In [ ]:
```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

In [ ]:
```
df = pd.read_csv('/content/sample_data/WP.csv')
df.head()
```

Out[ ]:

	Rank	CCA3	Country	Capital	Continent	2022 Population	2020 Population	2015 Population	2010 Population	2000 Population	1
0	36	AFG	Afghanistan	Kabul	Asia	41128771.000000	38972230.000000	33753499.000000	28189672.000000	19542982.000000	1
1	138	ALB	Albania	Tirana	Europe	2842321.000000	2866849.000000	2882481.000000	2913399.000000	3182021.000000	
2	34	DZA	Algeria	Algiers	Africa	44903225.000000	43451666.000000	39543154.000000	35856344.000000	30774621.000000	2
3	213	ASM	American Samoa	Pago Pago	Oceania	44273.000000	46189.000000	51368.000000	54849.000000	58230.000000	
4	203	AND	Andorra	Andorra la Vella	Europe	79824.000000	77700.000000	71746.000000	71519.000000	66097.000000	

In [ ]:
```
df.describe() #gives mean max and stdv count of data
```

Out[ ]:

	Rank	2022 Population	2020 Population	2015 Population	2010 Population	2000 Population	1990 Population	1980 Population	1970 Population	A
count	234.00	230.00	233.00	230.00	227.00	227.00	229.00	229.00	230.00	
mean	117.50	34632250.88	33600710.95	32066004.16	30270164.48	26840495.26	19330463.93	16282884.78	15866499.13	
std	67.69	137889172.44	135873196.61	131507146.34	126074183.54	113352454.57	81309624.96	69345465.54	68355859.75	1
min	1.00	510.00	520.00	564.00	596.00	651.00	700.00	733.00	752.00	
25%	59.25	419738.50	406471.00	394295.00	382726.50	329470.00	261928.00	223752.00	145880.50	
50%	117.50	5762857.00	5456681.00	5244415.00	4889741.00	4491202.00	3785847.00	3135123.00	2511718.00	
75%	175.75	22653719.00	21522626.00	19730853.75	16825852.50	15625467.00	11882762.00	9817257.00	8817329.00	
max	234.00	1425887337.00	1424929781.00	1393715448.00	1348191368.00	1264099069.00	1153704252.00	982372466.00	822534450.00	17

In [ ]:

```
pd.set_option('display.float_format', lambda x : '%.2f' % x)
df
```

Out[ ]:

	Rank	CCA3	Country	Capital	Continent	2022 Population	2020 Population	2015 Population	2010 Population	2000 Population	1990 Population	Popula
0	36	AFG	Afghanistan	Kabul	Asia	41128771.00	38972230.00	33753499.00	28189672.00	19542982.00	10694796.00	124866
1	138	ALB	Albania	Tirana	Europe	2842321.00	2866849.00	2882481.00	2913399.00	3182021.00	3295066.00	29416
2	34	DZA	Algeria	Algiers	Africa	44903225.00	43451666.00	39543154.00	35856344.00	30774621.00	25518074.00	187393
3	213	ASM	American Samoa	Pago Pago	Oceania	44273.00	46189.00	51368.00	54849.00	58230.00	47818.00	328
4	203	AND	Andorra	Andorra la Vella	Europe	79824.00	77700.00	71746.00	71519.00	66097.00	53569.00	356
...	...	...	...	...	...	...	...	...	...	...	...	
229	226	WLF	Wallis and Futuna	Mata-Utu	Oceania	11572.00	11655.00	12182.00	13142.00	14723.00	13454.00	113
230	172	ESH	Western Sahara	El Aaiún	Africa	575986.00	556048.00	491824.00	413296.00	270375.00	178529.00	1167
231	46	YEM	Yemen	Sanaa	Asia	33696614.00	32284046.00	28516545.00	24743946.00	18628700.00	13375121.00	92049
232	63	ZMB	Zambia	Lusaka	Africa	20017675.00	18927715.00	NaN	13792086.00	9891136.00	7686401.00	57204
233	74	ZWE	Zimbabwe	Harare	Africa	16320537.00	15669666.00	14154937.00	12839771.00	11834676.00	10113893.00	70499

234 rows × 17 columns

In [ ]:

