Data Science Team, Data Scientist (Methods)
Mini Data Case

Expectations: Fit one or more models using the dataset provided and provide a written interpretation of your model results, including at least one final evaluation metric and explain why you chose the metric you did. Please keep written interpretations brief, no longer than a one or two paragraphs. *Hint: utilize the provided data dictionary and context below.*

You may use any programming language to complete the task. Please include all code used during all phases of your analysis, including any visuals or tables that you think are helpful to convey your process and interpretation of the results. Along with the code, please include a brief explanation of your logic and thought process. We want to understand the way you think through problems and trade offs.

This mini case should take no more than 2 hours and is meant to test your abilities to both apply modeling concepts to spatial data **and** interpret the results and context behind the model. Your choice of model(s) is less critical than an explanation of why you selected them, how you evaluated them, and what they are telling you. A good choice for presenting your analysis may be a rendered Jupyter notebook or Rmarkdown file. Note that your code and analysis will be used only for evaluation purposes.

Please submit your data case via a private github repo and give permissions to the following users:

- ali-filipovic
- Sramaswamy-fraym
- Sandevaj
- jsgro-fraym
- cpaton8
- EChebelyon
- jtanwk
- Lmcindewar
- J-Levitt

Context: The dataset provided was derived from a 2016 DHS survey form Nigeria. The variable to be modeled, 'median_spend', represents the weighted average of median household spending within a DHS survey cluster. The columns 'lat' and 'long' contain the geographic coordinates of the cluster centroid. Finally, the dataset includes many possible predictor (independent) variables extracted from various raster datasets at the locations of the cluster centroids. Your task is to model the `median_spend` response given the associated feature set.