In [33]:
```
import pandas as pd
pd.set_option('display.max.rows',10)
```

In [34]:
```
df = pd.read_csv(r"C:\Users\kallzz\Desktop\Data Analytics Stuff\Data Analyst - Boot
df
```

Out[34]:

	Rank	CCA3	Country	Capital	Continent	2022 Population	2020 Population	2015 Population	201(Populatio
0	36	AFG	Afghanistan	Kabul	Asia	41128771.0	38972230.0	33753499.0	28189672.(
1	138	ALB	Albania	Tirana	Europe	2842321.0	2866849.0	2882481.0	2913399.(
2	34	DZA	Algeria	Algiers	Africa	44903225.0	43451666.0	39543154.0	35856344.(
3	213	ASM	American Samoa	Pago Pago	Oceania	44273.0	46189.0	51368.0	54849.(
4	203	AND	Andorra	Andorra la Vella	Europe	79824.0	77700.0	71746.0	71519.(
...	...	...	...	...	...	...	...	...	...
229	226	WLF	Wallis and Futuna	Mata-Utu	Oceania	11572.0	11655.0	12182.0	13142.(
230	172	ESH	Western Sahara	El Aaiún	Africa	575986.0	556048.0	491824.0	413296.(
231	46	YEM	Yemen	Sanaa	Asia	33696614.0	32284046.0	28516545.0	24743946.(
232	63	ZMB	Zambia	Lusaka	Africa	20017675.0	18927715.0	NaN	13792086.(
233	74	ZWE	Zimbabwe	Harare	Africa	16320537.0	15669666.0	14154937.0	12839771.(

234 rows × 17 columns

In [35]:
```
# to view data based on a condition
# df[df['Rank'] < 10]
df[df['Rank'] <= 10]
```

Out[35]:

	Rank	CCA3	Country	Capital	Continent	2022 Population	2020 Population	2015 Population
16	8	BGD	Bangladesh	Dhaka	Asia	1.711864e+08	1.674210e+08	1.578300e+08
27	7	BRA	Brazil	Brasilia	South America	2.153135e+08	2.131963e+08	2.051882e+08
41	1	CHN	China	Beijing	Asia	1.425887e+09	1.424930e+09	1.393715e+09
92	2	IND	India	New Delhi	Asia	1.417173e+09	1.396387e+09	1.322867e+09
93	4	IDN	Indonesia	Jakarta	Asia	2.755013e+08	2.718580e+08	2.590920e+08
131	10	MEX	Mexico	Mexico City	North America	1.275041e+08	1.259983e+08	1.201499e+08
149	6	NGA	Nigeria	Abuja	Africa	2.185412e+08	2.083274e+08	1.839958e+08
156	5	PAK	Pakistan	Islamabad	Asia	2.358249e+08	2.271967e+08	2.109693e+08
171	9	RUS	Russia	Moscow	Europe	1.447133e+08	1.456173e+08	1.446684e+08
221	3	USA	United States	Washington, D.C.	North America	3.382899e+08	3.359420e+08	3.246078e+08

In [36]:
```
# isin()

specific_counties = ['India', 'China']
df[df['Country'].isin(specific_counties)]
```

Out[36]:

	Rank	CCA3	Country	Capital	Continent	2022 Population	2020 Population	2015 Population	Popul
41	1	CHN	China	Beijing	Asia	1.425887e+09	1.424930e+09	1.393715e+09	1.34819
92	2	IND	India	New Delhi	Asia	1.417173e+09	1.396387e+09	1.322867e+09	1.24061

In [37]:
```
# contains() - wild card filter
df[df['Country'].str.contains('United')]
```

Out[37]:

	Rank	CCA3	Country	Capital	Continent	2022 Population	2020 Population	2015 Population	Popu
219	97	ARE	United Arab Emirates	Abu Dhabi	Asia	9441129.0	9287289.0	8916899.0	848
220	21	GBR	United Kingdom	London	Europe	67508936.0	67059474.0	65224364.0	6276
221	3	USA	United States	Washington, D.C.	North America	338289857.0	335942003.0	324607776.0	31118
222	200	VIR	United States Virgin Islands	Charlotte Amalie	North America	99465.0	100442.0	102803.0	10

