CS614 – Prof. Berardi
Final Project Description

Your final project for this course will be to download a data set and analyze it in accordance with the procedures we've been developing throughout the semester. This may include descriptive and inferential statistics, visualizations, interactive GUIs, machine learning, hierarchical modeling, and geospatial approaches. You should have clear, reasonable research questions, address outliers and missing data, and present the results as if you were presenting them to a supervisor, with some knowledge of data science.

The first step is to select a dataset from the UCI Machine Learning Repository. Your data should have the following attributes:

- Year between 2010 and 2019
- Multivariate data
- A minimum of 10 features
- A minimum of 500 instances/observations

Alternatives to the UCI repository or to these criteria will be assessed on case-by-case basis and should be directed to me ASAP.

There will be three components to this assignment:

1.) Select your data set, develop 2-3 research questions, and determine the approach most likely to achieve your goals. Summarize this document on the worksheet provided at the end of this document and submit to me by **November 22, 2020**.

2.) Prepare a written report summarizing your analytic approach and findings. The report can be written in LaTex, Word, or R Markdown. Where appropriate (e.g. to demonstrate a novel approach) you may insert code into your writeup, but generally this is not required and code should be provided in a separate, auxiliary document. There are no strict requirements for length, but somewhere within 10-20 pages (including figures/tables) will likely suffice for satisfying the requirements of this assignment. Your writeup should include an *Introduction* where you introduce to your data and describe the importance of your research question. You should then have a *Methods* section followed by a *Results* section. Lastly, there should be a *Conclusions* section, where you summarize your project and provide ideas for future work. Include figures, tables, and references as needed, although do not feel a need to perform a formal literature review that would produce a large number of references.

   Your report should be well-organized and use proper grammar. You will be graded on these criteria as well as the appropriateness of your analytic approach for your research question and your ability to convey your results. An electronic copy of your report and code (separate documents) must be uploaded to Canvas by **10:00AM** on **December 17, 2020**.

3.) Present an 7-8 minute talk to the class summarizing your results. A 2-3 minute Q&A session will follow each talk. Please be careful to stay within your allotted time (I recommend a timed practice run). The talks will occur during the final exam period on **December 16, 2020** and the order of presentations will be randomly determined.