In [5]:
```
#import data with read and print

import pandas as pd
df = pd.read_csv('worldpopulation.csv')
print(df)
```

```
     Rank CCA3            Country          Capital Continent  \
0      36  AFG        Afghanistan            Kabul      Asia
1     138  ALB            Albania           Tirana    Europe
2      34  DZA            Algeria          Algiers    Africa
3     213  ASM     American Samoa        Pago Pago   Oceania
4     203  AND            Andorra  Andorra la Vella    Europe
..    ...  ...                ...              ...       ...
229   226  WLF  Wallis and Futuna          Mata-Utu   Oceania
230   172  ESH     Western Sahara          El Aaiún    Africa
231    46  YEM              Yemen             Sanaa      Asia
232    63  ZMB             Zambia            Lusaka    Africa
233    74  ZWE           Zimbabwe            Harare    Africa

     2022 Population  2020 Population  2015 Population  2010 Population  \
0         41128771.0       38972230.0       33753499.0       28189672.0
1          2842321.0        2866849.0        2882481.0        2913399.0
2         44903225.0       43451666.0       39543154.0       35856344.0
3            44273.0          46189.0          51368.0          54849.0
4            79824.0          77700.0          71746.0          71519.0
..               ...              ...              ...              ...
229          11572.0          11655.0          12182.0          13142.0
230         575986.0         556048.0         491824.0         413296.0
231       33696614.0       32284046.0       28516545.0       24743946.0
232       20017675.0       18927715.0              NaN       13792086.0
233       16320537.0       15669666.0       14154937.0       12839771.0

     2000 Population  1990 Population  1980 Population  1970 Population  \
0         19542982.0       10694796.0       12486631.0       10752971.0
1          3182021.0        3295066.0        2941651.0        2324731.0
2         30774621.0       25518074.0       18739378.0       13795915.0
3            58230.0          47818.0          32886.0          27075.0
4            66097.0          53569.0          35611.0          19860.0
..               ...              ...              ...              ...
229          14723.0          13454.0          11315.0           9377.0
230         270375.0         178529.0         116775.0          76371.0
231       18628700.0       13375121.0        9204938.0        6843607.0
232        9891136.0        7686401.0        5720438.0        4281671.0
233       11834676.0       10113893.0        7049926.0        5202918.0

     Area (km²)  Density (per km²)  Growth Rate  World Population Percentage
0      652230.0            63.0587       1.0257                         0.52
1       28748.0            98.8702       0.9957                         0.04
2     2381741.0            18.8531       1.0164                         0.56
3         199.0           222.4774       0.9831                         0.00
4         468.0           170.5641       1.0100                         0.00
..          ...                ...          ...                          ...
229       142.0            81.4930       0.9953                         0.00
230    266000.0             2.1654       1.0184                         0.01
231    527968.0            63.8232       1.0217                         0.42
232    752612.0            26.5976       1.0280                         0.25
233    390757.0            41.7665       1.0204                         0.20

[234 rows x 17 columns]
```

In [6]:
```
#Viewing info on the dataset to see column, non-null and datatype overview
```

```
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 [7]:
```
#describe() gives some stats overview of the data

df.describe()
```

Out[7]:

	Rank	2022 Population	2020 Population	2015 Population	2010 Population	2000 Population	Pop
count	234.000000	2.300000e+02	2.330000e+02	2.300000e+02	2.270000e+02	2.270000e+02	2.2900
mean	117.500000	3.463225e+07	3.360071e+07	3.206600e+07	3.027016e+07	2.684050e+07	1.9330
std	67.694165	1.378892e+08	1.358732e+08	1.315071e+08	1.260742e+08	1.133525e+08	8.1309
min	1.000000	5.100000e+02	5.200000e+02	5.640000e+02	5.960000e+02	6.510000e+02	7.0000
25%	59.250000	4.197385e+05	4.064710e+05	3.942950e+05	3.827265e+05	3.294700e+05	2.6192
50%	117.500000	5.762857e+06	5.456681e+06	5.244415e+06	4.889741e+06	4.491202e+06	3.7858
75%	175.750000	2.265372e+07	2.152263e+07	1.973085e+07	1.682585e+07	1.562547e+07	1.1882
max	234.000000	1.425887e+09	1.424930e+09	1.393715e+09	1.348191e+09	1.264099e+09	1.1537

In [11]:
```
#isnull() shows the null and null value status as a bool