In [38]:
```
# set_index

df2 = df.set_index('Country')
df2
```

Out[38]:

Country	Rank	CCA3	Capital	Continent	2022 Population	2020 Population	2015 Population	2010 Population	Po
Afghanistan	36	AFG	Kabul	Asia	41128771.0	38972230.0	33753499.0	28189672.0	19
Albania	138	ALB	Tirana	Europe	2842321.0	2866849.0	2882481.0	2913399.0	3
Algeria	34	DZA	Algiers	Africa	44903225.0	43451666.0	39543154.0	35856344.0	30
American Samoa	213	ASM	Pago Pago	Oceania	44273.0	46189.0	51368.0	54849.0	
Andorra	203	AND	Andorra la Vella	Europe	79824.0	77700.0	71746.0	71519.0	
...	...	...	...	...	...	...	...	...	
Wallis and Futuna	226	WLF	Mata-Utu	Oceania	11572.0	11655.0	12182.0	13142.0	
Western Sahara	172	ESH	El Aaiún	Africa	575986.0	556048.0	491824.0	413296.0	
Yemen	46	YEM	Sanaa	Asia	33696614.0	32284046.0	28516545.0	24743946.0	18
Zambia	63	ZMB	Lusaka	Africa	20017675.0	18927715.0	NaN	13792086.0	9
Zimbabwe	74	ZWE	Harare	Africa	16320537.0	15669666.0	14154937.0	12839771.0	11

234 rows × 16 columns

In [39]:
```
# filter() -- to filter data at row and column level
df2.filter(items = ['Rank', 'CCA3'])
```

Out[39]:

Country	Rank	CCA3
Afghanistan	36	AFG
Albania	138	ALB
Algeria	34	DZA
American Samoa	213	ASM
Andorra	203	AND
...	...	...
Wallis and Futuna	226	WLF
Western Sahara	172	ESH
Yemen	46	YEM
Zambia	63	ZMB
Zimbabwe	74	ZWE

234 rows × 2 columns

In [40]:
```
# axis = 1 is a default values which fetch the coulmn wise data
df2.filter(items = ['Rank', 'CCA3'], axis = 1)
```

Out[40]:

Country	Rank	CCA3
Afghanistan	36	AFG
Albania	138	ALB
Algeria	34	DZA
American Samoa	213	ASM
Andorra	203	AND
...	...	...
Wallis and Futuna	226	WLF
Western Sahara	172	ESH
Yemen	46	YEM
Zambia	63	ZMB
Zimbabwe	74	ZWE

234 rows × 2 columns

In [41]:
```
df2.filter(items = ['Rank', 'CCA3'], axis = 0)
# axis = 0 is to fetch row data so items should have row data in filter function
```

Out[41]:

Country	Rank	CCA3	Capital	Continent	2022 Population	2020 Population	2015 Population	2010 Population	2 Popula

In [42]:
```
df2.filter(items = ['India', 'China'], axis = 0)
```

Out[42]:

Country	Rank	CCA3	Capital	Continent	2022 Population	2020 Population	2015 Population	201 Populatio
India	2	IND	New Delhi	Asia	1.417173e+09	1.396387e+09	1.322867e+09	1.240614e+0
China	1	CHN	Beijing	Asia	1.425887e+09	1.424930e+09	1.393715e+09	1.348191e+0

In [43]:
```
# like - wild card in filter function
df2.filter(like = 'United', axis = 0)
```

Out[43]:

Country	Rank	CCA3	Capital	Continent	2022 Population	2020 Population	2015 Population	201 Populatio
United Arab Emirates	97	ARE	Abu Dhabi	Asia	9441129.0	9287289.0	8916899.0	8481771.
United Kingdom	21	GBR	London	Europe	67508936.0	67059474.0	65224364.0	62760039.
United States	3	USA	Washington, D.C.	North America	338289857.0	335942003.0	324607776.0	311182845.
United States Virgin Islands	200	VIR	Charlotte Amalie	North America	99465.0	100442.0	102803.0	106142.

In [44]:
```
df2.loc['India']
```

Out[44]:
```
Rank                            2
CCA3                          IND
Capital                 New Delhi
Continent                    Asia
2022 Population       1417173173.0
                             ...
1970 Population        557501301.0
Area (km²)               3287590.0
Density (per km²)          431.0675
Growth Rate                1.0068
World Population Percentage 17.77
Name: India, Length: 16, dtype: object
```

In [45]:
```
df2.iloc[1]
```

Out[45]:
```
Rank                                138
CCA3                                ALB
Capital                          Tirana
Continent                        Europe
2022 Population                2842321.0
                                   ...
1970 Population                2324731.0
Area (km²)                       28748.0
Density (per km²)                98.8702
Growth Rate                       0.9957
World Population Percentage         0.04
Name: Albania, Length: 16, dtype: object
```

In [46]:
```
df[df['Rank'] < 10]
```

Out[46]:

	Rank	CCA3	Country	Capital	Continent	2022 Population	2020 Population	2015 Population
16	8	BGD	Bangladesh	Dhaka	Asia	1.711864e+08	1.674210e+08	1.578300e+08
27	7	BRA	Brazil	Brasilia	South America	2.153135e+08	2.131963e+08	2.051882e+08
41	1	CHN	China	Beijing	Asia	1.425887e+09	1.424930e+09	1.393715e+09
92	2	IND	India	New Delhi	Asia	1.417173e+09	1.396387e+09	1.322867e+09
93	4	IDN	Indonesia	Jakarta	Asia	2.755013e+08	2.718580e+08	2.590920e+08
149	6	NGA	Nigeria	Abuja	Africa	2.185412e+08	2.083274e+08	1.839958e+08
156	5	PAK	Pakistan	Islamabad	Asia	2.358249e+08	2.271967e+08	2.109693e+08
171	9	RUS	Russia	Moscow	Europe	1.447133e+08	1.456173e+08	1.446684e+08
221	3	USA	United States	Washington, D.C.	North America	3.382899e+08	3.359420e+08	3.246078e+08

In [47]:
```
# rank values are sorted in ascending order by default
df[df['Rank'] < 10].sort_values(by = 'Rank')
```

Out[47]:

	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
92	2	IND	India	New Delhi	Asia	1.417173e+09	1.396387e+09	1.322867e+09
221	3	USA	United States	Washington, D.C.	North America	3.382899e+08	3.359420e+08	3.246078e+08
93	4	IDN	Indonesia	Jakarta	Asia	2.755013e+08	2.718580e+08	2.590920e+08
156	5	PAK	Pakistan	Islamabad	Asia	2.358249e+08	2.271967e+08	2.109693e+08
149	6	NGA	Nigeria	Abuja	Africa	2.185412e+08	2.083274e+08	1.839958e+08
27	7	BRA	Brazil	Brasilia	South America	2.153135e+08	2.131963e+08	2.051882e+08
16	8	BGD	Bangladesh	Dhaka	Asia	1.711864e+08	1.674210e+08	1.578300e+08
171	9	RUS	Russia	Moscow	Europe	1.447133e+08	1.456173e+08	1.446684e+08

In [48]:
```
# rank values are sorted in ascending = True order by default
df[df['Rank'] < 10].sort_values(by = 'Rank',ascending = True)

# descending
df[df['Rank'] < 10].sort_values(by = 'Rank',ascending = False)
```

Out[48]:

	Rank	CCA3	Country	Capital	Continent	2022 Population	2020 Population	2015 Population
171	9	RUS	Russia	Moscow	Europe	1.447133e+08	1.456173e+08	1.446684e+08
16	8	BGD	Bangladesh	Dhaka	Asia	1.711864e+08	1.674210e+08	1.578300e+08
27	7	BRA	Brazil	Brasilia	South America	2.153135e+08	2.131963e+08	2.051882e+08
149	6	NGA	Nigeria	Abuja	Africa	2.185412e+08	2.083274e+08	1.839958e+08
156	5	PAK	Pakistan	Islamabad	Asia	2.358249e+08	2.271967e+08	2.109693e+08
93	4	IDN	Indonesia	Jakarta	Asia	2.755013e+08	2.718580e+08	2.590920e+08
221	3	USA	United States	Washington, D.C.	North America	3.382899e+08	3.359420e+08	3.246078e+08
92	2	IND	India	New Delhi	Asia	1.417173e+09	1.396387e+09	1.322867e+09
41	1	CHN	China	Beijing	Asia	1.425887e+09	1.424930e+09	1.393715e+09

In [49]:
```
# sort values based on more than one column
df[df['Rank'] < 10].sort_values(by = ['Rank', 'Country'],ascending = True)

# Rank is given first priority for sorting as it is a first value in by and Country
```

Out[49]:

	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
92	2	IND	India	New Delhi	Asia	1.417173e+09	1.396387e+09	1.322867e+09
221	3	USA	United States	Washington, D.C.	North America	3.382899e+08	3.359420e+08	3.246078e+08
93	4	IDN	Indonesia	Jakarta	Asia	2.755013e+08	2.718580e+08	2.590920e+08
156	5	PAK	Pakistan	Islamabad	Asia	2.358249e+08	2.271967e+08	2.109693e+08
149	6	NGA	Nigeria	Abuja	Africa	2.185412e+08	2.083274e+08	1.839958e+08
27	7	BRA	Brazil	Brasilia	South America	2.153135e+08	2.131963e+08	2.051882e+08
16	8	BGD	Bangladesh	Dhaka	Asia	1.711864e+08	1.674210e+08	1.578300e+08
171	9	RUS	Russia	Moscow	Europe	1.447133e+08	1.456173e+08	1.446684e+08

In [50]:
```
# sorting order will change based on by values
df[df['Rank'] < 10].sort_values(by = ['Country', 'Rank'],ascending = True)
```

Out[50]:

	Rank	CCA3	Country	Capital	Continent	2022 Population	2020 Population	2015 Population
16	8	BGD	Bangladesh	Dhaka	Asia	1.711864e+08	1.674210e+08	1.578300e+08
27	7	BRA	Brazil	Brasilia	South America	2.153135e+08	2.131963e+08	2.051882e+08
41	1	CHN	China	Beijing	Asia	1.425887e+09	1.424930e+09	1.393715e+09
92	2	IND	India	New Delhi	Asia	1.417173e+09	1.396387e+09	1.322867e+09
93	4	IDN	Indonesia	Jakarta	Asia	2.755013e+08	2.718580e+08	2.590920e+08
149	6	NGA	Nigeria	Abuja	Africa	2.185412e+08	2.083274e+08	1.839958e+08
156	5	PAK	Pakistan	Islamabad	Asia	2.358249e+08	2.271967e+08	2.109693e+08
171	9	RUS	Russia	Moscow	Europe	1.447133e+08	1.456173e+08	1.446684e+08
221	3	USA	United States	Washington, D.C.	North America	3.382899e+08	3.359420e+08	3.246078e+08

In [51]:
```
# sorting order will change based on by and ascending values
df[df['Rank'] < 10].sort_values(by = ['Country', 'Rank'],ascending = [True, False])
```

Out[51]:

	Rank	CCA3	Country	Capital	Continent	2022 Population	2020 Population	2015 Population
16	8	BGD	Bangladesh	Dhaka	Asia	1.711864e+08	1.674210e+08	1.578300e+08
27	7	BRA	Brazil	Brasilia	South America	2.153135e+08	2.131963e+08	2.051882e+08
41	1	CHN	China	Beijing	Asia	1.425887e+09	1.424930e+09	1.393715e+09
92	2	IND	India	New Delhi	Asia	1.417173e+09	1.396387e+09	1.322867e+09
93	4	IDN	Indonesia	Jakarta	Asia	2.755013e+08	2.718580e+08	2.590920e+08
149	6	NGA	Nigeria	Abuja	Africa	2.185412e+08	2.083274e+08	1.839958e+08
156	5	PAK	Pakistan	Islamabad	Asia	2.358249e+08	2.271967e+08	2.109693e+08
171	9	RUS	Russia	Moscow	Europe	1.447133e+08	1.456173e+08	1.446684e+08
221	3	USA	United States	Washington, D.C.	North America	3.382899e+08	3.359420e+08	3.246078e+08

In [ ]:

In [ ]: