Word doc sections:

-Dataset A-Dean

- Dataset B-Tshiamo

- Dataset C- Caitlin

- Dataset D-Zoë

Each Word section contains:

Intro:

* explain shortly what is the theme of your dataset group(water/literacy impact on health)

Modelling:

* Start with Power BI visualizations to explore relationships, clusters, etc.
* Choose two models and explain why they are suitable: using the visualizations from power BI-> NB
* **Add parameters -also explain why**

Evaluate model performance:

* After modelling in R, visualise the datasets in power BI.
* Use performance statistics to compare the two models.
* Justify why you chose the best performing model.

R markdown:

1. Show all the code except libraries and warnings.
2. Change your dataset if necessary
3. Do the modelling and add important comments- all graphs/visualization needs to go in the word doc.
4. Include performance stats of each model
5. Knit your script to a pdf and send it on **teams**.

Notes:

Avoid using lists and bullets & show your knowledge of the data.

Teams meeting @ 12 on Friday: share your work so we can fill in any gaps and fine tune

If you need assistance please ask before Thursday

We will submit:

Word doc: explaining the modelling process of group a-d

4 R scripts: modelling and performance stats of each dataset group.