df.isnull()
```

Out[11]:

	Rank	CCA3	Country	Capital	Continent	2022 Population	2020 Population	2015 Population	2010 Population	I
0	False	False	False	False	False	False	False	False	False	
1	False	False	False	False	False	False	False	False	False	
2	False	False	False	False	False	False	False	False	False	
3	False	False	False	False	False	False	False	False	False	
4	False	False	False	False	False	False	False	False	False	
...	...	...	...	...	...	...	...	...	...	
229	False	False	False	False	False	False	False	False	False	
230	False	False	False	False	False	False	False	False	False	
231	False	False	False	False	False	False	False	False	False	
232	False	False	False	False	False	False	False	True	False	
233	False	False	False	False	False	False	False	False	False	

234 rows × 17 columns

In [12]:
```
#to show to sum of the null

df.isnull().sum()
```

Out[12]:
```
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 [13]:
```
#to see the number of unique values in the data against the column
df.nunique()
```

Out[13]:
```
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 [14]:
```
#sorting by a particular column. By default, we will have  the sort by ascendding c

df.sort_values('2022 Population')
```

Out[14]:

	Rank	CCA3	Country	Capital	Continent	2022 Population	2020 Population	2015 Population
226	234	VAT	Vatican City	Vatican City	Europe	5.100000e+02	5.200000e+02	5.640000e+02
209	233	TKL	Tokelau	Nukunonu	Oceania	1.871000e+03	1.827000e+03	1.454000e+03
150	232	NIU	Niue	Alofi	Oceania	1.934000e+03	1.942000e+03	1.847000e+03
64	231	FLK	Falkland Islands	Stanley	South America	3.780000e+03	3.747000e+03	3.408000e+03
137	230	MSR	Montserrat	Brades	North America	4.390000e+03	4.500000e+03	5.059000e+03
...	...	...	...	...	...	...	...	...
41	1	CHN	China	Beijing	Asia	1.425887e+09	1.424930e+09	1.393715e+09
62	159	SWZ	Eswatini	Mbabane	Africa	NaN	1.180655e+06	1.133936e+06
154	120	NOR	Norway	Oslo	Europe	NaN	5.379839e+06	NaN
157	222	PLW	Palau	Ngerulmud	Oceania	NaN	1.797200e+04	1.779400e+04
207	155	TLS	Timor-Leste	Dili	Asia	NaN	1.299995e+06	1.205813e+06

234 rows × 17 columns

In [16]:
```
#To sort by descending order, we use ascending= "False"
df.sort_values('2022 Population', ascending = False)
```

Out[16]:

	Rank	CCA3	Country	Capital	Continent	2022 Population	2020 Population	2015 Population	
41	1	CHN	China	Beijing	Asia	1.425887e+09	1.424930e+09	1.393715e+09	1
92	2	IND	India	New Delhi	Asia	1.417173e+09	1.396387e+09	1.322867e+09	1
221	3	USA	United States	Washington, D.C.	North America	3.382899e+08	3.359420e+08	3.246078e+08	3
93	4	IDN	Indonesia	Jakarta	Asia	2.755013e+08	2.718580e+08	2.590920e+08	2
156	5	PAK	Pakistan	Islamabad	Asia	2.358249e+08	2.271967e+08	2.109693e+08	1
...	...	...	...	...	...	...	...	...	
226	234	VAT	Vatican City	Vatican City	Europe	5.100000e+02	5.200000e+02	5.640000e+02	5
62	159	SWZ	Eswatini	Mbabane	Africa	NaN	1.180655e+06	1.133936e+06	1
154	120	NOR	Norway	Oslo	Europe	NaN	5.379839e+06	NaN	4
157	222	PLW	Palau	Ngerulmud	Oceania	NaN	1.797200e+04	1.779400e+04	1
207	155	TLS	Timor-Leste	Dili	Asia	NaN	1.299995e+06	1.205813e+06	1

234 rows × 17 columns

In [17]:
```
#To check the relationship btw my data, i use correlation "corr()"

df.corr()
```

```
C:\Users\aniwa\AppData\Local\Temp\ipykernel_15848\1470766260.py:3: FutureWarning:
The default value of numeric_only in DataFrame.corr is deprecated. In a future ver
sion, it will default to False. Select only valid columns or specify the value of
numeric_only to silence this warning.
  df.corr()
```

Out[17]:

	Rank	2022 Population	2020 Population	2015 Population	2010 Population	2000 Population	1990 Population	P
Rank	1.000000	-0.357989	-0.356283	-0.352850	-0.350009	-0.342095	-0.333014	
2022 Population	-0.357989	1.000000	0.999946	0.999492	0.998633	0.994602	0.991124	
2020 Population	-0.356283	0.999946	1.000000	0.999764	0.999108	0.995582	0.992557	
2015 Population	-0.352850	0.999492	0.999764	1.000000	0.999784	0.997343	0.994968	
2010 Population	-0.350009	0.998633	0.999108	0.999784	1.000000	0.998595	0.996716	
2000 Population	-0.342095	0.994602	0.995582	0.997343	0.998595	1.000000	0.999192	
1990 Population	-0.333014	0.991124	0.992557	0.994968	0.996716	0.999192	1.000000	
1980 Population	-0.328620	0.986675	0.988392	0.991346	0.993628	0.997247	0.999257	
1970 Population	-0.335737	0.973135	0.975245	0.979413	0.983050	0.990959	0.997158	
Area (km^2)	-0.384854	0.452812	0.454902	0.458101	0.461498	0.473137	0.524392	
Density (per km^2)	0.128994	-0.028761	-0.027802	-0.027740	-0.026635	-0.026341	-0.032348	
Growth Rate	-0.220197	-0.022059	-0.025976	-0.031150	-0.038262	-0.052041	-0.066718	
World Population Percentage	-0.358464	0.999999	0.999944	0.999489	0.998630	0.994598	0.991102	

In [37]:
```
#To group by a parameter and show the mean values

df2 = df.groupby("Continent").mean()
print(df2)
```

Continent	Rank	2022 Population	2020 Population	2015 Population
Africa	92.157895	2.545588e+07	2.387144e+07	2.141970e+07
Asia	77.560000	9.632739e+07	9.495513e+07	8.916500e+07
Europe	124.500000	1.505537e+07	1.491584e+07	1.502745e+07
North America	160.925000	1.500740e+07	1.485591e+07	1.425960e+07
Oceania	188.521739	2.046386e+06	1.910149e+06	1.756664e+06
South America	97.571429	3.120119e+07	3.082357e+07	2.950960e+07

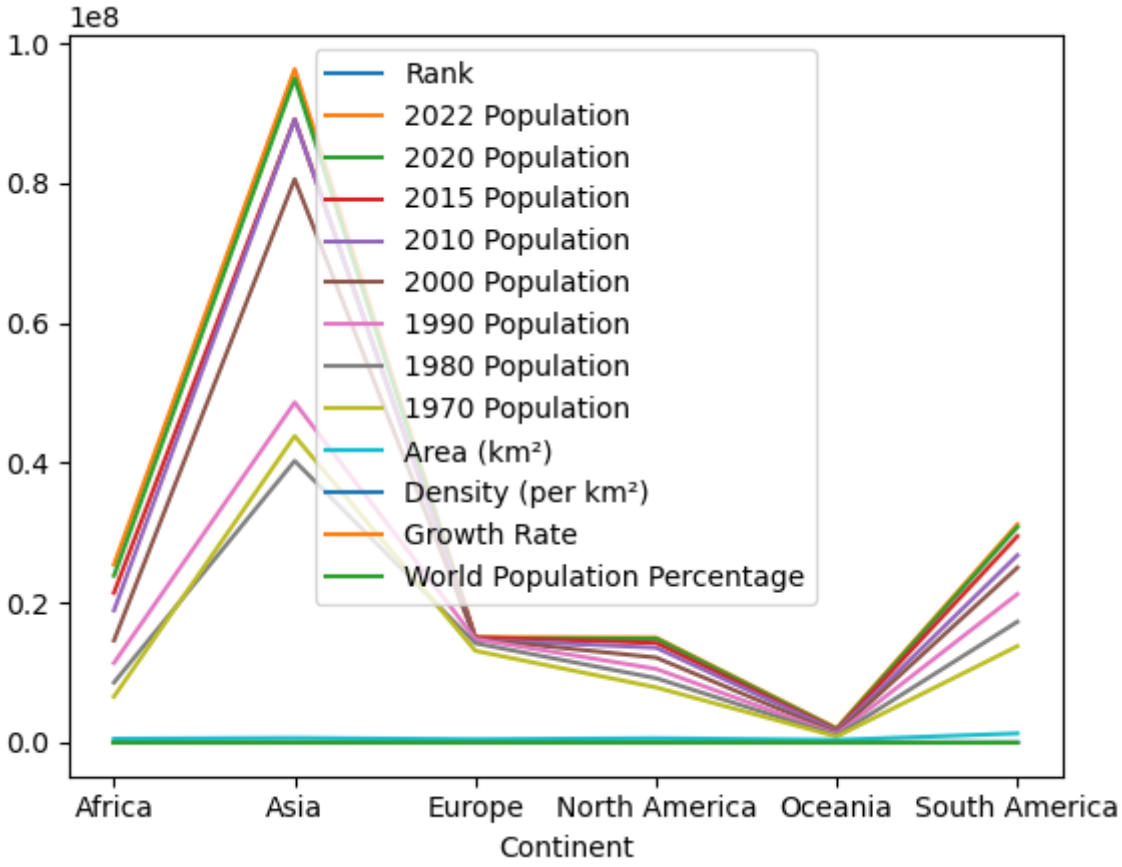
Continent	2010 Population	2000 Population	1990 Population
Africa	1.889820e+07	1.459837e+07	1.137696e+07
Asia	8.908777e+07	8.058084e+07	4.864000e+07
Europe	1.471228e+07	1.481769e+07	1.478520e+07
North America	1.356802e+07	1.215174e+07	1.053166e+07
Oceania	1.613164e+06	1.357512e+06	1.162775e+06
South America	2.678940e+07	2.501589e+07	2.122474e+07

Continent	1980 Population	1970 Population	Area (km²)
Africa	8.586032e+06	6.567175e+06	5.378793e+05
Asia	4.027833e+07	4.383988e+07	6.427628e+05
Europe	1.420000e+07	1.311848e+07	4.602082e+05
North America	9.207334e+06	7.885865e+06	6.061044e+05
Oceania	9.965322e+05	8.469683e+05	3.702209e+05
South America	1.727064e+07	1.378194e+07	1.301303e+06

Continent	Density (per km²)	Growth Rate	World Population Percentage
Africa	126.406569	1.021180	0.313509
Asia	1025.024136	1.009384	1.183800
Europe	663.324742	1.002256	0.186600
North America	272.494118	1.004285	0.187750
Oceania	132.543065	1.007383	0.023913
South America	20.971979	1.007957	0.391429

In [40]:
```
#Having the data on a chart
df2.plot()
```

Out[40]:
```
<Axes: xlabel='Continent'>
```



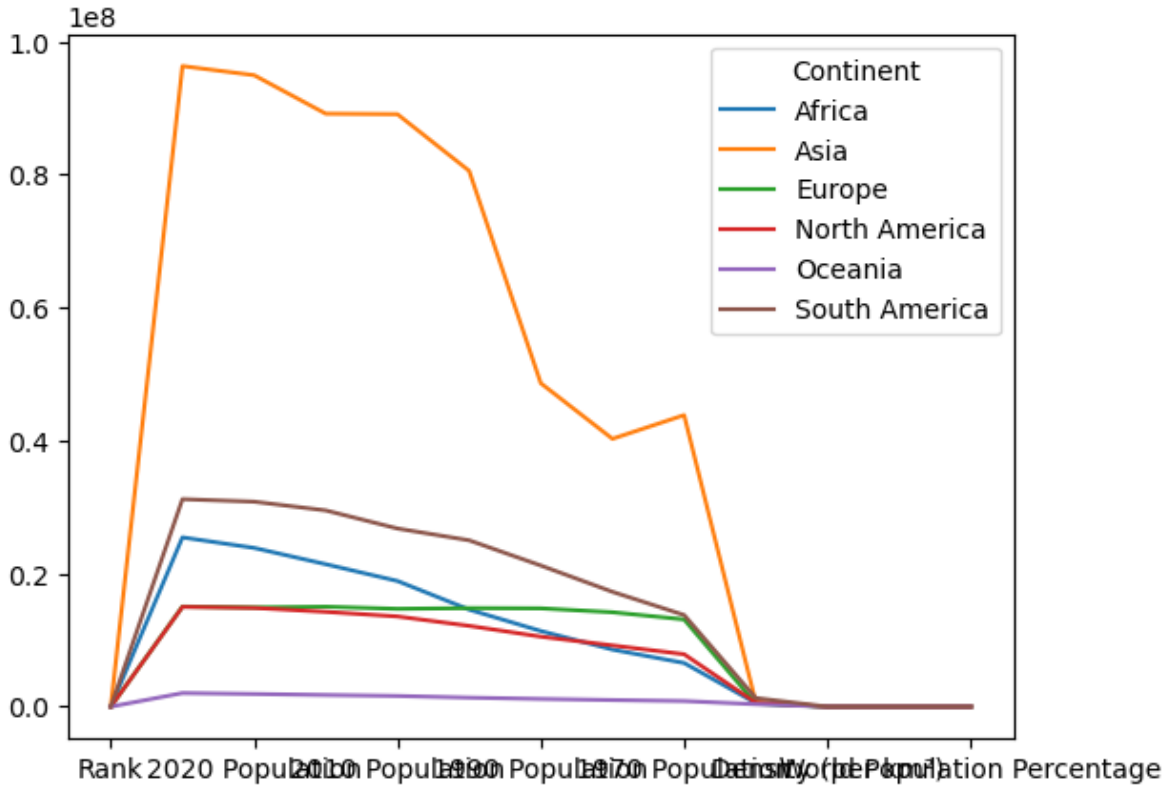
In [44]:
```
#Te above doesnt plot niely so i can transpose the axis
df3 = df2.transpose()
print(df3)
```

