Brad Pawlow, Sebastian Reid, Noah Slosberg

5/8/23 - 5/12/23

STAT 390

Progress Report 4

**Goals from Last Week:**

* Task 1, perform EDA on kaggle assignment and obtain model with public score less than 9.00 RMSE
* Task 2, discuss potential prediction questions for the MCMF data set
* Task 3, Incorporate new educational metrics to better model educational performance in CPS schools
* Task 4, Continue to clean data to account for new problems discovered in dataset
* Task 5, Begin working on prediction model

**Progress/Accomplishments**:

* Completed Task 1, with best public score of 8.33, working on fixing overfitting issues
* For Task 2, we will provide analysis on the number of educational programs that should be available based on school location and quality
* For task 3, we are still discussing what metrics best contextualize the issue
* Conducted 2 meetings with the team to discuss/generate EDA plots, direction of project, and data cleaning techniques.
* Completed Task 4 on schedule.
* Began Task 5, will continue to work on prediction model over the next week

**Problems/Challenges**:

* We are still in the process of understanding which educational metrics would be best to understand the problems in the CPS system. Many metrics can be biased and are unreliable, so we are reading the necessary literature to better understand this problem.
* Encountered delays in starting to perform PCA and factor analysis due to more focus on EDA steps
* Kaggle competition model is overfitting, we are working on trying to fix this

**Plan/Next Steps**:

* Continue working on Kaggle project, compensate for overfitting
* Better understand what educational metrics are required for project
  + Literature analysis of issue, understand what current research says about topic
* Remove new issues with row duplication found in data
  + Combine into multi-factor columns
* Make early prediction model so we can understand what factors play the most into the availability of educational programs
* Prepare for MCMF progress presentation in two weeks