In [ ]:
```
# EDA(Exploratory Data Analysis) for World Population from 1970-2022
```

In [1]:
```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

In [15]:
```
df = pd.read_csv(r"C:\Users\Maj Mortuza\Downloads\world_population.csv")
df
```

Out[15]:

	Rank	CCA3	Country	Capital	Continent	2022 Population	2020 Population	2015 Population	2010 Population	2000 Population	1990 Population	1980 Population	197 Populatio
0	36	AFG	Afghanistan	Kabul	Asia	41128771.0	38972230.0	33753499.0	28189672.0	19542982.0	10694796.0	12486631.0	10752971.
1	138	ALB	Albania	Tirana	Europe	2842321.0	2866849.0	2882481.0	2913399.0	3182021.0	3295066.0	2941651.0	2324731.
2	34	DZA	Algeria	Algiers	Africa	44903225.0	43451666.0	39543154.0	35856344.0	30774621.0	25518074.0	18739378.0	13795915.
3	213	ASM	American Samoa	Pago Pago	Oceania	44273.0	46189.0	51368.0	54849.0	58230.0	47818.0	32886.0	27075.
4	203	AND	Andorra	Andorra la Vella	Europe	79824.0	77700.0	71746.0	71519.0	66097.0	53569.0	35611.0	19860.
...	...	...	...	...	...	...	...	...	...	...	...	...	...
229	226	WLF	Wallis and Futuna	Mata-Utu	Oceania	11572.0	11655.0	12182.0	13142.0	14723.0	13454.0	11315.0	9377
230	172	ESH	Western Sahara	El Aaiún	Africa	575986.0	556048.0	491824.0	413296.0	270375.0	178529.0	116775.0	76371.
231	46	YEM	Yemen	Sanaa	Asia	33696614.0	32284046.0	28516545.0	24743946.0	18628700.0	13375121.0	9204938.0	6843607.
232	63	ZMB	Zambia	Lusaka	Africa	20017675.0	18927715.0	NaN	13792086.0	9891136.0	7686401.0	5720438.0	4281671.
233	74	ZWE	Zimbabwe	Harare	Africa	16320537.0	15669666.0	14154937.0	12839771.0	11834676.0	10113893.0	7049926.0	5202918.

234 rows × 17 columns

In [16]:
```
# Keep 2 digit after decimal
pd.set_option('display.float_format', lambda x: '%.2f' % x)
```

In [17]:
```
df
```

Out[17]:

	Rank	CCA3	Country	Capital	Continent	2022 Population	2020 Population	2015 Population	2010 Population	2000 Population	1990 Population	1980 Population	
0	36	AFG	Afghanistan	Kabul	Asia	41128771.00	38972230.00	33753499.00	28189672.00	19542982.00	10694796.00	12486631.00	1(
1	138	ALB	Albania	Tirana	Europe	2842321.00	2866849.00	2882481.00	2913399.00	3182021.00	3295066.00	2941651.00	2
2	34	DZA	Algeria	Algiers	Africa	44903225.00	43451666.00	39543154.00	35856344.00	30774621.00	25518074.00	18739378.00	1:
3	213	ASM	American Samoa	Pago Pago	Oceania	44273.00	46189.00	51368.00	54849.00	58230.00	47818.00	32886.00	
4	203	AND	Andorra	Andorra la Vella	Europe	79824.00	77700.00	71746.00	71519.00	66097.00	53569.00	35611.00	
...	...	...	...	...	...	...	...	...	...	...	...	...	