```
Continent                        Africa          Asia        Europe  \
Rank                       9.215789e+01  7.756000e+01  1.245000e+02
2022 Population            2.545588e+07  9.632739e+07  1.505537e+07
2020 Population            2.387144e+07  9.495513e+07  1.491584e+07
2015 Population            2.141970e+07  8.916500e+07  1.502745e+07
2010 Population            1.889820e+07  8.908777e+07  1.471228e+07
2000 Population            1.459837e+07  8.058084e+07  1.481769e+07
1990 Population            1.137696e+07  4.864000e+07  1.478520e+07
1980 Population            8.586032e+06  4.027833e+07  1.420000e+07
1970 Population            6.567175e+06  4.383988e+07  1.311848e+07
Area (km²)                 5.378793e+05  6.427628e+05  4.602082e+05
Density (per km²)          1.264066e+02  1.025024e+03  6.633247e+02
Growth Rate                1.021180e+00  1.009384e+00  1.002256e+00
World Population Percentage  3.135088e-01  1.183800e+00  1.866000e-01

Continent                  North America       Oceania  South America
Rank                        1.609250e+02  1.885217e+02   9.757143e+01
2022 Population             1.500740e+07  2.046386e+06   3.120119e+07
2020 Population             1.485591e+07  1.910149e+06   3.082357e+07
2015 Population             1.425960e+07  1.756664e+06   2.950960e+07
2010 Population             1.356802e+07  1.613164e+06   2.678940e+07
2000 Population             1.215174e+07  1.357512e+06   2.501589e+07
1990 Population             1.053166e+07  1.162775e+06   2.122474e+07
1980 Population             9.207334e+06  9.965322e+05   1.727064e+07
1970 Population             7.885865e+06  8.469683e+05   1.378194e+07
Area (km²)                  6.061044e+05  3.702209e+05   1.301303e+06
Density (per km²)           2.724941e+02  1.325431e+02   2.097198e+01
Growth Rate                 1.004285e+00  1.007383e+00   1.007957e+00
World Population Percentage   1.877500e-01  2.391304e-02   3.914286e-01
```

In [45]:
```
#I will plot the new transpose axix
df3.plot()
```

Out[45]:
```
<Axes: >
```



In [ ]: