**Project Summary/Overview**

A two-part integrated project on investigating access to safe and affordable drinking water (SDG 6: Clean water & Sanitation) using google sheets. The first part of the project covers understanding the features in our data and the second involves transforming our data.

**Tools Used**

* Google Sheets/Microsoft Excel

**Datasets**

* **Project 1**

WHO/UNICEF JMP (Joint Monitoring Programme for Water Supply, Sanitation, and Hygiene) Estimates on the use of water dataset for 2020.

* + *This dataset has a total of 16 features (or columns), 12 of which are service-level percentage shares.*

***NB:*** *Refer to data dictionary for more in-depth descriptions of what each feature or column represents*

**Project 1: Understanding the Data**

The first part of the project covers understanding the features in our dataset. This involved:

* Data cleaning
  + value
* Investigating population size (**New**: **value\_cnt**)
  + A summary of the dataset population size and estimated world population, which includes urban percentage share and the percentage difference between all of the features.
  + A line chart of the national population versus the urban and rural population shares.
* Investigating access by area (**New**: pop\_u\_val, pop\_r, pop\_n (m))
  + The maximum, minimum, mean, mode, median, first and third quartiles, the interquartile range, and the standard deviation for each of the 12 water access features.
  + A box and whisker plot for all 12 water access features.
* Investigating access by population size (**New: wat\_bas\_n (rounded))**
  + Three 100% stacked column charts, one each for national, rural, and urban population size or percentage versus the four different service levels.
* Investigating access by income group (**New**: pop\_u (rounded), pop\_r (rounded))
  + A pivot table for income group versus the sum of the national population and the averages for the urban population, basic, limited, unimproved, and surface access shares.
  + A visualisation (of choice) for the income group versus the different average shares in the created pivot table.

**If possible: Input MCQ as question to answer [some of them] or do them as insights drawn**

**Project 2: Transforming the Data**