```
df.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                      234 non-null    object 
 3   Capital                      234 non-null    object 
 4   Continent                    234 non-null    object 
 5   2022 Population              230 non-null    float64
 6   2020 Population              233 non-null    float64
 7   2015 Population              230 non-null    float64
 8   2010 Population              227 non-null    float64
 9   2000 Population              227 non-null    float64
 10  1990 Population              229 non-null    float64
 11  1980 Population              229 non-null    float64
 12  1970 Population              230 non-null    float64
 13  Area (km²)                   232 non-null    float64
 14  Density (per km²)            230 non-null    float64
 15  Growth Rate                  232 non-null    float64
 16  World Population Percentage  234 non-null    float64
dtypes: float64(12), int64(1), object(4)
memory usage: 31.2+ KB
```

In [ ]:

```
df.isnull().sum() #gives count of null values in column
```

Out[ ]:

	0
Rank	0
CCA3	0
Country	0
Capital	0
Continent	0
2022 Population	4
2020 Population	1
2015 Population	4
2010 Population	7
2000 Population	7
1990 Population	5
1980 Population	5
1970 Population	4
Area (km²)	2
Density (per km²)	4
Growth Rate	2
World Population Percentage	0

dtype: int64

In [ ]:

```
df.nunique() #checks for unique values in columns outputs it as list
#Here Each Country ,Rank and capital is same since all of them are unique
# But World Population Percentage       is less because many country can have same percentage but not same name
```

Out[ ]:

	0
Rank	234
CCA3	234
Country	234
Capital	234
Continent	6
2022 Population	230
2020 Population	233
2015 Population	230
2010 Population	227
2000 Population	227
1990 Population	229
1980 Population	229
1970 Population	230
Area (km²)	231
Density (per km²)	230
Growth Rate	178
World Population Percentage	70

dtype: int64

In [ ]:

```
df.sort_values(by="2022 Population").head()
```

Out[ ]:

	Rank	CCA3	Country	Capital	Continent	2022 Population	2020 Population	2015 Population	2010 Population	2000 Population	1990 Population	1980 Population
226	234	VAT	Vatican City	Vatican City	Europe	510.00	520.00	564.00	596.00	651.00	700.00	733.00
209	233	TKL	Tokelau	Nukunonu	Oceania	1871.00	1827.00	1454.00	1367.00	1666.00	1669.00	1647.00
150	232	NIU	Niue	Alofi	Oceania	1934.00	1942.00	1847.00	1812.00	2074.00	2533.00	3637.00
64	231	FLK	Falkland Islands	Stanley	South America	3780.00	3747.00	3408.00	3187.00	3080.00	2332.00	2240.00
137	230	MSR	Montserrat	Brades	North America	4390.00	4500.00	5059.00	4938.00	5138.00	10805.00	11452.00

In [ ]:

