This project is a based on a collaboration between Manoranjan Kumar and myself (Gary Sepulveda) as part of a final project in the University of Texas @ San Antonio Master of Science Data Analytics program. We chose a current (as of 8/7/18) Kaggle challenge (PASSNYC) as our topic. This basically is a challenge in the NYC school district to ultimately help underserved student achieve academic success by taking and being accepted into some the best specialized High School in the city. We used the following approach and is seen throughout the R code attached:

1. Gather Kaggle Data Sources and Merge with Other Data Sources
2. Determine Response and Initial Predictor Set
3. Feature Engineering
4. Data Cleaning & Preprocessing
5. Model Tuning
6. Multi Model Development
7. Model Comparison
8. Model Selection Criteria
9. Final Model - PLS
10. Conclusion (not in R code but was provided in the final presentation in class)