PROJECT PART 1

Data:
1.Import data
2.Understand the data

- Look at the rows and columns (what data do they have?)
- Calculate total national population size
- Calculate total urban population size
- Calculate total rural population size
- Calculate the urban share to national population
- Calculate the rural share to national population
- Create a visualization to compare the share of the national population living in urban vs rural areas
- Calculate the measures of central tendency and spread of the four national service levels.
- Visualize the five-number summary of the four access features across the three different types of areas.
- Visualize the national access to water on all four levels based on the national population size.
- Visualize the urban access to water on all four levels based on the urban population
- Visualize the rural access to water on all four levels based on the rural population.
- Summarize the dataset to group and investigate by the four income groups.(Create a pivot table.)

  Questions to answer

  1. What is the percentage difference between the dataset and the estimated world urban population size (the total number of people living in urban areas) for 2020?
  2. What is the interquartile range of the estimated rural share of people with surface service feature, wat_sur_r?
  3. Based on the created pivot table, what is the national average percentage of access to limited services (wat_lim_n) for low-income countries?

PROJECT PART 2

Transforming the data

- Which years are represented for which countries.
- Calculate the average difference in years for data entries per country.
- calculate the average year difference across all countries
- Calculate the minimum and maximum year difference
- create a histogram of the year column.

- Calculate ARC per country, i.e. only calculate the ARC between two years of the same country name.(ARC_x = (wat_bas_x(n+1) - wat_bas_x(n))/(year(n+1) - year(n)))
- Calculate the average, minimum, and maximum for each of the ARC values for access to basic service level for each of the three population groups.
- Calculate the number of countries per area that have full access and Annual Rates of Change equal to zero, smaller than zero, and greater than zero.
- calculate the number of countries that have full access per population
- Calculate the number of countries that have ARC values equal to zero that doesn't already have full access for each of the population types: national, rural, and urban.
- Calculate the number of countries where ARC < 0 and doesn't have full access for each of the population types: national, rural, and urban.
- Calculate the number of countries where ARC > 0 and doesn't have full access for each of the population types.
- Calculate the difference between the Annual Rates of Change between rural and urban populations per country.
- Create a new feature called ARC_diff in the dataset sheet that calculates the difference between the rural ARC (ARC_r) and urban ARC (ARC_u) for every second row since these rows are empty.
- Create a histogram of the newly created ARC_diff feature.
- Import the Regions.csv into a new sheet.
- Add a new column to the original dataset called region and use any LOOKUP function to add the region based on the country name.
- In the summary sheet, use any preferred method(s) to calculate:
  a. The number of countries per region.
  b. The average Annual Rates of Change on a national level per region.
  c. The average Annual Rates of Change in rural areas per region.
  d. The average Annual Rates of Change in urban areas per region.
- Create a visualization that represents the national ARC versus the rural ARC, as well as the region and national population size.

Questions to answer

1. What is the average Annual Rates of Change (ARC) of access to basic water services for rural populations (ARC_r) across all countries?
2. How many countries' national populations had a 0% Annual Rates of Change, excluding countries that have 100% access, across the time period?
3. Which two countries had the highest *absolute difference* between urban and rural Annual Rates of Change?
4. On average, which region saw the greatest improvement in access to basic water services on a national level (considering the Annual Rates of Change) over the dataset time period?