

```
In [14]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
```

```
In [15]: df = pd.read_csv('C:/Users/DELL/Downloads/worldpopulationdata (1).csv')
```

```
In [16]: df.head()
```

Out[16]:

	Series Name	Series Code	Country Name	Country Code	2022	2021	2020	2019	2018	2017	...
0	Population, total	SP.POP.TOTL	Afghanistan	AFG	41128771.0	40099462.0	38972230.0	37769499.0	36686784.0	35643418.0	...
1	Population, total	SP.POP.TOTL	Albania	ALB	2775634.0	2811666.0	2837849.0	2854191.0	2866376.0	2873457.0	...
2	Population, total	SP.POP.TOTL	Algeria	DZA	44903225.0	44177969.0	43451666.0	42705368.0	41927007.0	41136546.0	...
3	Population, total	SP.POP.TOTL	American Samoa	ASM	44273.0	45035.0	46189.0	47321.0	48424.0	49463.0	...
4	Population, total	SP.POP.TOTL	Andorra	AND	79824.0	79034.0	77700.0	76343.0	75013.0	73837.0	...

5 rows × 26 columns

```
In [17]: df.tail()
```

Out[17]:

	Series Name	Series Code	Country Name	Country Code	2022	2021	2020	2019	2018	2017	...
1080	Population, male (% of total population)	SP.POP.TOTL.MA.ZS	Virgin Islands (U.S.)	VIR	46.613382	46.764444	46.914637	47.057307	47.185912	47.314214	...
1081	Population, male (% of total population)	SP.POP.TOTL.MA.ZS	West Bank and Gaza	PSE	49.893678	49.877839	49.858957	49.835542	49.811374	49.785969	...
1082	Population, male (% of total population)	SP.POP.TOTL.MA.ZS	Yemen, Rep.	YEM	50.519031	50.538516	50.554317	50.571320	50.596614	50.616964	...
1083	Population, male (% of total population)	SP.POP.TOTL.MA.ZS	Zambia	ZMB	49.344602	49.344951	49.338301	49.326233	49.309087	49.288400	...
1084	Population, male (% of total population)	SP.POP.TOTL.MA.ZS	Zimbabwe	ZWE	47.214139	47.167153	47.130679	47.099796	47.076238	47.051613	...

5 rows × 26 columns

```
In [18]: df.columns
```

```
Out[18]: Index(['Series Name', 'Series Code', 'Country Name', 'Country Code', '2022',
              '2021', '2020', '2019', '2018', '2017', '2016', '2015', '2014', '2013',
              '2012', '2011', '2010', '2009', '2008', '2007', '2006', '2005', '2004',
              '2003', '2002', '2001'],
              dtype='object')
```

```
In [19]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1085 entries, 0 to 1084
Data columns (total 26 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Series Name     1085 non-null   object
1   Series Code     1085 non-null   object
2   Country Name    1085 non-null   object
3   Country Code    1085 non-null   object
4   2022            1085 non-null   float64
5   2021            1085 non-null   float64
6   2020            1085 non-null   float64
7   2019            1085 non-null   float64
8   2018            1085 non-null   float64
9   2017            1085 non-null   float64
10  2016            1085 non-null   float64
11  2015            1085 non-null   float64
12  2014            1085 non-null   float64
13  2013            1085 non-null   float64
14  2012            1085 non-null   float64
15  2011            1085 non-null   float64
16  2010            1085 non-null   float64
17  2009            1085 non-null   float64
18  2008            1085 non-null   float64
19  2007            1085 non-null   float64
20  2006            1085 non-null   float64
21  2005            1085 non-null   float64
22  2004            1085 non-null   float64
23  2003            1085 non-null   float64
24  2002            1085 non-null   float64
25  2001            1085 non-null   float64
dtypes: float64(22), object(4)
memory usage: 220.5+ KB
```

```
In [20]: df.duplicated().sum()
```

```
Out[20]: 0
```

```
In [21]: print(df['Country Name'].unique())
print("\nTotal no of unique countries:",df['Country Name'].nunique())
```

```
['Afghanistan' 'Albania' 'Algeria' 'American Samoa' 'Andorra' 'Angola'
'Antigua and Barbuda' 'Argentina' 'Armenia' 'Aruba' 'Australia' 'Austria'
'Azerbaijan' 'Bahamas, The' 'Bahrain' 'Bangladesh' 'Barbados' 'Belarus'
'Belgium' 'Belize' 'Benin' 'Bermuda' 'Bhutan' 'Bolivia'
'Bosnia and Herzegovina' 'Botswana' 'Brazil' 'British Virgin Islands'
'Brunei Darussalam' 'Bulgaria' 'Burkina Faso' 'Burundi' 'Cabo Verde'
'Cambodia' 'Cameroon' 'Canada' 'Cayman Islands'
'Central African Republic' 'Chad' 'Channel Islands' 'Chile' 'China'
'Colombia' 'Comoros' 'Congo, Dem. Rep.' 'Congo, Rep.' 'Costa Rica'
'Cote d'Ivoire' 'Croatia' 'Cuba' 'Curacao' 'Cyprus' 'Czechia' 'Denmark'
'Djibouti' 'Dominica' 'Dominican Republic' 'Ecuador' 'Egypt, Arab Rep.'
'El Salvador' 'Equatorial Guinea' 'Eritrea' 'Estonia' 'Eswatini'
'Ethiopia' 'Faroe Islands' 'Fiji' 'Finland' 'France' 'French Polynesia'
'Gabon' 'Gambia, The' 'Georgia' 'Germany' 'Ghana' 'Gibraltar' 'Greece'
'Greenland' 'Grenada' 'Guam' 'Guatemala' 'Guinea' 'Guinea-Bissau'
'Guyana' 'Haiti' 'Honduras' 'Hong Kong SAR, China' 'Hungary' 'Iceland'
'India' 'Indonesia' 'Iran, Islamic Rep.' 'Iraq' 'Ireland' 'Isle of Man'
'Israel' 'Italy' 'Jamaica' 'Japan' 'Jordan' 'Kazakhstan' 'Kenya'
'Kiribati' 'Korea, Dem. People's Rep.' 'Korea, Rep.' 'Kosovo' 'Kuwait'
'Kyrgyz Republic' 'Lao PDR' 'Latvia' 'Lebanon' 'Lesotho' 'Liberia'
'Libya' 'Liechtenstein' 'Lithuania' 'Luxembourg' 'Macao SAR, China'
'Madagascar' 'Malawi' 'Malaysia' 'Maldives' 'Mali' 'Malta'
'Marshall Islands' 'Mauritania' 'Mauritius' 'Mexico'
'Micronesia, Fed. Sts.' 'Moldova' 'Monaco' 'Mongolia' 'Montenegro'
'Morocco' 'Mozambique' 'Myanmar' 'Namibia' 'Nauru' 'Nepal' 'Netherlands'
'New Caledonia' 'New Zealand' 'Nicaragua' 'Niger' 'Nigeria'
'North Macedonia' 'Northern Mariana Islands' 'Norway' 'Oman' 'Pakistan'
'Palau' 'Panama' 'Papua New Guinea' 'Paraguay' 'Peru' 'Philippines'
'Poland' 'Portugal' 'Puerto Rico' 'Qatar' 'Romania' 'Russian Federation'
'Rwanda' 'Samoa' 'San Marino' 'Sao Tome and Principe' 'Saudi Arabia'
'Senegal' 'Serbia' 'Seychelles' 'Sierra Leone' 'Singapore'
'Sint Maarten (Dutch part)' 'Slovak Republic' 'Slovenia'
'Solomon Islands' 'Somalia' 'South Africa' 'South Sudan' 'Spain'
'Sri Lanka' 'St. Kitts and Nevis' 'St. Lucia' 'St. Martin (French part)'
'St. Vincent and the Grenadines' 'Sudan' 'Suriname' 'Sweden'
'Switzerland' 'Syrian Arab Republic' 'Tajikistan' 'Tanzania' 'Thailand'
'Timor-Leste' 'Togo' 'Tonga' 'Trinidad and Tobago' 'Tunisia' 'Turkiye'
'Turkmenistan' 'Turks and Caicos Islands' 'Tuvalu' 'Uganda' 'Ukraine'
'United Arab Emirates' 'United Kingdom' 'United States' 'Uruguay'
'Uzbekistan' 'Vanuatu' 'Venezuela, RB' 'Vietnam' 'Virgin Islands (U.S.)'
'West Bank and Gaza' 'Yemen, Rep.' 'Zambia' 'Zimbabwe']
```

```
Total no of unique countries: 217
```

```
In [22]: #Dropping unnecessary columns
df.drop(['Series Name', 'Country Code'], axis=1, inplace=True)
```

```
In [23]: df.columns
```

```
Out[23]: Index(['Series Code', 'Country Name', '2022', '2021', '2020', '2019', '2018',
              '2017', '2016', '2015', '2014', '2013', '2012', '2011', '2010', '2009',
              '2008', '2007', '2006', '2005', '2004', '2003', '2002', '2001'],
              dtype='object')
```

```
In [29]: #Extracting top ten countries with respect to total population
```

```
total_population_data = df[df['Series Code'] == 'SP.POP.TOTL']
# Sort data based on the total population for 2022
total_population_sorted = total_population_data.sort_values(by="2022", ascending=False)
# Get the top ten countries with the highest total population for 2022
total_top_ten_countries_of_2022 = total_population_sorted.head(10)
print("Top ten countries of total population\n")
print(total_top_ten_countries[['Country Name']])
```

```
#Extracting top ten countries with respect to total population
```

```
total_population_data = df[df['Series Code'] == 'SP.POP.TOTL']
# Sort data based on the total population for 2020
total_population_sorted = total_population_data.sort_values(by="2020", ascending=False)
# Get the top ten countries with the highest total population for 2020
total_top_ten_countries_of_2020 = total_population_sorted.head(10)
print("Top ten countries of total population\n")
print(total_top_ten_countries[['Country Name']])
```

