Final Project Overview

Work with your lab groups and use one of the areas below to answer a broad question(s) related to a given dataset. Some dataset resources are list below for you to potential use, but you are also welcomed to use a dataset you have used in the past or is not a part of the listed resources. Do not use a dataset from class. Similar to the mid-term please present a cleanly knitted final presentation that walks the reader through your project step by step.

General topics we will/have covered that can be a focus of the final project (feel free to combine topics or extend them)

* Data Visualization – interactive or static
* Text Mining
* kNN
* Clustering - Kmeans
* Decision Trees
* Ensemble – Random Forrest or other Ensemble Tree Method

Generate a publishable Rmarkdown document with the following sections:

* Question and background information on the data and why you are asking this question(s). References to previous research/evidence generally would be nice to include. – You must present your question to me during office hours, either next week on 26th or the following week on the 3rd
* Exploratory Data Analysis – Initial summary statistics and graphs with an emphasis on variables you believe to be important for your analysis.
* Methods – Techniques you are using to address your question and the results of those methods.
* Evaluation of your model – Select appropriate metrics and explain the output as it relates to your question.
* Fairness assessment – if necessary, should you happen to have any protected classes.
* Conclusions – What can you say about the results of the methods section as it relates to your question given the limitations to your model.
* Future work – What additional analysis is needed or what limited your analysis on this project.

Potential Data Sources:

Google Dataset Search: <https://datasetsearch.research.google.com/>

Covid 19 - <https://github.com/XinerNing/CGDV.github.io/blob/master/dataSource/index.md>

data.world - <https://data.world/>

UCI ML Repository - <http://archive.ics.uci.edu/ml/index.php>

Data is Plural - <https://docs.google.com/spreadsheets/d/1wZhPLMCHKJvwOkP4juclhjFgqIY8fQFMemwKL2c64vk/edit#gid=0>