```python
df.sort_values(by="2022 Population", ascending = False).head(5)
```

Out[ ]:

	Rank	CCA3	Country	Capital	Continent	2022 Population	2020 Population	2015 Population	2010 Population	2000 Population	Popu
41	1	CHN	China	Beijing	Asia	1425887337.00	1424929781.00	1393715448.00	1348191368.00	1264099069.00	1153704
92	2	IND	India	New Delhi	Asia	1417173173.00	1396387127.00	1322866505.00	1240613620.00	1059633675.00	
221	3	USA	United States	Washington, D.C.	North America	338289857.00	335942003.00	324607776.00	311182845.00	282398554.00	248083
93	4	IDN	Indonesia	Jakarta	Asia	275501339.00	271857970.00	259091970.00	244016173.00	214072421.00	182159
156	5	PAK	Pakistan	Islamabad	Asia	235824862.00	227196741.00	210969298.00	194454498.00	154369924.00	115414

In [ ]:

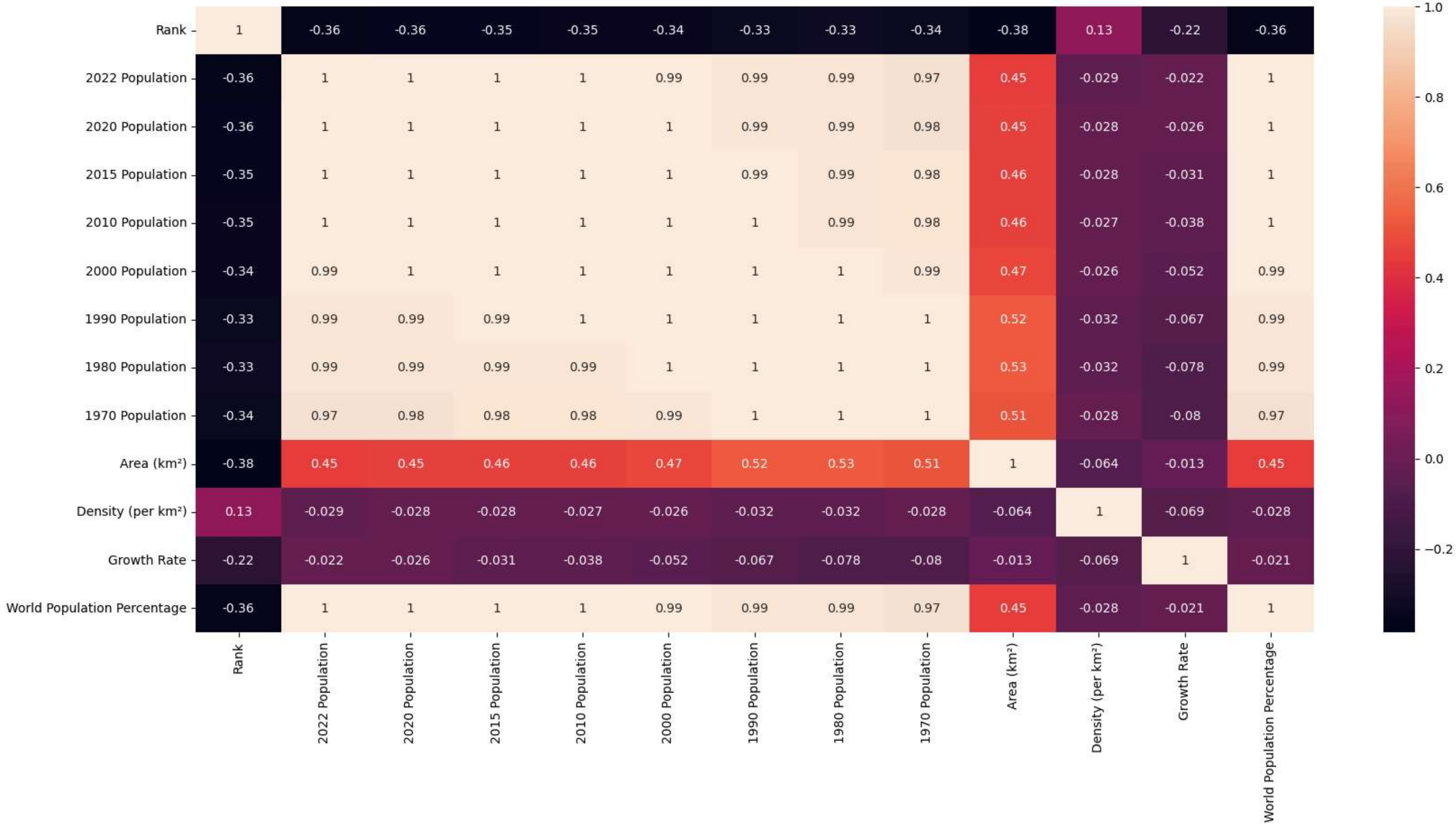
```python
#Finding Correlation

df.corr(numeric_only = True) #always add numeric_only = True since textual data can clash
```

Out[ ]:

	Rank	2022 Population	2020 Population	2015 Population	2010 Population	2000 Population	1990 Population	1980 Population	1970 Population	Area (km^2)	Density (per km^2)	Grow Ra
Rank	1.00	-0.36	-0.36	-0.35	-0.35	-0.34	-0.33	-0.33	-0.34	-0.38	0.13	-0.
2022 Population	-0.36	1.00	1.00	1.00	1.00	0.99	0.99	0.99	0.97	0.45	-0.03	-0.
2020 Population	-0.36	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.98	0.45	-0.03	-0.
2015 Population	-0.35	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.98	0.46	-0.03	-0.
2010 Population	-0.35	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.98	0.46	-0.03	-0.
2000 Population	-0.34	0.99	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.47	-0.03	-0.
1990 Population	-0.33	0.99	0.99	0.99	1.00	1.00	1.00	1.00	1.00	0.52	-0.03	-0.
1980 Population	-0.33	0.99	0.99	0.99	0.99	1.00	1.00	1.00	1.00	0.53	-0.03	-0.
1970 Population	-0.34	0.97	0.98	0.98	0.98	0.99	1.00	1.00	1.00	0.51	-0.03	-0.
Area (km^2)	-0.38	0.45	0.45	0.46	0.46	0.47	0.52	0.53	0.51	1.00	-0.06	-0.
Density (per km^2)	0.13	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.06	1.00	-0.
Growth Rate	-0.22	-0.02	-0.03	-0.03	-0.04	-0.05	-0.07	-0.08	-0.08	-0.01	-0.07	1.
World Population Percentage	-0.36	1.00	1.00	1.00	1.00	0.99	0.99	0.99	0.97	0.45	-0.03	-0.

In [ ]:
```python
#Using Heatmap to find Correlation between data
# Here there is no  correlation between Area , Density, Growth rate
sns.heatmap(df.corr(numeric_only = True),annot=True)
plt.rcParams['figure.figsize'] = (20,9) #changed figure size since it was numbers were not visble
plt.show()
```



In [ ]:
```python
#Use case --> find continents where population has grown faster
```