229	226	WLF	Wallis and Futuna	Mata-Utu	Oceania	11572.00	11655.00	12182.00	13142.00	14723.00	13454.00	11315.00	
230	172	ESH	Western Sahara	El Aaiún	Africa	575986.00	556048.00	491824.00	413296.00	270375.00	178529.00	116775.00	
231	46	YEM	Yemen	Sanaa	Asia	33696614.00	32284046.00	28516545.00	24743946.00	18628700.00	13375121.00	9204938.00	(
232	63	ZMB	Zambia	Lusaka	Africa	20017675.00	18927715.00	NaN	13792086.00	9891136.00	7686401.00	5720438.00	4
233	74	ZWE	Zimbabwe	Harare	Africa	16320537.00	15669666.00	14154937.00	12839771.00	11834676.00	10113893.00	7049926.00	5

234 rows × 17 columns

In [10]:
```
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
```

In [80]:
```
# Filter out the numeric data
df.select_dtypes(include='number')
```

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

In [82]:
```
# Filter out the string data
df.select_dtypes(include='object')
```

Out[82]:

	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

In [11]:
```
df.describe()
```

Out[11]:

	Rank	2022 Population	2020 Population	2015 Population	2010 Population	2000 Population	1990 Population	1980 Population	1970 Population	Area (kr
count	234.00	230.00	233.00	230.00	227.00	227.00	229.00	229.00	230.00	232.
mean	117.50	34632250.88	33600710.95	32066004.16	30270164.48	26840495.26	19330463.93	16282884.78	15866499.13	581663.
std	67.69	137889172.44	135873196.61	131507146.34	126074183.54	113352454.57	81309624.96	69345465.54	68355859.75	1769133.
min	1.00	510.00	520.00	564.00	596.00	651.00	700.00	733.00	752.00	1.
25%	59.25	419738.50	406471.00	394295.00	382726.50	329470.00	261928.00	223752.00	145880.50	2567.
50%	117.50	5762857.00	5456681.00	5244415.00	4889741.00	4491202.00	3785847.00	3135123.00	2511718.00	77141.
75%	175.75	22653719.00	21522626.00	19730853.75	16825852.50	15625467.00	11882762.00	9817257.00	8817329.00	414643.
max	234.00	1425887337.00	1424929781.00	1393715448.00	1348191368.00	1264099069.00	1153704252.00	982372466.00	822534450.00	17098242.

In [13]:
```
# Check the number of missing values in each column
df.isnull().sum()
```

Out[13]:
```
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 [18]:
```
# Check the unique value in each column
df.nunique()
```

Out[18]:
```
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 [26]:
```
df.sort_values(by="World Population Percentage", ascending = False).head(10)
```

Out[26]:

	Rank	CCA3	Country	Capital	Continent	2022 Population	2020 Population	2015 Population	2010 Population	2000 Population	1990 Population
