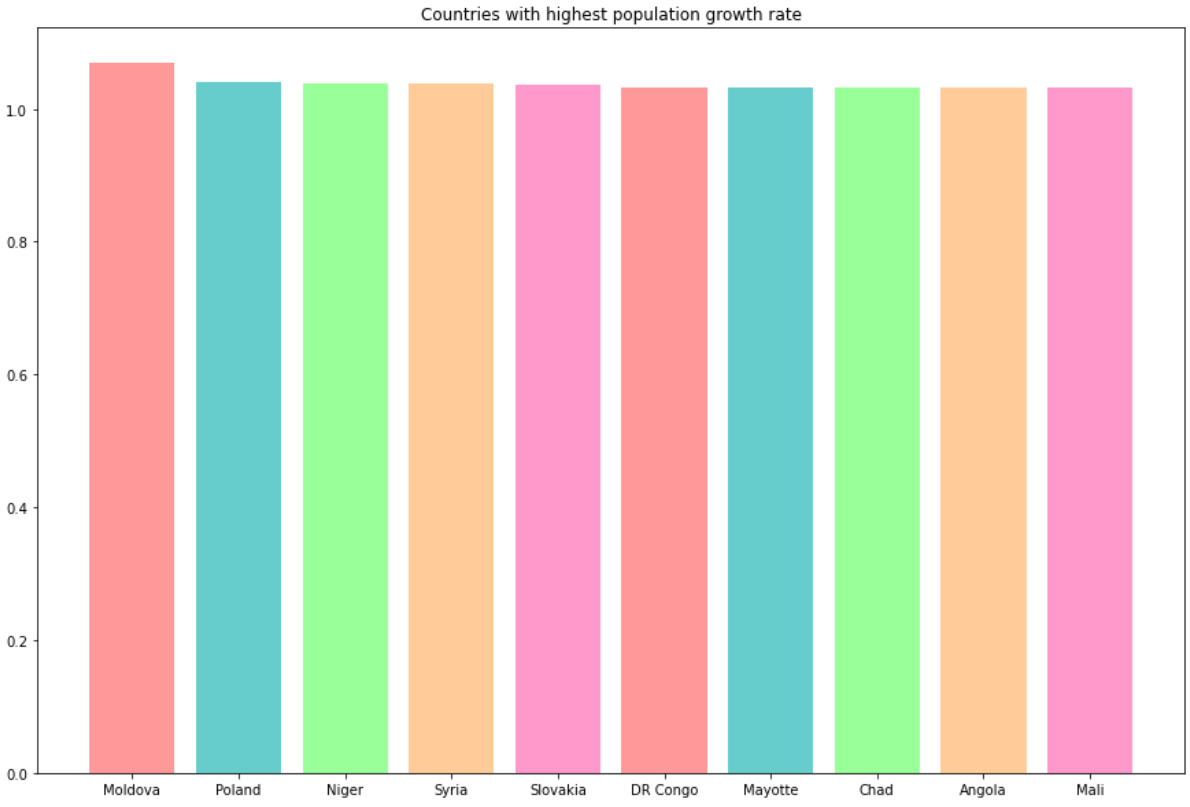
In [55]:
```
pop_sorted = pop.sort_values(by='Density (per km²)', ascending=False)

# Select the top 5 most densely populated countries
top_5_densely_populated = pop_sorted.head(5)

# Select the bottom 5 least densely populated countries
bottom_5_densely_populated = pop_sorted.tail(5)

plt.figure(figsize=(12, 6))
plt.subplot(2, 1, 1)

# Create the bar plot for top 5 most densely populated countries
plt.bar(top_5_densely_populated['Country/Territory'], top_5_densely_populate
plt.xlabel('')
plt.ylabel('Population Density (per km²)')
plt.title('Top 5 Most Densely Populated Countries')
plt.xticks(rotation=45, ha='right')

# Set up the figure and axes for the bottom plot
plt.subplot(2, 1, 2)

# Create the bar plot for bottom 5 least densely populated countries
plt.bar(bottom_5_densely_populated['Country/Territory'], bottom_5_densely_po
plt.xlabel('')
plt.ylabel('Population Density (per km²)')
plt.title('Bottom 5 Least Densely Populated Countries')
plt.xticks(rotation=45, ha='right')

# Adjust layout
plt.tight_layout()

# Show the plots
plt.show()
```





Countries with highest population growth rate

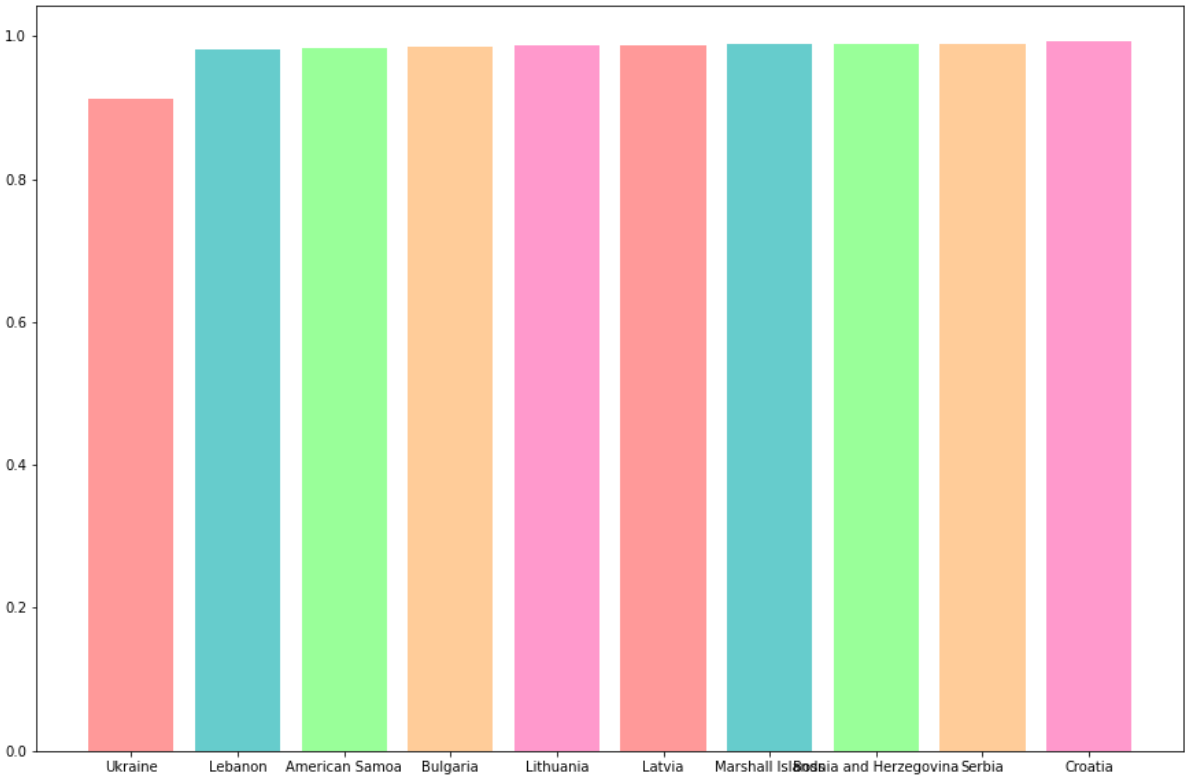
In [35]:
```
#Top 10 countries in population growth rate

growth_rate=pop.sort_values(by='Growth Rate')

top_growth_rate=growth_rate.tail(10)
top_growth_rate=top_growth_rate.sort_values(by='Growth Rate',ascending=False
plt.figure(figsize=(15,10))
plt.bar(top_growth_rate['Country/Territory'],top_growth_rate['Growth Rate'],
plt.title("Countries with highest population growth rate")
plt.show()
```



Countries with least population growth rate

In [64]:
```
#least 10 countriest in population growth rate
```

```
growth_rate=pop.sort_values(by='Growth Rate')

low_growth_rate=growth_rate.head(10)
plt.figure(figsize=(15,10))
plt.bar(low_growth_rate['Country/Territory'],low_growth_rate['Growth Rate'],
plt.show()
```

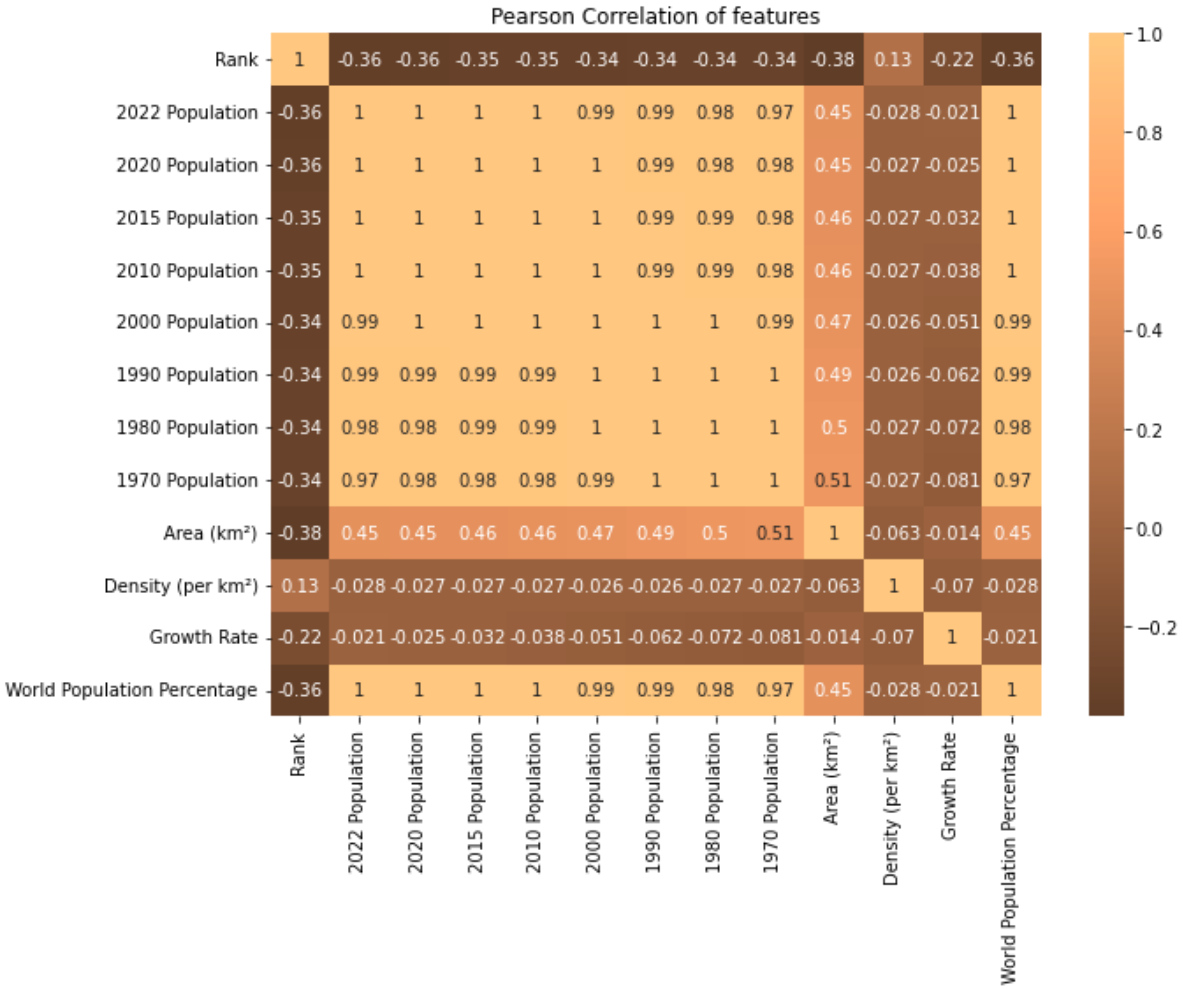


In [36]:
```
fig = px.choropleth(pop,
                    locations='Country/Territory',
                    locationmode='country names',
                    color='Growth Rate',
                    color_continuous_scale='Viridis',  # You can change the
                    title='World Population Growth Rate (2022-2010) by Count
                    hover_name='Country/Territory')

# Show the plot
fig.show()
```

World Population Growth Rate (2022-2010) by Country

Correlation between features

In [37]:
```
# Calculate the correlation matrix
correlation_matrix = pop.corr()

# Set up the figure and axes
plt.figure(figsize=(10, 8))

# Create the heatmap
sns.heatmap(correlation_matrix, annot=True, cmap='copper', center=0)

# Add title
plt.title('Pearson Correlation of features')

# Show the plot
plt.tight_layout()
plt.show()
```



1. There is a strong positive correlation between a country's population and its world population percentage. Similarly, there's a high positive correlation between a country's name and its CCA3 code, which is expected.

2. A moderate positive correlation exists between a country's area and its population, as well as between its area and its world population percentage.

3. There is a moderate negative correlation between a country's growth rate and its continent. Additionally, moderate negative correlations are observed between a country's rank and both its world population percentage, its area, and its population.

In [ ]: