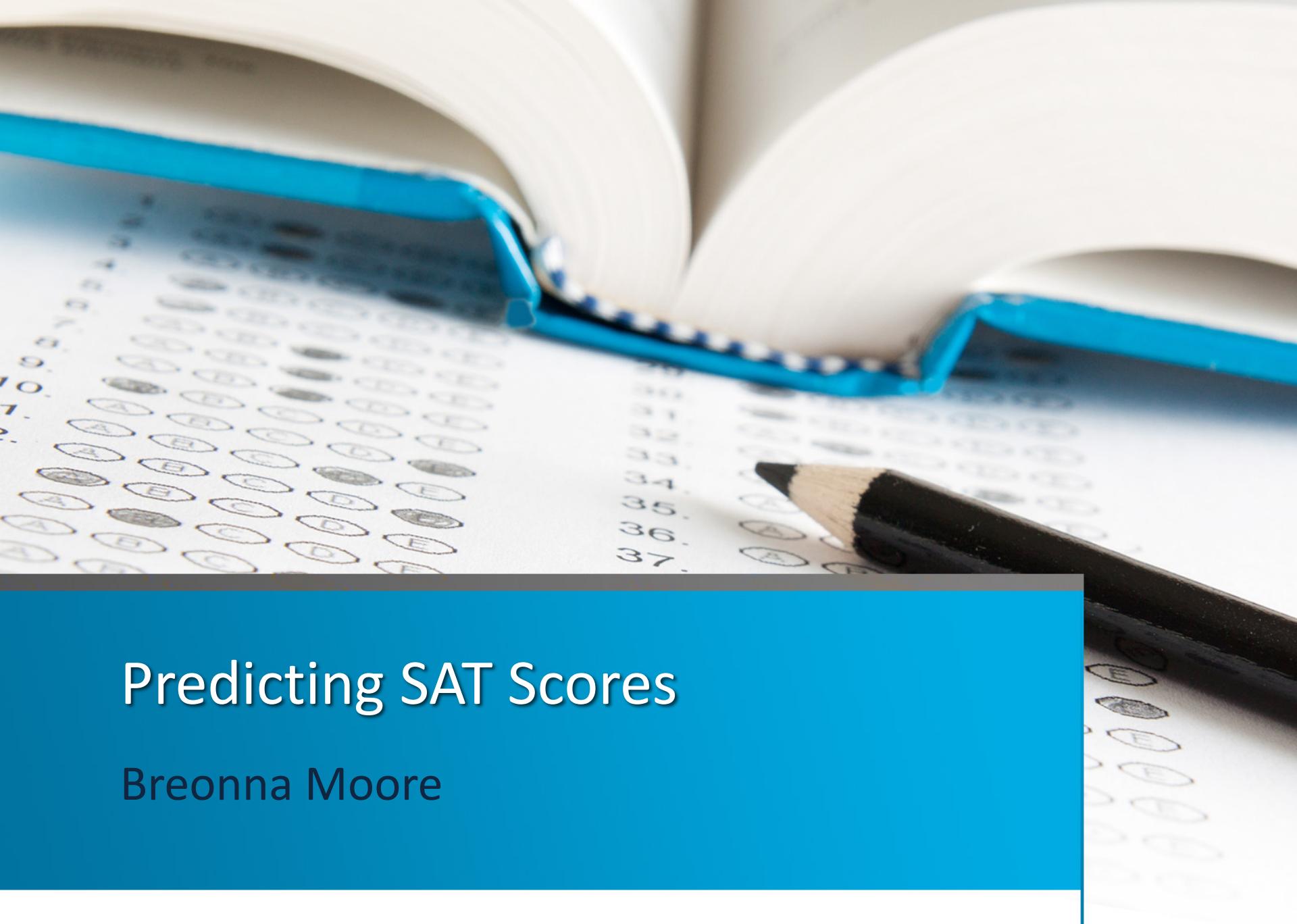


Predicting SAT Scores

Breonna Moore



Motivations

- Equal Opportunity
- Texas Public School Districts
- Non-profit SAT programs



Data Collection and Cleaning

Data

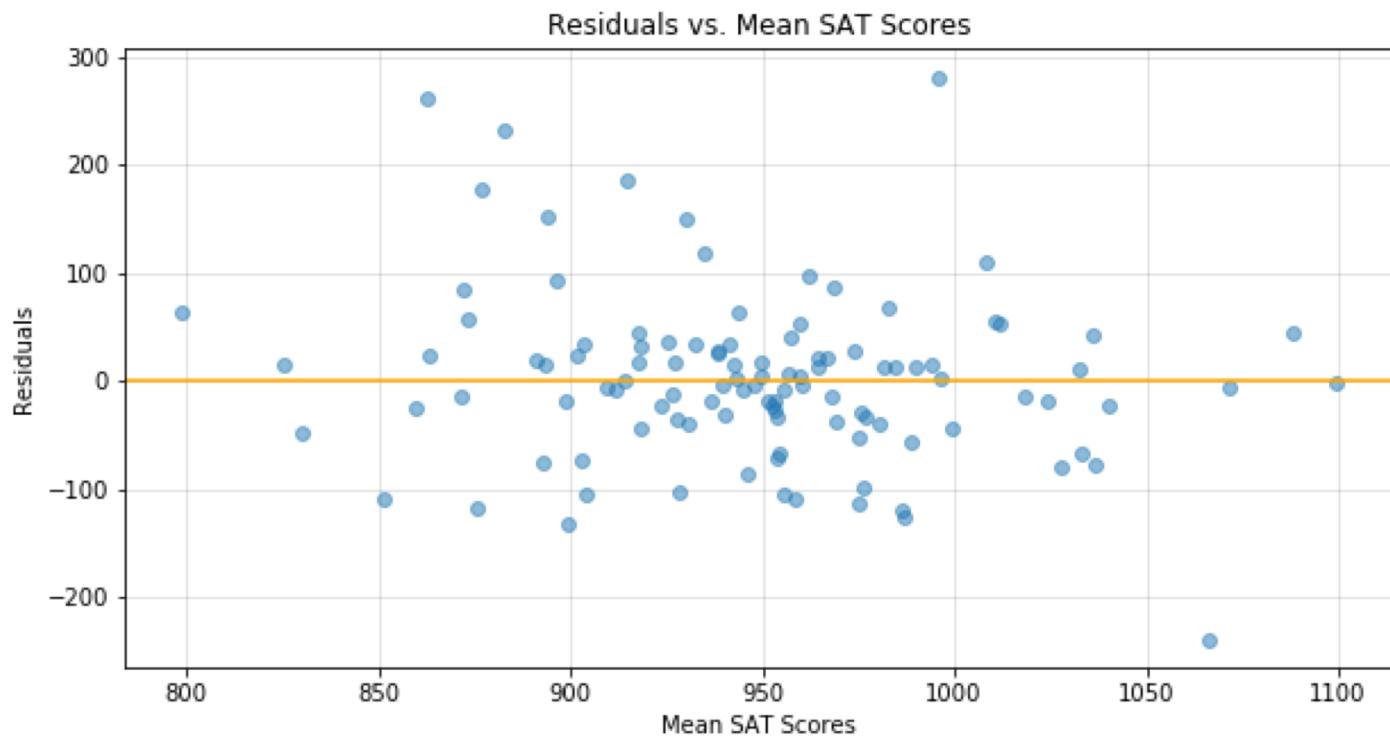
- Public School Districts
- Mean SAT scores
- Student Data
- School Data
- Parental Data

Tools

- BeautifulSoup
- Jupyter Notebook
- Pandas, NumPy
- Matplotlib
- Scikit-learn
- StatsModels