41	1	CHN	China	Beijing	Asia	1425887337.00	1424929781.00	1393715448.00	1348191368.00	1264099069.00	1153704252.00
92	2	IND	India	New Delhi	Asia	1417173173.00	1396387127.00	1322866505.00	1240613620.00	1059633675.00	NaN
221	3	USA	United States	Washington, D.C.	North America	338289857.00	335942003.00	324607776.00	311182845.00	282398554.00	248083732.00
93	4	IDN	Indonesia	Jakarta	Asia	275501339.00	271857970.00	259091970.00	244016173.00	214072421.00	182159874.00
156	5	PAK	Pakistan	Islamabad	Asia	235824862.00	227196741.00	210969298.00	194454498.00	154369924.00	115414069.00
149	6	NGA	Nigeria	Abuja	Africa	218541212.00	208327405.00	183995785.00	160952853.00	122851984.00	95214257.00
27	7	BRA	Brazil	Brasilia	South America	215313498.00	213196304.00	205188205.00	196353492.00	175873720.00	150706446.00
16	8	BGD	Bangladesh	Dhaka	Asia	171186372.00	167420951.00	157830000.00	148391139.00	129193327.00	107147651.00
171	9	RUS	Russia	Moscow	Europe	144713314.00	145617329.00	144668389.00	143242599.00	146844839.00	148005704.00
131	10	MEX	Mexico	Mexico City	North America	127504125.00	125998302.00	120149897.00	112532401.00	97873442.00	81720428.00

In [53]:
```
df.groupby("Continent").mean("numeric_only").sort_values(by='2022 Population', ascending=False)
```

Out[53]:

Continent	Rank	2022 Population	2020 Population	2015 Population	2010 Population	2000 Population	1990 Population	1980 Population	1970 Population	Area (km²)	Density (per km²)
Asia	77.56	96327387.31	94955134.37	89165003.64	89087770.00	80580835.11	48639995.33	40278333.33	43839877.83	642762.82	1025.02
South America	97.57	31201186.29	30823574.50	29509599.71	26789395.54	25015888.69	21224743.93	17270643.29	13781939.71	1301302.85	20.97
Africa	92.16	25455879.68	23871435.26	21419703.57	18898197.31	14598365.95	11376964.52	8586031.98	6567175.27	537879.30	126.41
Europe	124.50	15055371.82	14915843.92	15027454.12	14712278.68	14817685.71	14785203.94	14200004.52	13118479.82	460208.22	663.32
North America	160.93	15007403.40	14855914.82	14259596.25	13568016.28	12151739.60	10531660.62	9207334.03	7885865.15	606104.45	272.49
Oceania	188.52	2046386.32	1910148.96	1756664.48	1613163.65	1357512.09	1162774.87	996532.17	846968.26	370220.91	132.54

In [33]:
```
# Check Oceania continent
df[df['Continent'].str.contains('Oceania')]
```

Out[33]:

	Rank	CCA3	Country	Capital	Continent	2022 Population	2020 Population	2015 Population	2010 Population	2000 Population	1990 Population	1980 Population
