

## Feature Selection and Dimension Reduction

In this project, you will investigate feature selection and dimension reduction on ONE of three datasets.

**Due Date:** December 5, 2020, by midnight.

**Datasets:** *algae\_blooms.csv*, *HR\_2016\_Census\_simple.csv*, *GlobalCitiesPBI.csv*

### Tasks:

For ONE of the datasets:

1. Establish 2-3 questions that you could try to answer with the data.
2. Based on your questions, and using filter/wrapper/embedded, supervised/unsupervised feature selection methods, provide 3-5 subsets of features that would do a good job of representing the original dataset (use some of the methods described in class, but other methods as well, such as described in the *Feature Selection and Dimension Reduction* review article).
3. Learn at least 3 (reduced) manifolds for the original data (consult the *Feature Selection and Dimension Reduction* review article for suggestions and examples).
4. How would you validate your results?

**Hints/Suggestions:** this project will require a fair amount of independent learning. Do not hesitate to ask me questions as needed.