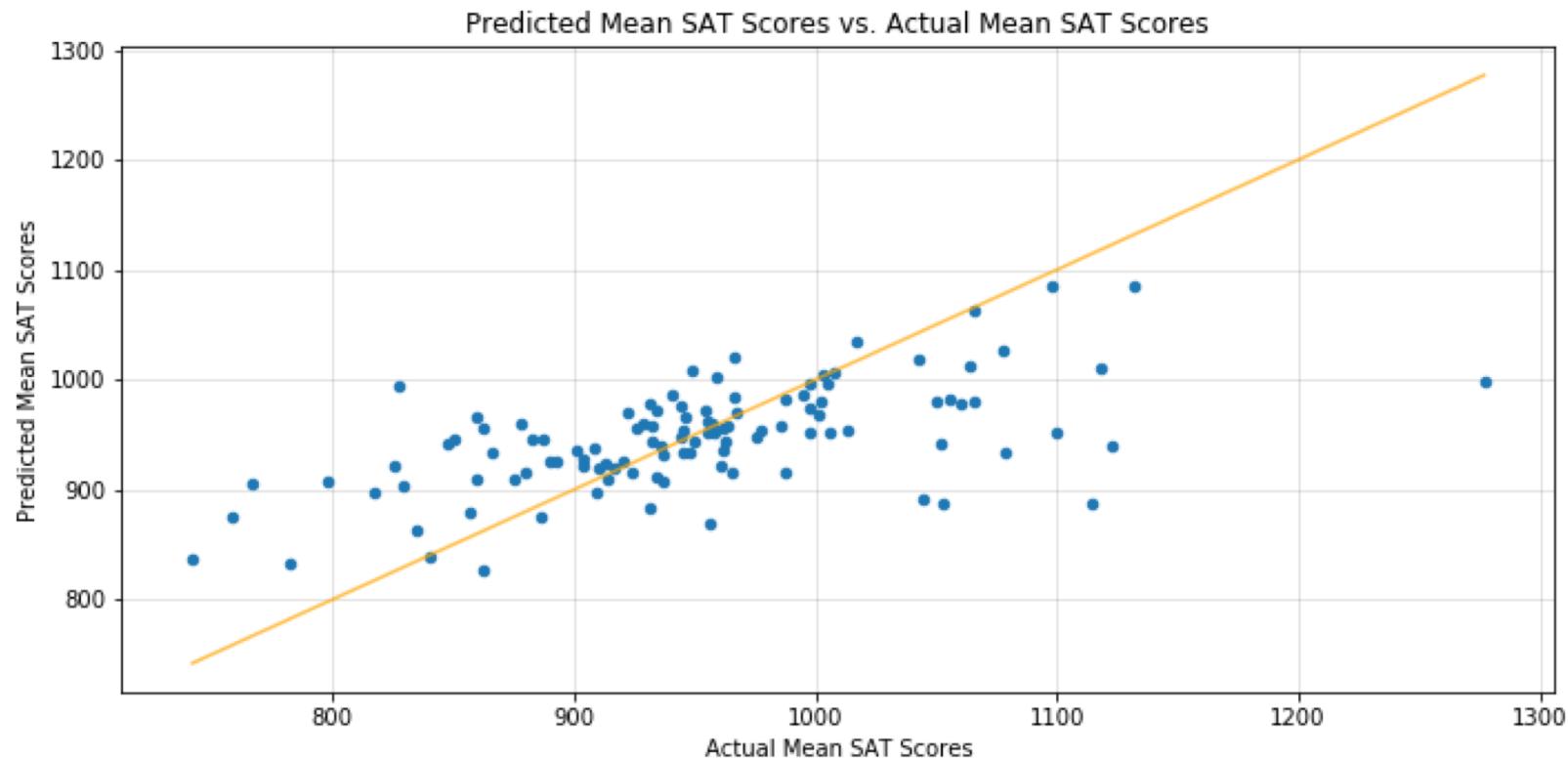
First Iteration

- Ordinary Least Squares Model
- Val. data $R^2 : 0.179$, Train data $R^2 : 0.472$
- MAE : 54.759



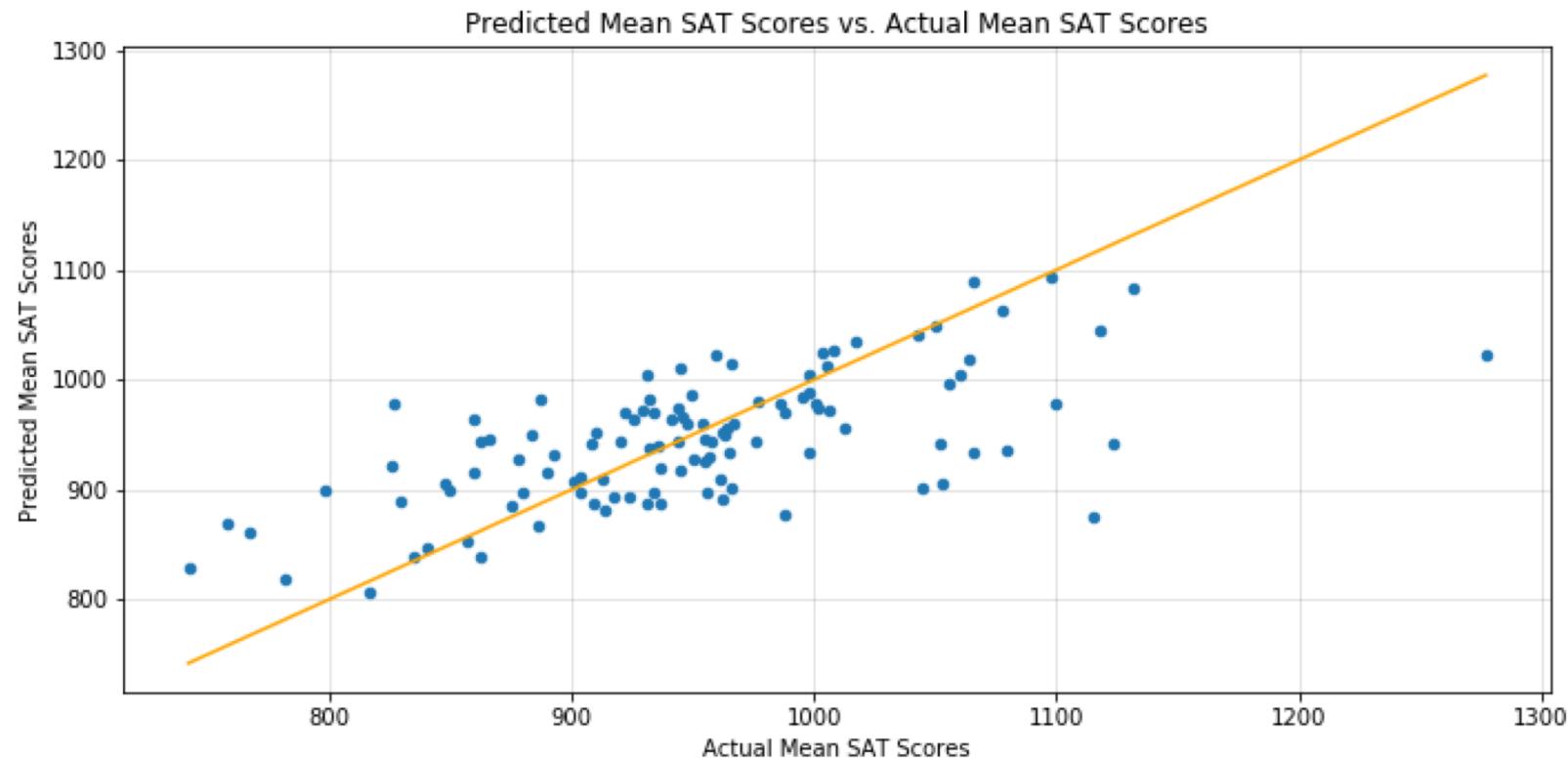
Second Iteration

- Lasso Model
- Val. data R^2 : 0.347, Train data R^2 : 0.428
- MAE : 48.815



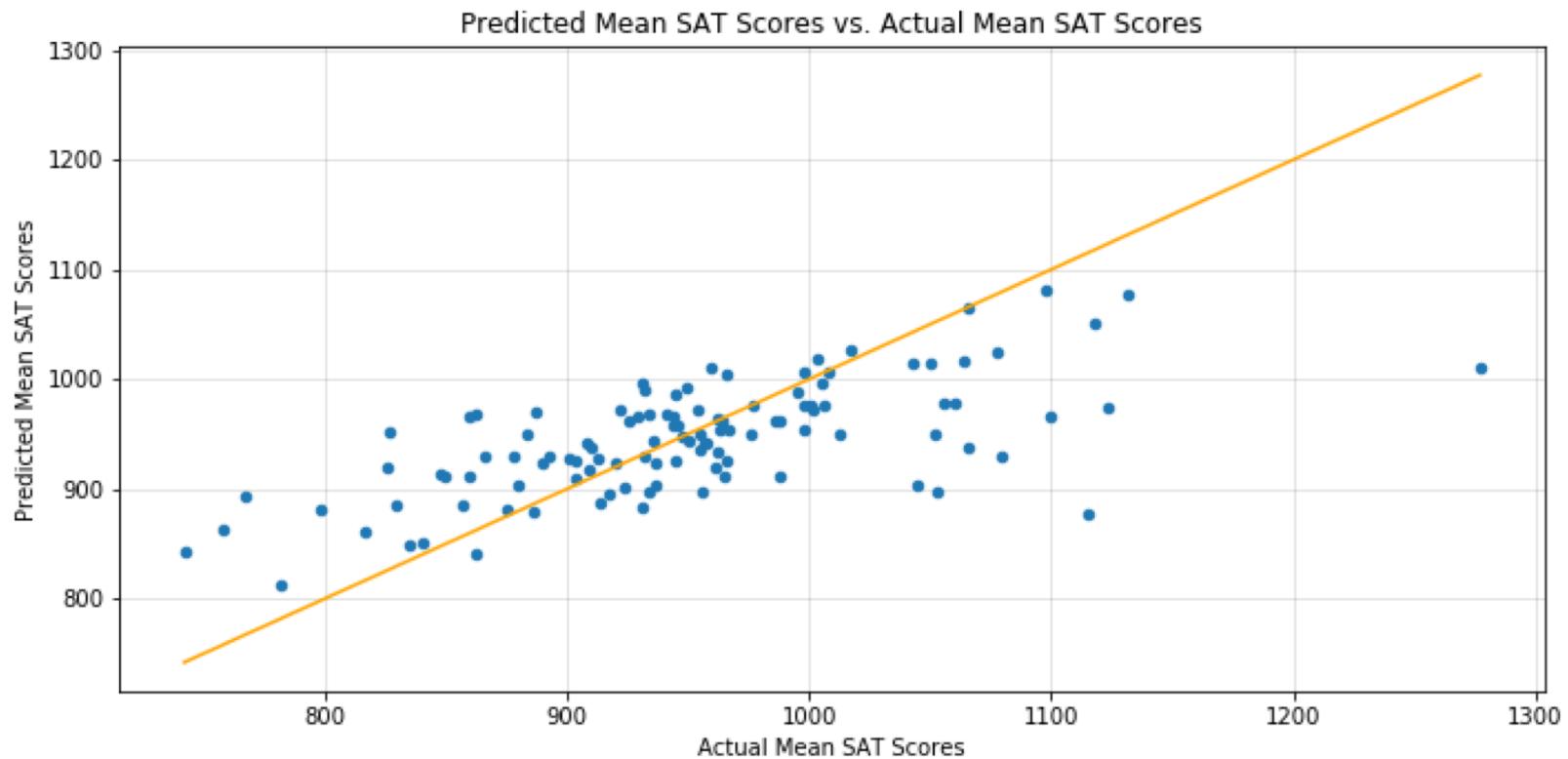
Third Iteration

- Ordinary Least Squares Model
- Val. data R^2 : 0.408, Train data R^2 : 0.441
- MAE : 47.211



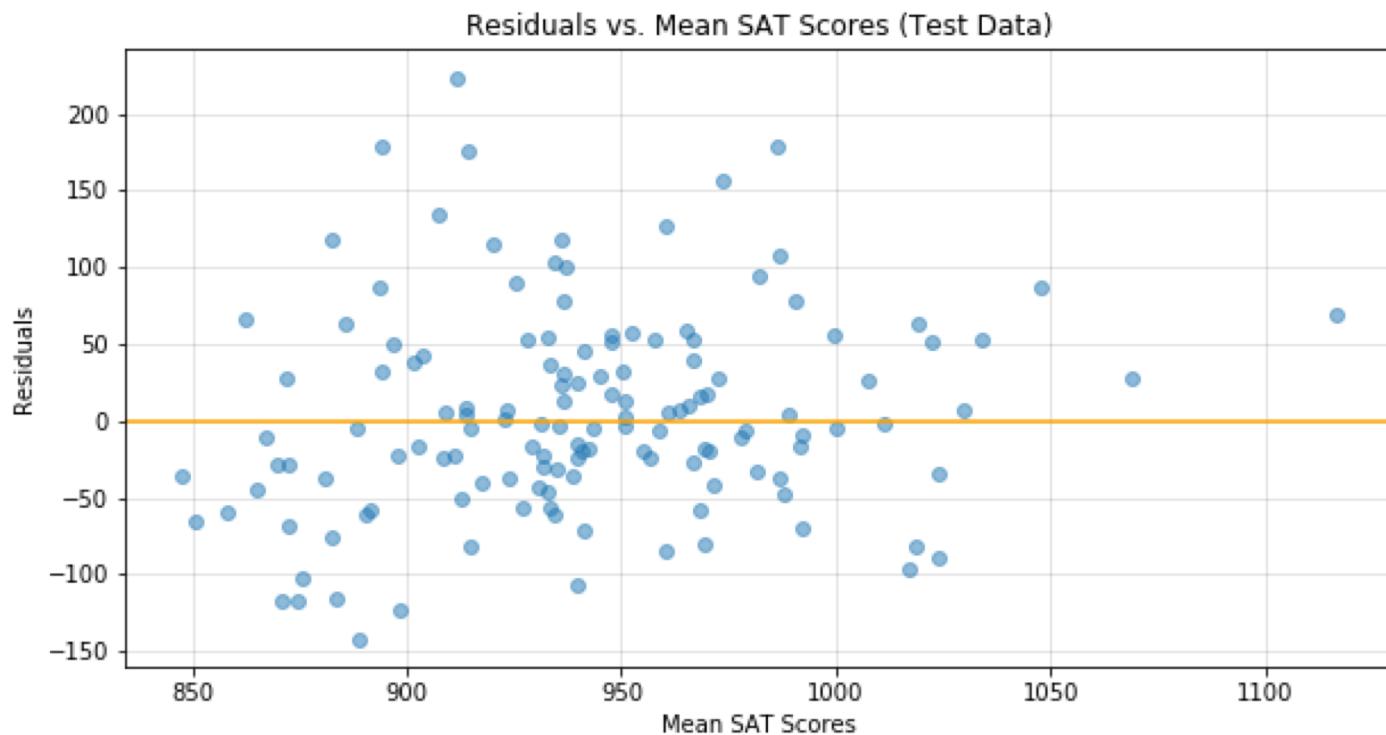
Fourth Iteration

- Lasso Model
- Val. data R^2 : 0.421, Train data R^2 : 0.439
- MAE : 46.739



Conclusion

- Test data R^2 : 0.442, Train data R^2 : 0.439
- MAE : 49.312, SAT scores: 400 – 1600
- Graduate, Earnings, Disadvantaged, Dropout

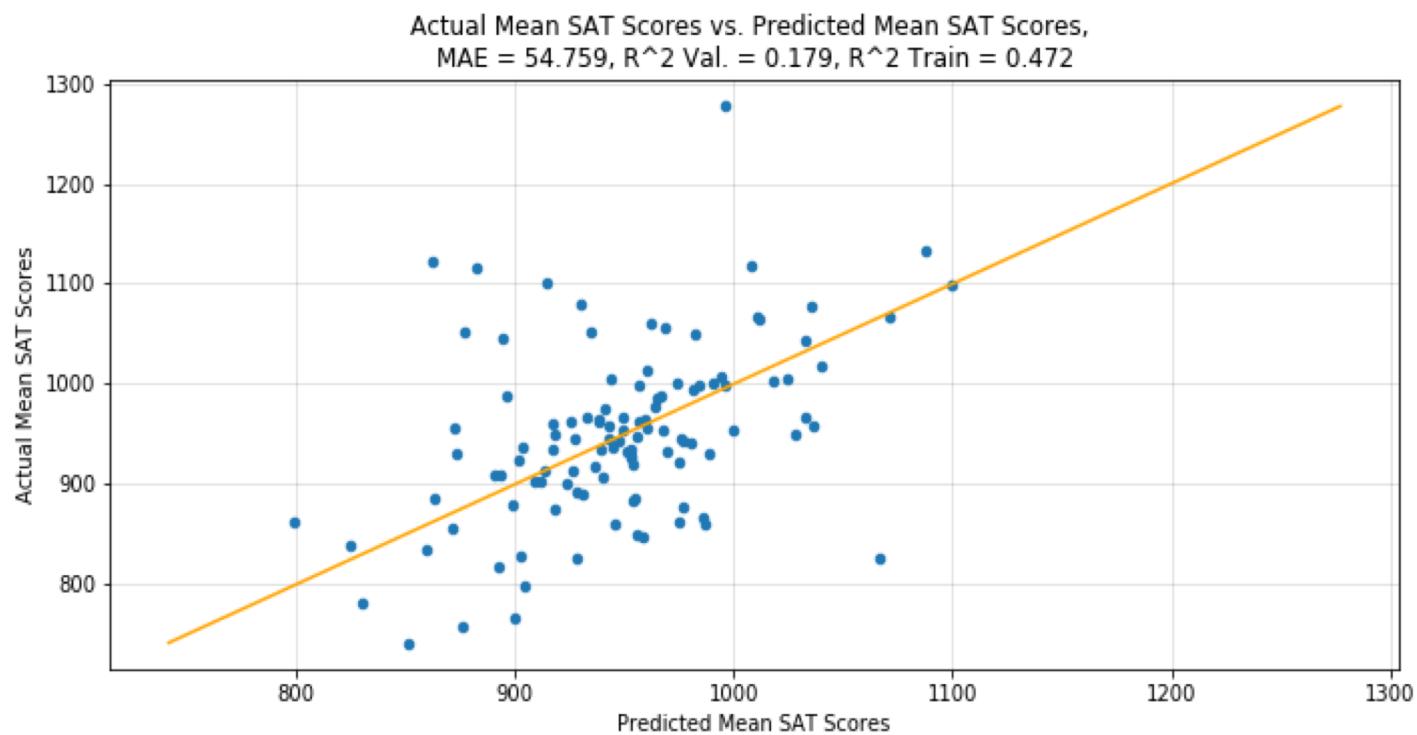


Future Work

- Individual Scores and Data
- More actionable features
- When to intervene?
- How to intervene?

Appendix

Model 1- OLS



Appendix

Model 4 - Lasso