Top ten countries of total population

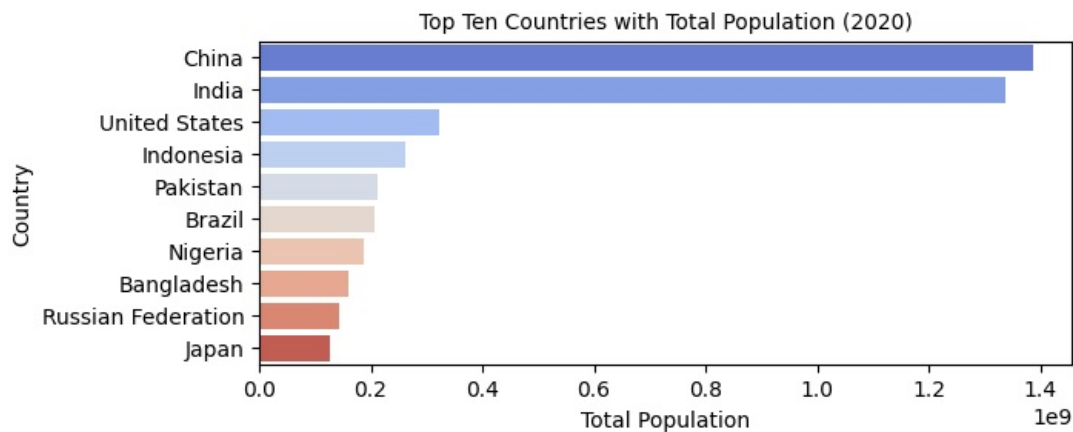
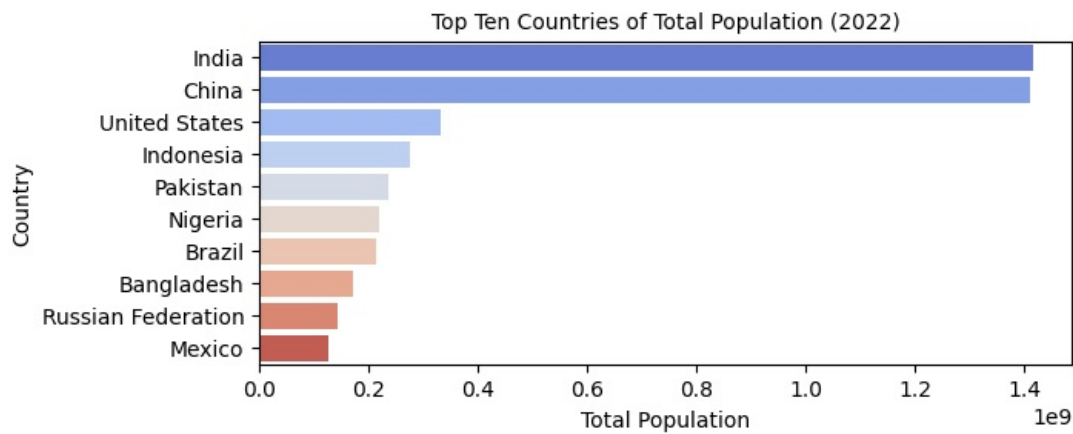
	Country Name
41	China
89	India
206	United States
90	Indonesia
149	Pakistan
26	Brazil
144	Nigeria
15	Bangladesh
161	Russian Federation
98	Japan

Top ten countries of total population

	Country Name
41	China
89	India
206	United States
90	Indonesia
149	Pakistan
26	Brazil
144	Nigeria
15	Bangladesh
161	Russian Federation
98	Japan

```
In [30]: # Create the bar plot
plt.figure(figsize=(15, 6))
plt.subplot(2,2,1)
sns.barplot(x="2022", y="Country Name", data=total_top_ten_countries_of_2022, palette="coolwarm")
plt.title("Top Ten Countries of Total Population (2022)", fontsize=10)
plt.xlabel("Total Population", fontsize=10)
plt.ylabel("Country", fontsize=10)
plt.show()

plt.figure(figsize=(15, 6))
plt.subplot(2,2,2)
sns.barplot(x="2020", y="Country Name", data=total_top_ten_countries_of_2020, palette="coolwarm")
plt.title("Top Ten Countries with Total Population (2020)", fontsize=10)
plt.xlabel("Total Population", fontsize=10)
plt.ylabel("Country", fontsize=10)
plt.show()
```



```
In [31]: #Extraction of bottom-10 countries with respect to total population
```

```
In [36]: #for 2022
# Sort data based on the total population for 2022
total_population_sorted1 = total_population_data.sort_values(by="2022", ascending=True)

# Get the bottom ten countries with the highest total population for 2022
total_bottom_ten_countries_of_2022 = total_population_sorted1.head(10)
print("Bottom ten countries of total population\n")
print(total_bottom_ten_countries[['Country Name']])
```

Bottom ten countries of total population

	Country Name
201	Tuvalu
137	Nauru
150	Palau
27	British Virgin Islands
183	St. Martin (French part)
75	Gibraltar
164	San Marino
130	Monaco
114	Liechtenstein
124	Marshall Islands

```
In [37]: #for 2020
# Sort data based on the total population for 2008
total_population_sorted1 = total_population_data.sort_values(by="2008", ascending=True)

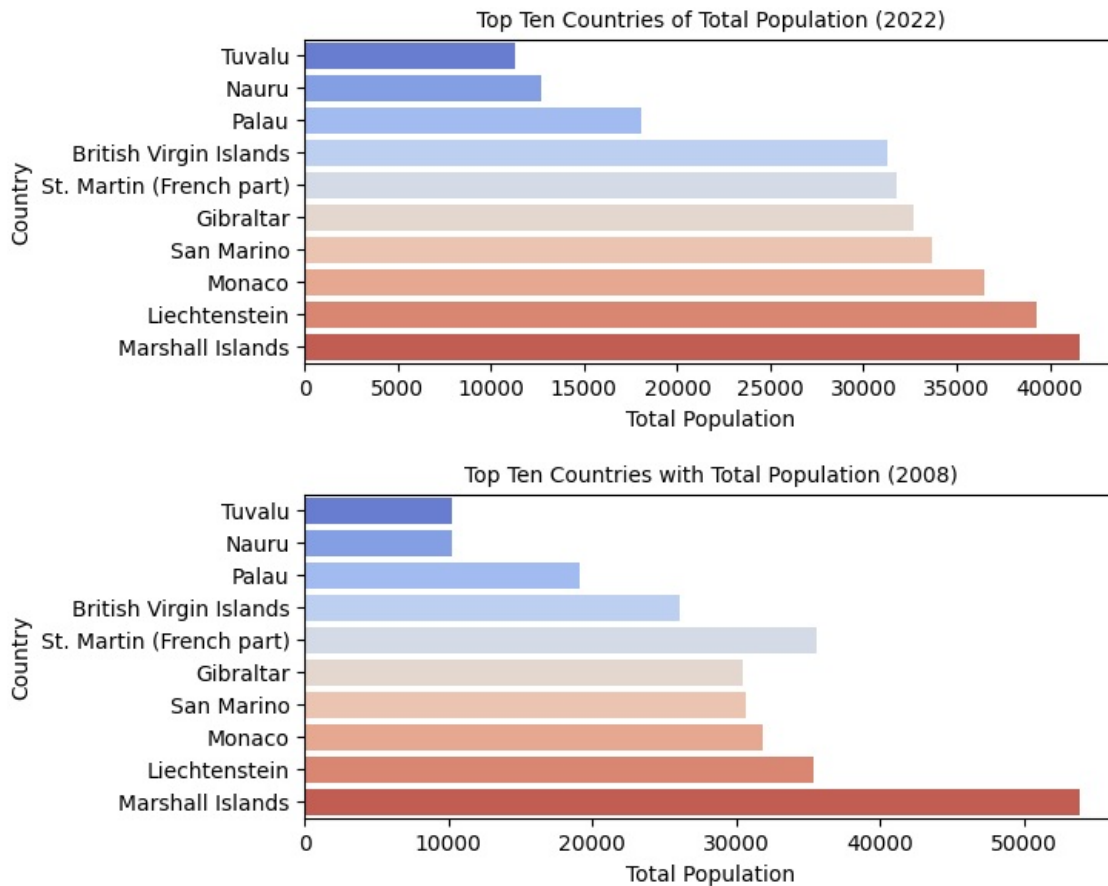
# Get the bottom ten countries with the highest total population for 2008
total_bottom_ten_countrie_s = total_population_sorted1.head(10)
print("Bottom ten countries of total population\n")
print(total_bottom_ten_countries[['Country Name']])
```

Bottom ten countries of total population

	Country Name
201	Tuvalu
137	Nauru
150	Palau
27	British Virgin Islands
183	St. Martin (French part)
75	Gibraltar
164	San Marino
130	Monaco
114	Liechtenstein
124	Marshall Islands

```
In [39]: # Create the bar plot
plt.figure(figsize=(15, 6))
plt.subplot(2,2,1)
sns.barplot(x="2022", y="Country Name", data=total_bottom_ten_countries_of_2022, palette="coolwarm")
plt.title("Top Ten Countries of Total Population (2022)", fontsize=10)
plt.xlabel("Total Population", fontsize=10)
plt.ylabel("Country", fontsize=10)
plt.show()

plt.figure(figsize=(15, 6))
plt.subplot(2,2,2)
sns.barplot(x="2008", y="Country Name", data=total_bottom_ten_countries, palette="coolwarm")
plt.title("Top Ten Countries with Total Population (2008)", fontsize=10)
plt.xlabel("Total Population", fontsize=10)
plt.ylabel("Country", fontsize=10)
plt.show()
```



```
In [40]: #Extraction of top ten countries with highest male population
```

```
In [56]: #Top 10 Countries with Male and Female Populations (2022)
```

```
In [ ]:
```

```
In [41]: # Filter data for male population
male_population_data = df[df["Series Code"] == "SP.POP.TOTL.MA.IN"]

# Sort data based on the male population for 2022
male_population_sorted = male_population_data.sort_values(by="2022", ascending=False)

# Get the top ten countries with the highest male population for 2022
male_top_ten_countries = male_population_sorted.head(10)
print("Top ten countries of male population")
print(male_top_ten_countries[['Country Name']])
```

Top ten countries of male population

	Country Name
523	India
475	China
640	United States
524	Indonesia
583	Pakistan
578	Nigeria
460	Brazil
449	Bangladesh
595	Russian Federation
561	Mexico

```
In [43]: # Filter data for female population
female_population_data = df[df["Series Code"] == "SP.POP.TOTL.FE.IN"]

# Sort data based on the female population for 2022
female_population_sorted = female_population_data.sort_values(by="2022", ascending=False)

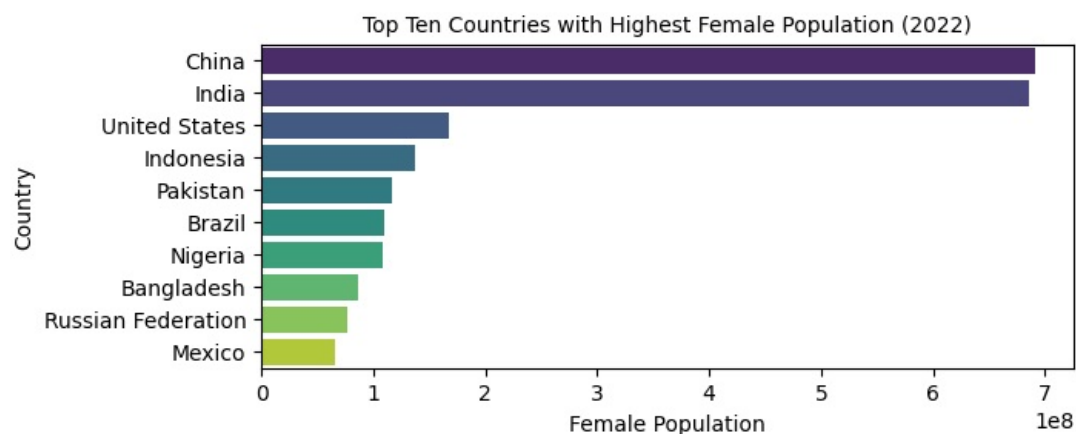
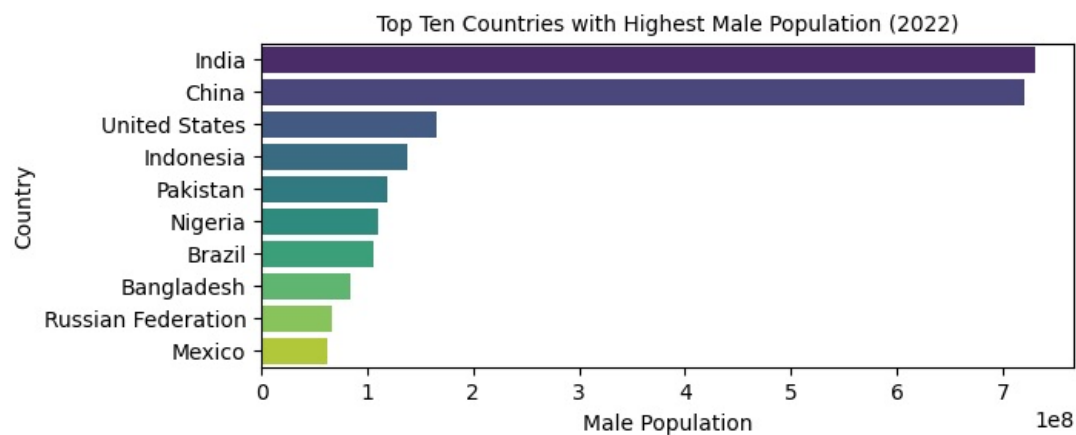
# Get the top ten countries with the highest female population for 2022
female_top_ten_countries = female_population_sorted.head(10)
print("Top ten countries of female population")
print(female_top_ten_countries[['Country Name']])
```

Top ten countries of female population

	Country Name
258	China
306	India
423	United States
307	Indonesia
366	Pakistan
243	Brazil
361	Nigeria
232	Bangladesh
378	Russian Federation
344	Mexico

```
In [45]: # Create the bar plot
plt.figure(figsize=(15, 6))
plt.subplot(2,2,1)
sns.barplot(x="2022", y="Country Name", data=male_top_ten_countries, palette="viridis")
plt.title("Top Ten Countries with Highest Male Population (2022)",size=10)
plt.xlabel("Male Population",size=10)
plt.ylabel("Country",size=10)
plt.show()

plt.figure(figsize=(15, 6))
plt.subplot(2,2,2)
sns.barplot(x="2022", y="Country Name", data=female_top_ten_countries, palette="viridis")
plt.title("Top Ten Countries with Highest Female Population (2022)",size=10)
plt.xlabel("Female Population",size=10)
plt.ylabel("Country",size=10)
plt.show()
```

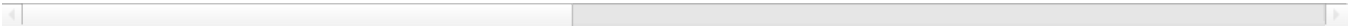


```
In [52]: #Top 10 Countries with Male and Female Populations (2022)
# Merge male and female population data on 'Country Name'
merged_data = pd.merge(male_population_data, female_population_data, on="Country Name", suffixes=("_male", "_female"))
```

Out[52]:

	Series Code_male	Country Name	2022_male	2021_male	2020_male	2019_male	2018_male	2017_male	2016_male	2015_r
0	SP.POP.TOTL.MA.IN	Afghanistan	20766442.0	20254878.0	19692301.0	19090409.0	18549862.0	18028696.0	17520861.0	170714
1	SP.POP.TOTL.MA.IN	Albania	1384548.0	1404454.0	1419264.0	1428828.0	1435881.0	1440219.0	1442176.0	14448
2	SP.POP.TOTL.MA.IN	Algeria	22862237.0	22497244.0	22132899.0	21756903.0	21362603.0	20961313.0	20556314.0	201522
3	SP.POP.TOTL.MA.IN	American Samoa	21873.0	22289.0	22921.0	23535.0	24134.0	24701.0	25240.0	257
4	SP.POP.TOTL.MA.IN	Andorra	40786.0	40361.0	39615.0	38842.0	38071.0	37380.0	36628.0	361
...	...	...	...	...	...	...	...	...	...	...
212	SP.POP.TOTL.MA.IN	Virgin Islands (U.S.)	49137.0	49510.0	49866.0	50196.0	50489.0	50759.0	50999.0	512
213	SP.POP.TOTL.MA.IN	West Bank and Gaza	2516444.0	2455361.0	2394860.0	2334948.0	2275925.0	2217868.0	2173706.0	21256
214	SP.POP.TOTL.MA.IN	Yemen, Rep.	17023203.0	16668432.0	16320979.0	15953578.0	15578957.0	15202496.0	14820156.0	144391
215	SP.POP.TOTL.MA.IN	Zambia	9877642.0	9609004.0	9338613.0	9066397.0	8794716.0	8525934.0	8260471.0	80003
216	SP.POP.TOTL.MA.IN	Zimbabwe	7705601.0	7543690.0	7385220.0	7231989.0	7086002.0	6940631.0	6796658.0	66528

217 rows × 47 columns



In [53]:

```
# Calculate the total population for each country (male + female)
merged_data["Total Population"] = merged_data["2022_male"] + merged_data["2022_female"]
```

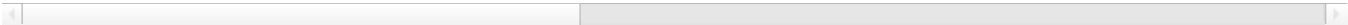
In [54]:

```
merged_data.head()
```

Out[54]:

	Series Code_male	Country Name	2022_male	2021_male	2020_male	2019_male	2018_male	2017_male	2016_male	2015_ma
0	SP.POP.TOTL.MA.IN	Afghanistan	20766442.0	20254878.0	19692301.0	19090409.0	18549862.0	18028696.0	17520861.0	17071446
1	SP.POP.TOTL.MA.IN	Albania	1384548.0	1404454.0	1419264.0	1428828.0	1435881.0	1440219.0	1442176.0	1444890
2	SP.POP.TOTL.MA.IN	Algeria	22862237.0	22497244.0	22132899.0	21756903.0	21362603.0	20961313.0	20556314.0	20152232
3	SP.POP.TOTL.MA.IN	American Samoa	21873.0	22289.0	22921.0	23535.0	24134.0	24701.0	25240.0	25739
4	SP.POP.TOTL.MA.IN	Andorra	40786.0	40361.0	39615.0	38842.0	38071.0	37380.0	36628.0	36188

5 rows × 48 columns

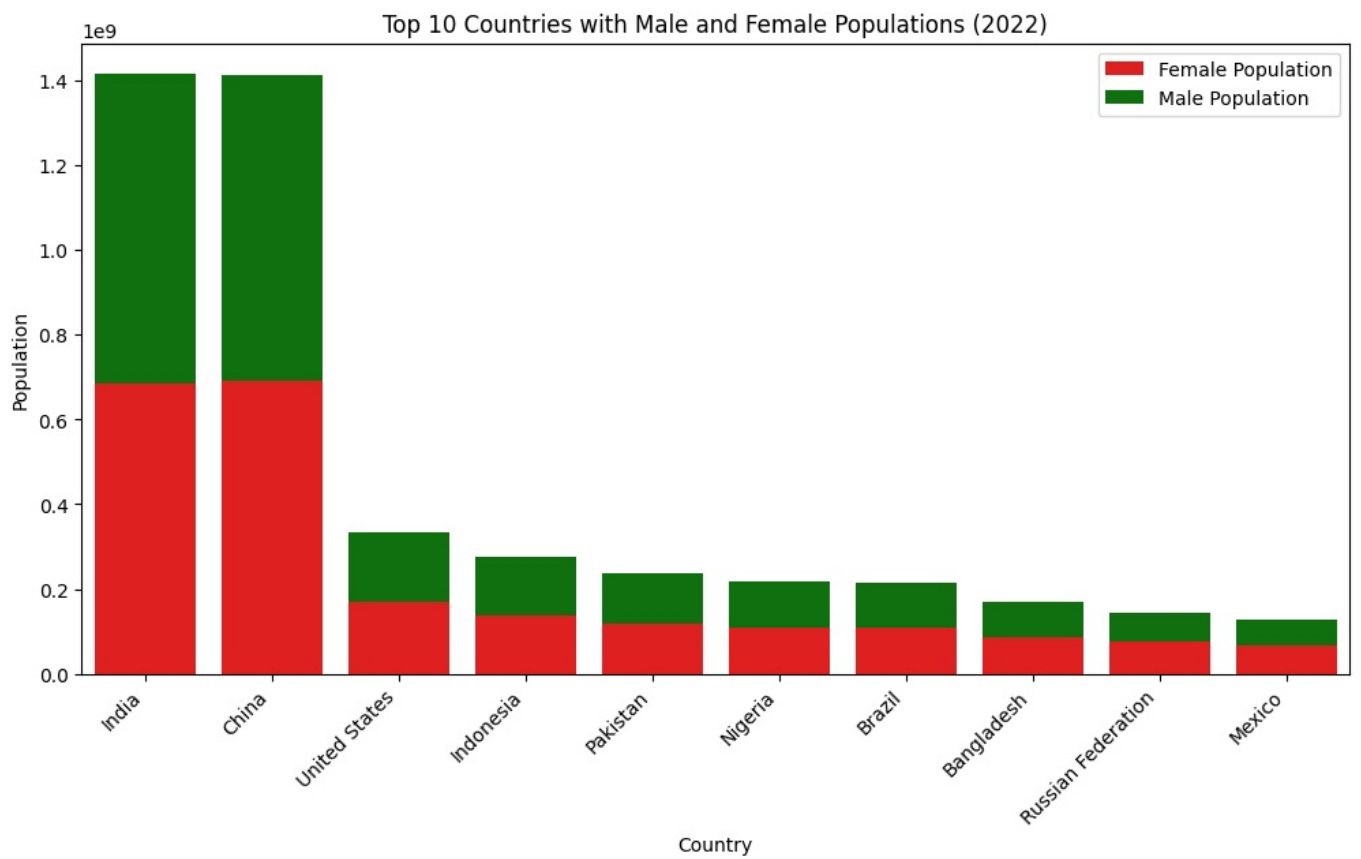


In [58]:

```
# Sort data based on total population in descending order
sorted_data = merged_data.sort_values(by="Total Population", ascending=False)

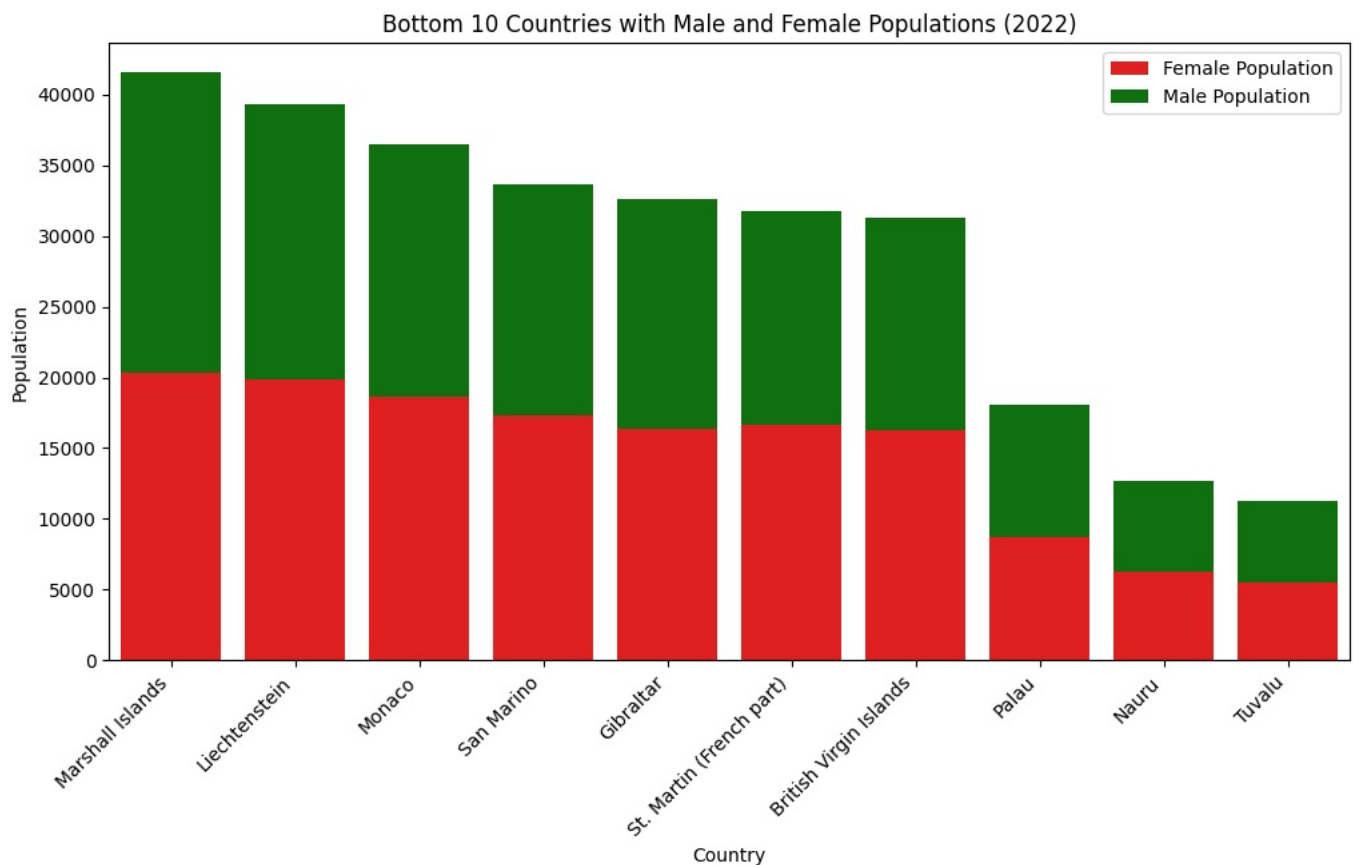
# Select the top 10 countries with the highest total population
top_10_countries = sorted_data.head(10)
# Create the stacked bar plot
plt.figure(figsize=(12, 6))

sns.barplot(x="Country Name", y="2022_female", data=top_10_countries, color="red", label="Female Population")
sns.barplot(x="Country Name", y="2022_male", data=top_10_countries, bottom=top_10_countries["2022_female"], color="blue", label="Male Population")
plt.title("Top 10 Countries with Male and Female Populations (2022)")
plt.xlabel("Country")
plt.ylabel("Population")
plt.legend()
plt.xticks(rotation=45, ha="right")
plt.show()
```



```
In [70]: #Bottom 10 Countries with Male and Female Populations (2022)
# Select the top 10 countries with the highest total population
bottom_10_countries = sorted_data.tail(10)
# Create the stacked bar plot
plt.figure(figsize=(12, 6))

sns.barplot(x="Country Name", y="2022_female", data=bottom_10_countries, color="red", label="Female Population")
sns.barplot(x="Country Name", y="2022_male", data=bottom_10_countries, bottom=bottom_10_countries["2022_female"])
plt.title("Bottom 10 Countries with Male and Female Populations (2022)")
plt.xlabel("Country")
plt.ylabel("Population")
plt.legend()
plt.xticks(rotation=45, ha="right")
plt.show()
```





In [ ]:

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