3	213	ASM	American Samoa	Pago Pago	Oceania	44273.00	46189.00	51368.00	54849.00	58230.00	47818.00	32886.00
11	55	AUS	Australia	Canberra	Oceania	26177413.00	25670051.00	23820236.00	22019168.00	19017963.00	17048003.00	14706322.00
44	223	COK	Cook Islands	Avarua	Oceania	17011.00	17029.00	17695.00	17212.00	15897.00	17123.00	17651.00
66	162	FJI	Fiji	Suva	Oceania	929766.00	920422.00	917200.00	905169.00	832509.00	780430.00	644582.00
70	183	PYF	French Polynesia	Papeete	Oceania	306279.00	301920.00	291787.00	283788.00	250927.00	211089.00	163591.00
81	191	GUM	Guam	Hagåtña	Oceania	171774.00	169231.00	167978.00	164905.00	160188.00	138263.00	110286.00
107	192	KIR	Kiribati	Tarawa	Oceania	131232.00	126463.00	116707.00	107995.00	88826.00	75124.00	60813.00
126	215	MHL	Marshall Islands	Majuro	Oceania	41569.00	43413.00	49410.00	53416.00	54224.00	46047.00	31988.00
132	194	FSM	Micronesia	Palikir	Oceania	114164.00	112106.00	109462.00	107588.00	111709.00	98603.00	76299.00
142	225	NRU	Nauru	Yaren	Oceania	12668.00	12315.00	11185.00	10241.00	10377.00	9598.00	7635.00
145	185	NCL	New Caledonia	Nouméa	Oceania	289950.00	286403.00	283032.00	261426.00	221537.00	177264.00	148599.00
146	123	NZL	New Zealand	Wellington	Oceania	5185288.00	5061133.00	4590590.00	4346338.00	3855266.00	3397389.00	3147168.00
150	232	NIU	Niue	Alofi	Oceania	1934.00	1942.00	1847.00	1812.00	2074.00	2533.00	3637.00
153	210	NFK	Northern Mariana Islands	Saipan	Oceania	49551.00	49587.00	51514.00	54087.00	80338.00	48002.00	17613.00
157	222	PLW	Palau	Ngerulmud	Oceania	NaN	17972.00	17794.00	18540.00	19726.00	15293.00	12252.00
160	93	PNG	Papua New Guinea	Port Moresby	Oceania	10142619.00	9749640.00	8682174.00	7583269.00	5508297.00	3864972.00	3104788.00
179	188	WSM	Samoa	Apia	Oceania	222382.00	214929.00	203571.00	194672.00	184008.00	168186.00	164905.00
191	166	SLB	Solomon Islands	Honiara	Oceania	724273.00	691191.00	612660.00	540394.00	429978.00	324171.00	233668.00

	Rank	CCA3	Country	Capital	Continent	2022 Population	2020 Population	2015 Population	2010 Population	2000 Population	1990 Population	1980 Population
209	233	TKL	Tokelau	Nukunonu	Oceania	1871.00	1827.00	1454.00	1367.00	1666.00	1669.00	1647.00
210	197	TON	Tonga	Nuku‘alofa	Oceania	106858.00	105254.00	106122.00	107383.00	102603.00	98727.00	96708.00
216	227	TUV	Tuvalu	Funafuti	Oceania	11312.00	11069.00	10877.00	10550.00	9638.00	9182.00	7731.00
225	181	VUT	Vanuatu	Port-Vila	Oceania	326740.00	311685.00	276438.00	245453.00	192074.00	150882.00	118156.00
