

Homework Assignment

In this assignment, you will analyze the dataset on world population from Kaggle.

<https://www.kaggle.com/datasets/iamsouravbanerjee/world-population-dataset>

This data set with 235 rows and 17 columns:

1. Rank: Rank by Population.
2. CCA3: 3 Digit Country/Territories Code.
3. Country: Name of the Country/Territories.
4. Capital: Name of the Capital.
5. Continent: Name of the Continent.
6. 2022 Population: Population of the Country/Territories in the year 2022.
7. 2020 Population: Population of the Country/Territories in the year 2020.
8. 2015 Population: Population of the Country/Territories in the year 2015.

9. 2010 Population: Population of the Country/Territories in the year 2010.
10. 2000 Population: Population of the Country/Territories in the year 2000.
11. 1990 Population: Population of the Country/Territories in the year 1990.
12. 1980 Population: Population of the Country/Territories in the year 1980.
13. 1970 Population: Population of the Country/Territories in the year 1970.
14. Area (km^2): Area size of the Country/Territories in square kilometer.
15. Density (per km^2): Population Density per square kilometer.
16. Growth Rate: Population Growth Rate by Country/Territories.
17. World Population Percentage: The population percentage by each Country/Territories.

Questions:

1. load the dataset csv file as a dataframe using Pandas
2. compute annual rate of population growth over 50 years from 1970 to 2020. Express the result as percentage rounded to 2 decimal points. What 5 countries had the highest annual growth and which had the lowest annual growth?
3. compute annual rate of population growth over 10 years from 2010 to 2020. Express the result as percentage rounded to 2 decimal points. What 5 countries had the highest annual growth and which had the lowest annual growth? Are the countries the same as in the previous question.
4. for 2020 what are the 5 countries with the highest and what are the 5 countries with the lowest population?
5. for 1970 what are the 5 countries with the highest and what are the 5 countries with the lowest population? Which countries remained in 2022?
6. for 2020, compute the mean μ and the quartiles Q_1 (25%), Q_2 (50% or median M), and Q_3 (75%).
7. find 3 countries "around" Q_1 , M , Q_3 and μ
8. consider the columns for years 2020, 2010, 2000, 1990 and 1980. For each columns compute q_1 and q_3 . For each

of these years, construct 5 histograms by considering only countries with populations from q_1 to q_3 for that year. use the following colors: 2020 (in green), 2010 (in blue), 2000 (in cyan), 1990 (in black), and 1980 (in red). Write each histogram to a separate pdf file (with a descriptive name).

9. examine your histograms. Any interesting observations?
10. compute the rank by population density
11. compute the rank of countries (by population) in 1970 and compute the difference in rank from 1970 to 2020. Which 5 countries experienced the largest positive change (by value, not percentage) in rank and which 5 countries experiences the largest negative change in rank
12. take the population in 2020 (all countries). Compute the first digit and compute the % of occurrence of this digit.