```
df.groupby('Continent').mean(numeric_only=True)
```

Out[ ]:

	Rank	2022 Population	2020 Population	2015 Population	2010 Population	2000 Population	1990 Population	1980 Population	1970 Population	Area (km²)	D
Continent											
Africa	92.16	25455879.68	23871435.26	21419703.57	18898197.31	14598365.95	11376964.52	8586031.98	6567175.27	537879.30	
Asia	77.56	96327387.31	94955134.37	89165003.64	89087770.00	80580835.11	48639995.33	40278333.33	43839877.83	642762.82	1
Europe	124.50	15055371.82	14915843.92	15027454.12	14712278.68	14817685.71	14785203.94	14200004.52	13118479.82	460208.22	
North America	160.93	15007403.40	14855914.82	14259596.25	13568016.28	12151739.60	10531660.62	9207334.03	7885865.15	606104.45	
Oceania	188.52	2046386.32	1910148.96	1756664.48	1613163.65	1357512.09	1162774.87	996532.17	846968.26	370220.91	
South America	97.57	31201186.29	30823574.50	29509599.71	26789395.54	25015888.69	21224743.93	17270643.29	13781939.71	1301302.85	

In [ ]:

```
df[df['Continent'].str.contains('Oceania')].head()
#Checks the country included in oceania
```

Out[ ]:

	Rank	CCA3	Country	Capital	Continent	2022 Population	2020 Population	2015 Population	2010 Population	2000 Population	1990 Population	19[illegible] Populatio[illegible]
3	213	ASM	American Samoa	Pago Pago	Oceania	44273.00	46189.00	51368.00	54849.00	58230.00	47818.00	32886.(
11	55	AUS	Australia	Canberra	Oceania	26177413.00	25670051.00	23820236.00	22019168.00	19017963.00	17048003.00	14706322.(
44	223	COK	Cook Islands	Avarua	Oceania	17011.00	17029.00	17695.00	17212.00	15897.00	17123.00	17651.(
66	162	FJI	Fiji	Suva	Oceania	929766.00	920422.00	917200.00	905169.00	832509.00	780430.00	644582.(
70	183	PYF	French Polynesia	Papeete	Oceania	306279.00	301920.00	291787.00	283788.00	250927.00	211089.00	163591.(

In [ ]:

```
df.groupby('Continent').mean(numeric_only=True).sort_values(by="2022 Population", ascending = False)
```

Out[ ]:

Continent	Rank	2022 Population	2020 Population	2015 Population	2010 Population	2000 Population	1990 Population	1980 Population	1970 Population	Area (km^2)	D
Asia	77.56	96327387.31	94955134.37	89165003.64	89087770.00	80580835.11	48639995.33	40278333.33	43839877.83	642762.82	1
South America	97.57	31201186.29	30823574.50	29509599.71	26789395.54	25015888.69	21224743.93	17270643.29	13781939.71	1301302.85	
Africa	92.16	25455879.68	23871435.26	21419703.57	18898197.31	14598365.95	11376964.52	8586031.98	6567175.27	537879.30	
Europe	124.50	15055371.82	14915843.92	15027454.12	14712278.68	14817685.71	14785203.94	14200004.52	13118479.82	460208.22	
North America	160.93	15007403.40	14855914.82	14259596.25	13568016.28	12151739.60	10531660.62	9207334.03	7885865.15	606104.45	
Oceania	188.52	2046386.32	1910148.96	1756664.48	1613163.65	1357512.09	1162774.87	996532.17	846968.26	370220.91	

In [ ]:

```
df.columns
```

Out[ ]:

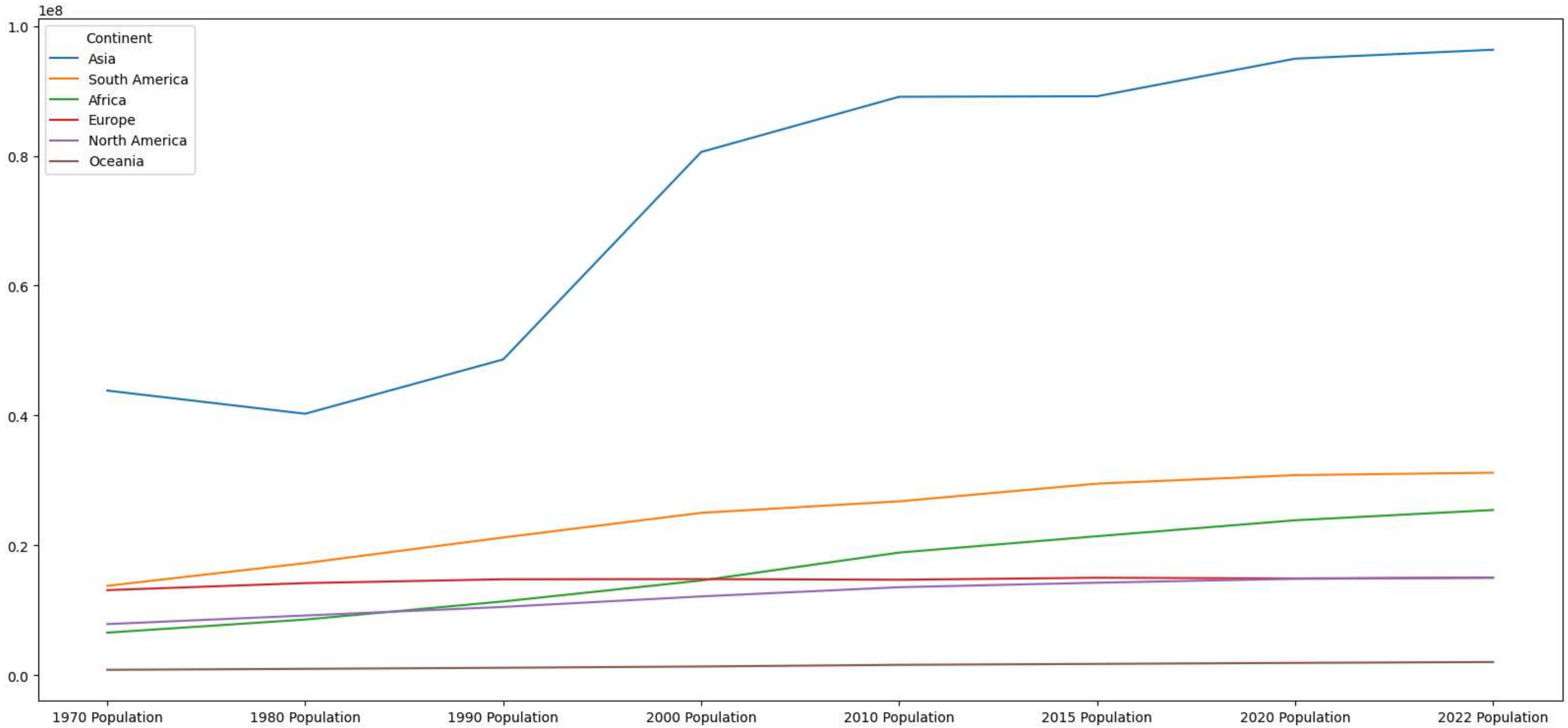
```
Index(['Rank', 'CCA3', 'Country', 'Capital', 'Continent', '2022 Population',
       '2020 Population', '2015 Population', '2010 Population',
       '2000 Population', '1990 Population', '1980 Population',
       '1970 Population', 'Area (km²)', 'Density (per km²)', 'Growth Rate',
       'World Population Percentage'],
      dtype='object')
```

In [ ]:

```
df2 = df.groupby('Continent')[['1970 Population',
       '1980 Population', '1990 Population', '2000 Population',
       '2010 Population', '2015 Population', '2020 Population',
       '2022 Population']].mean(numeric_only=True).sort_values(by="2022 Population", ascending = False)
df3 = df2.transpose()
df3.plot()
```

Out[ ]:

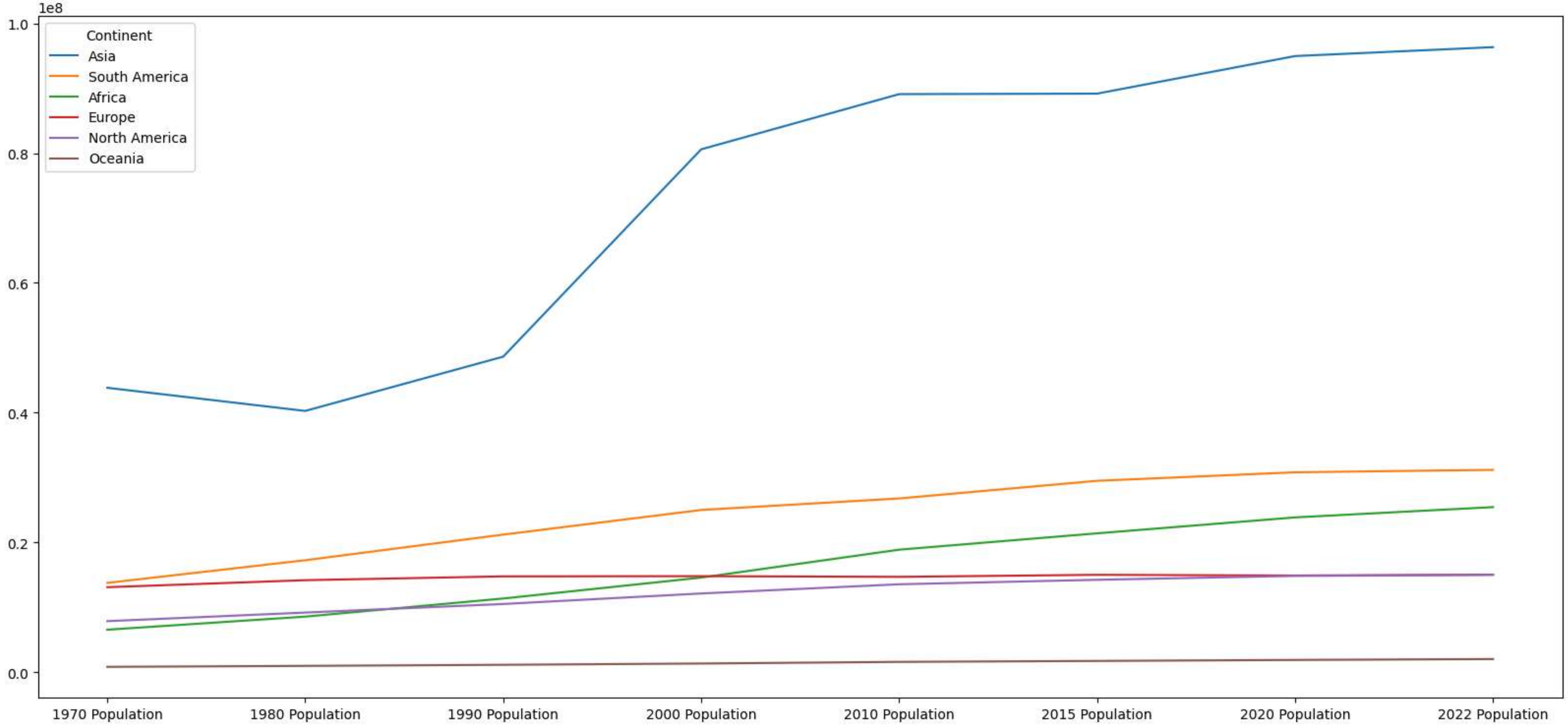
```
<Axes: >
```



In [ ]:

```
#used df.columns[::-1]
df2 = df.groupby('Continent')[df.columns[-5:-13:-1]].mean(numeric_only=True).sort_values(by="2022 Population", ascending = Fal
df3 = df2.transpose()
df3.plot()
```

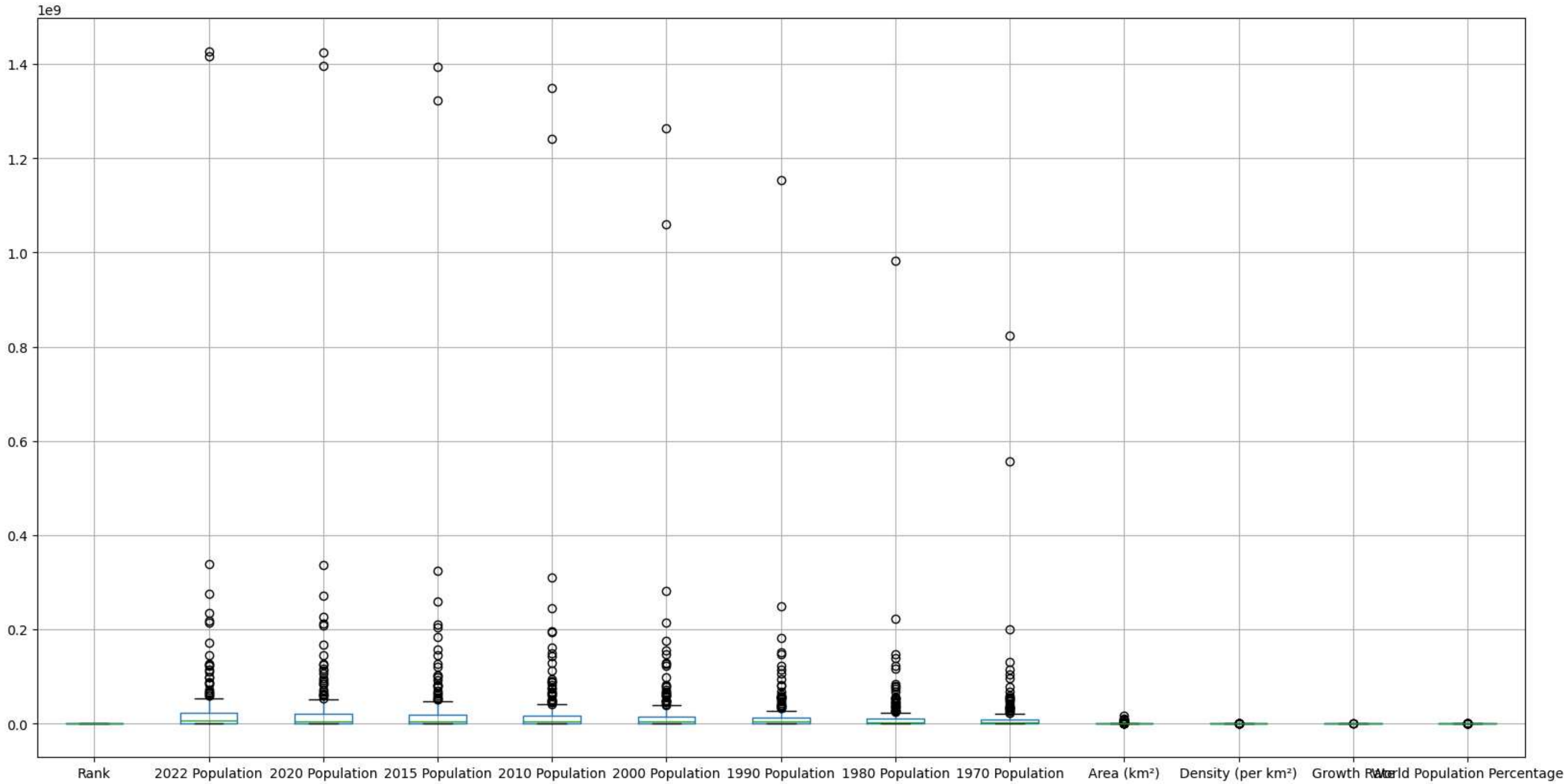
Out[ ]: <Axes: >



In [ ]:

```
df.boxplot(figsize=(20,10)) #find outliers
#since this is population we do not check(fix) outliers
```

Out[ ]: <Axes: >



In [ ]:
```
df.select_dtypes(include = 'object')
```

Out[ ]:

	CCA3	Country	Capital	Continent
0	AFG	Afghanistan	Kabul	Asia
1	ALB	Albania	Tirana	Europe
2	DZA	Algeria	Algiers	Africa
3	ASM	American Samoa	Pago Pago	Oceania
4	AND	Andorra	Andorra la Vella	Europe
...	...	...	...	...
229	WLF	Wallis and Futuna	Mata-Utu	Oceania
230	ESH	Western Sahara	El Aaiún	Africa
231	YEM	Yemen	Sanaa	Asia
232	ZMB	Zambia	Lusaka	Africa
233	ZWE	Zimbabwe	Harare	Africa

234 rows × 4 columns

In [ ]:

```
df.select_dtypes(include = 'number')
```

Out[ ]:

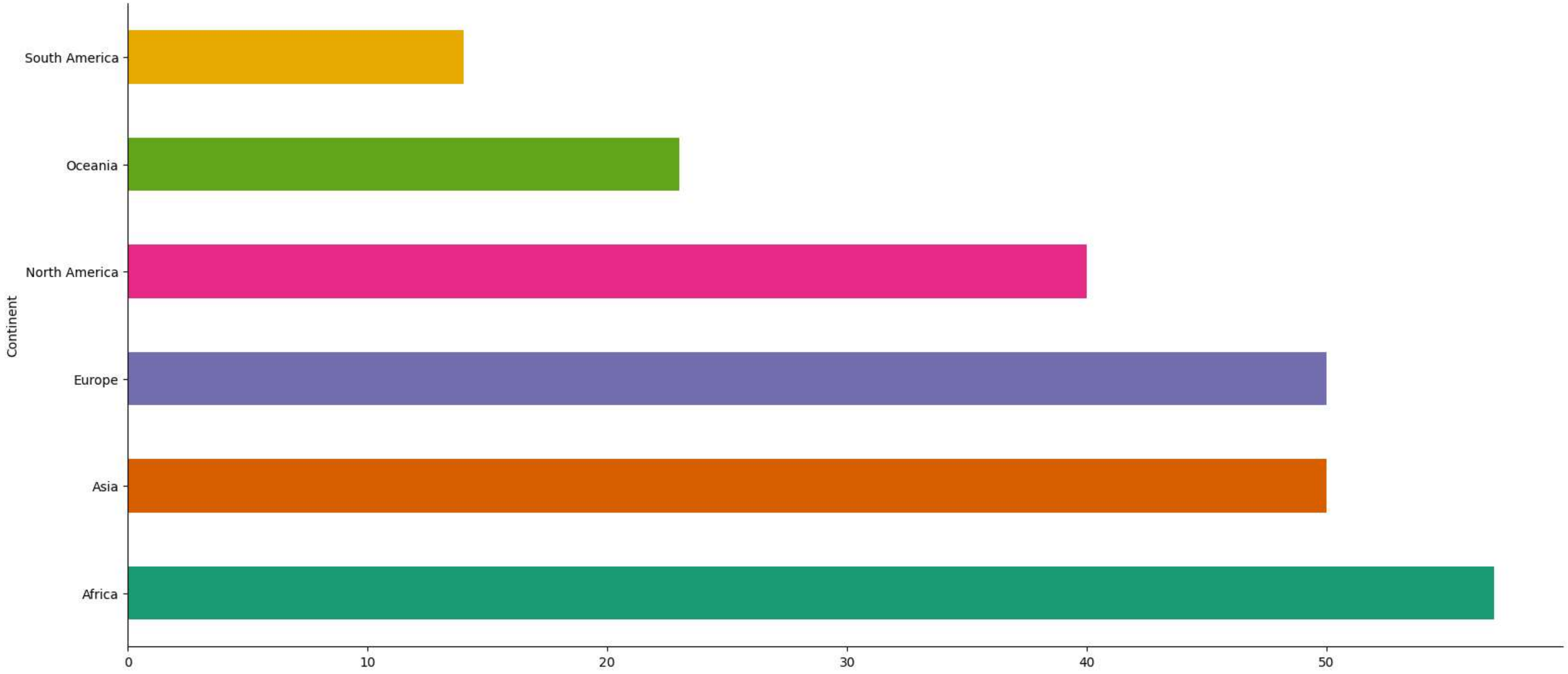
	Rank	2022 Population	2020 Population	2015 Population	2010 Population	2000 Population	1990 Population	1980 Population	1970 Population	Area (km²)	Density (per km²)
0	36	41128771.00	38972230.00	33753499.00	28189672.00	19542982.00	10694796.00	12486631.00	10752971.00	652230.00	63.06
1	138	2842321.00	2866849.00	2882481.00	2913399.00	3182021.00	3295066.00	2941651.00	2324731.00	28748.00	98.87
2	34	44903225.00	43451666.00	39543154.00	35856344.00	30774621.00	25518074.00	18739378.00	13795915.00	2381741.00	18.85
3	213	44273.00	46189.00	51368.00	54849.00	58230.00	47818.00	32886.00	27075.00	199.00	222.48
4	203	79824.00	77700.00	71746.00	71519.00	66097.00	53569.00	35611.00	19860.00	468.00	170.56
...	...	...	...	...	...	...	...	...	...	...	...
229	226	11572.00	11655.00	12182.00	13142.00	14723.00	13454.00	11315.00	9377.00	142.00	81.49
230	172	575986.00	556048.00	491824.00	413296.00	270375.00	178529.00	116775.00	76371.00	266000.00	2.17
231	46	33696614.00	32284046.00	28516545.00	24743946.00	18628700.00	13375121.00	9204938.00	6843607.00	527968.00	63.82
232	63	20017675.00	18927715.00	NaN	13792086.00	9891136.00	7686401.00	5720438.00	4281671.00	752612.00	26.60
233	74	16320537.00	15669666.00	14154937.00	12839771.00	11834676.00	10113893.00	7049926.00	5202918.00	390757.00	41.77

234 rows × 13 columns

Size of Continent

In [ ]:

```
from matplotlib import pyplot as plt
import seaborn as sns
_df_12.groupby('Continent').size().plot(kind='barh', color=sns.palettes.mpl_palette('Dark2'))
plt.gca().spines[['top', 'right',]].set_visible(False)
```



⭐ World Population Analysis Project by **Prathamesh Kedar**⭐

Tools Used: Python · Pandas · NumPy · Matplotlib