229	226	WLF	Wallis and Futuna	Mata-Utu	Oceania	11572.00	11655.00	12182.00	13142.00	14723.00	13454.00	11315.00

In [86]:
```
df2.columns
```

Out[86]:
```
Index(['1970 Population', '1980 Population', '1990 Population',
       '2000 Population', '2010 Population', '2015 Population',
       '2020 Population', '2022 Population'],
      dtype='object')
```

In [83]:
```
df2 = df.groupby("Continent")[['1970 Population', '1980 Population', '1990 Population',
       '2000 Population', '2010 Population', '2015 Population',
       '2020 Population', '2022 Population']].mean("numeric_only").sort_values(by='2022 Population', ascending=False)
df2
```

Out[83]:

Continent	1970 Population	1980 Population	1990 Population	2000 Population	2010 Population	2015 Population	2020 Population	2022 Population
Asia	43839877.83	40278333.33	48639995.33	80580835.11	89087770.00	89165003.64	94955134.37	96327387.31
South America	13781939.71	17270643.29	21224743.93	25015888.69	26789395.54	29509599.71	30823574.50	31201186.29
Africa	6567175.27	8586031.98	11376964.52	14598365.95	18898197.31	21419703.57	23871435.26	25455879.68
Europe	13118479.82	14200004.52	14785203.94	14817685.71	14712278.68	15027454.12	14915843.92	15055371.82
North America	7885865.15	9207334.03	10531660.62	12151739.60	13568016.28	14259596.25	14855914.82	15007403.40
Oceania	846968.26	996532.17	1162774.87	1357512.09	1613163.65	1756664.48	1910148.96	2046386.32

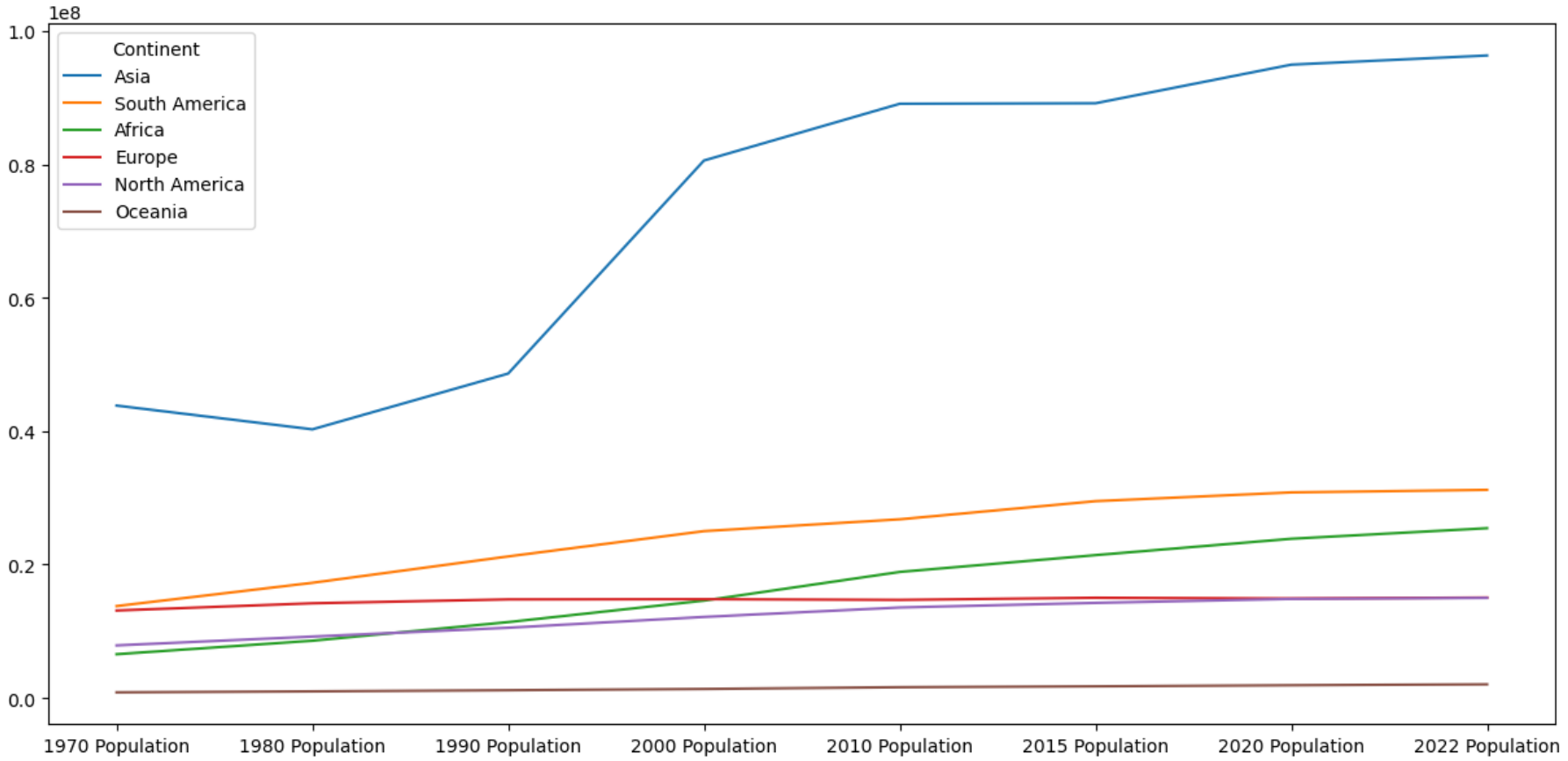
In [84]:
```
# Switch column to index and vice verse
df3 = df2.transpose()
df3
```

Out[84]:

Continent	Asia	South America	Africa	Europe	North America	Oceania
1970 Population	43839877.83	13781939.71	6567175.27	13118479.82	7885865.15	846968.26
1980 Population	40278333.33	17270643.29	8586031.98	14200004.52	9207334.03	996532.17
1990 Population	48639995.33	21224743.93	11376964.52	14785203.94	10531660.62	1162774.87
2000 Population	80580835.11	25015888.69	14598365.95	14817685.71	12151739.60	1357512.09
2010 Population	89087770.00	26789395.54	18898197.31	14712278.68	13568016.28	1613163.65
2015 Population	89165003.64	29509599.71	21419703.57	15027454.12	14259596.25	1756664.48
2020 Population	94955134.37	30823574.50	23871435.26	14915843.92	14855914.82	1910148.96
2022 Population	96327387.31	31201186.29	25455879.68	15055371.82	15007403.40	2046386.32

In [85]:
```
plt.rcParams['figure.figsize']=(15, 7)
df3.plot()
```

Out[85]: <Axes: >



In [88]:
```
# Check outlier
df.boxplot(figsize=(25,10))
```

Out[88]: <Axes: >