Data Extraction of countries data

In this project, we are going extract the following data from Countries CSV file.

1. which country has the highest population
2. what is the capital of the country with highest population
3. which country has the least population
4. what is the capital of the country with least population
5. give me top 5 countries with highest democratic score
6. how many total regions are there
7. how many countries lie in Eastern Europe region
8. who is the political leader of the 2nd highest populated country
9. how many countries are there whoes political leaders are unknown
10. how many country have Republic in their full name
11. which country in african region has highest population

In [1]:
```
## Importing NumPy & Pandas library
import numpy as np
import pandas as pd
```

In [3]:
```
## importing data from CSV
df = pd.read_csv("Countries.csv")
```

In [4]:
```
df
```

Out[4]:

	country	country_long	currency	capital_city	region	continent	demonym	latitude	longitude	agricultural_land
0	Afghanistan	Islamic State of Afghanistan	Afghan afghani	Kabul	Southern Asia	Asia	Afghan	33.000000	65.000000	383560.0
1	Albania	Republic of Albania	Albanian lek	Tirana	Southern Europe	Europe	Albanian	41.000000	20.000000	11655.5
2	Algeria	People's Democratic Republic of Algeria	Algerian dinar	Algiers	Northern Africa	Africa	Algerian	28.000000	3.000000	413588.0
3	Andorra	Principality of Andorra	Euro	Andorra la Vella	Southern Europe	Europe	Andorran	42.500000	1.500000	187.2
4	Angola	People's Republic of Angola	Angolan kwanza	Luanda	Middle Africa	Africa	Angolan	-12.500000	18.500000	569525.0
...	...	...	...	...	...	...	...	...	...	...
189	Vietnam	Socialist Republic of Vietnam	Vietnamese dong	Hanoi	South-Eastern Asia	Asia	Vietnamese	16.166667	107.833333	123600.0
190	West Bank and Gaza	West Bank and Gaza	Israeli new shekel	Ramallah	Western Asia	Asia	Palestinian	31.900000	35.200000	4449.0
191	Yemen	Republic of Yemen	Yemeni rial	Sana'a	Western Asia	Asia	Yemeni	15.000000	48.000000	234520.0
192	Zambia	Republic of Zambia	New Zambian kwacha	Lusaka	Eastern Africa	Africa	Zambian	-15.000000	30.000000	238360.0
193	Zimbabwe	Republic of Zimbabwe	U.S. dollar	Harare	Eastern Africa	Africa	Zimbabwean	-20.000000	30.000000	162000.0

194 rows × 64 columns

In [ ]:
```
## data has 194 rows & 64 columns
df.shape
```

Out[ ]:
```
(194, 64)
```

In [6]:
```
## these are the 64 columns
df.columns
```

Out[6]:
```
Index(['country', 'country_long', 'currency', 'capital_city', 'region',
       'continent', 'demonym', 'latitude', 'longitude', 'agricultural_land',
       'forest_area', 'land_area', 'rural_land', 'urban_land',
       'central_government_debt_pct_gdp', 'expense_pct_gdp', 'gdp',
       'inflation', 'self_employed_pct', 'tax_revenue_pct_gdp',
       'unemployment_pct', 'vulnerable_employment_pct',
       'electricity_access_pct', 'alternative_nuclear_energy_pct',
       'electricty_production_coal_pct',
       'electricty_production_hydroelectric_pct',
       'electricty_production_gas_pct', 'electricty_production_nuclear_pct',
       'electricty_production_oil_pct', 'electricty_production_renewable_pct',
       'energy_imports_pct', 'fossil_energy_consumption_pct',
       'renewable_energy_consumption_pct', 'co2_emissions',
       'methane_emissions', 'nitrous_oxide_emissions',
       'greenhouse_other_emissions', 'urban_population_under_5m',
       'health_expenditure_pct_gdp', 'health_expenditure_capita',
       'hospital_beds', 'hiv_incidence', 'suicide_rate', 'armed_forces',
       'internally_displaced_persons', 'military_expenditure_pct_gdp',
       'birth_rate', 'death_rate', 'fertility_rate', 'internet_pct',
       'life_expectancy', 'net_migration', 'population_female',
       'population_male', 'population', 'women_parliament_seats_pct',
       'rural_population', 'urban_population', 'press', 'democracy_score',
       'democracy_type', 'median_age', 'political_leader', 'title'],
      dtype='object')
```

In [7]:
```
##first 5 rows
df.head()
```

Out[7]:

	country	country_long	currency	capital_city	region	continent	demonym	latitude	longitude	agricultural_land	...	popul
0	Afghanistan	Islamic State of Afghanistan	Afghan afghani	Kabul	Southern Asia	Asia	Afghan	33.0	65.0	383560.0	...	4112
1	Albania	Republic of Albania	Albanian lek	Tirana	Southern Europe	Europe	Albanian	41.0	20.0	11655.5	...	277
2	Algeria	People's Democratic Republic of Algeria	Algerian dinar	Algiers	Northern Africa	Africa	Algerian	28.0	3.0	413588.0	...	4490
3	Andorra	Principality of Andorra	Euro	Andorra la Vella	Southern Europe	Europe	Andorran	42.5	1.5	187.2	...	7
4	Angola	People's Republic of Angola	Angolan kwanza	Luanda	Middle Africa	Africa	Angolan	-12.5	18.5	569525.0	...	3558

5 rows × 64 columns

In [8]:
```
# 1.which country has the highest population
df[df["population"]==df["population"].max()]["country"]
```

Out[8]:
```
75    India
Name: country, dtype: object
```

In [9]:
```
# 2 what is the capital of the country with highest population
df[df["population"]==df["population"].max()]["capital_city"]
```

Out[9]:
```
75    New Delhi
Name: capital_city, dtype: object
```

In [10]:
```
# 3 which country has the least population
df[df["population"]==df["population"].min()]["country"]
```

Out[10]:
```
179    Tuvalu
Name: country, dtype: object
```

In [11]:
```
# 4 what is the capital of the country with least population
df[df["population"]==df["population"].min()]["capital_city"]
```

Out[11]:
```
179    Funafuti
Name: capital_city, dtype: object
```

In [12]:
```
# 5 give me top 5 countries with highest democratic score
df.sort_values("democracy_score",ascending=False)["country"].head(5)
```

Out[12]:
```
127         Norway
74         Iceland
164         Sweden
122    New Zealand
46         Denmark
Name: country, dtype: object
```

In [19]:
```
# 6 how many total regions are there
df["region"].nunique()
```

Out[19]:
```
22
```

In [26]:
```
# how many countries lie in Eastern Europe region
df[df["region"]=="Eastern Europe"]["country"].nunique()
```

Out[26]:
```
10
```

In [31]:
```
# 8 who is the political leader of the 2nd highest populated country
df.nlargest(2,"population").iloc[-1]["political_leader"]
```

Out[31]:
```
'Xi Jinping'
```

In [33]:
```
# 9 how many countries are there whoes political leaders are unknown
df["political_leader"].isna().sum()
```

Out[33]:
```
np.int64(7)
```

In [37]:
```
# 10 how many country have Republic in their full name
df["country_long"].str.contains("Republic",case=False,na=False).sum()
```

Out[37]:
```
np.int64(125)
```

In [51]:
```
# 11 which country in african region has highest population
african_countries = df[df["region"].str.contains("africa",case=False,na=False)]
result = african_countries.nlargest(1,"population")["country"]
result
```

Out[51]:
```
125    Nigeria
Name: country, dtype: object
```

Thank you