

Project Description The objective of the project is to predict performance in New York City public high schools and identify opportunities for schools to improve their system. We will use demographic data as well as information regarding financial and educational resources provided to each school to determine which characteristics have larger impact in academic performance.

Data We will use open source data provided by the New York City Department of Education and College Board. All information was recollected for the 2014-2015 school year:

- Scores: SAT scores for NYC public high schools as well as enrollment demographics.
- High School Directory: Basic descriptive information about each public high school including special classes, extracurricular activities and additional services.
- High School Programs: Directory of NYC High School programs.
- Quality Results: Quality metrics for NYC public high schools.
- Expenditure: Budget allocation to NYC high schools including teacher's salaries and text books.

Approach

1. *Data cleaning/wrangling*: The first phase of the project will focus on cleaning the data and combining all data sets to produce a unique data frame. This phase will also include the creation of new variables using text analytics.
2. *Data exploration*: A brief exploration phase will allow us to identify the relevant variables that should be used during the modeling stage.
3. *Modeling* This phase will consist of testing out a variety of analytical models. Potential areas to explore include cluster analysis, decision trees, regression models, and random forest. One approach includes a two stage modelling process, with an initial stage performing clustering analysis and then exploring different models within the clusters.
4. *Analysis and evaluation* The final stage will consist of model evaluation and comparison to pick the best model.

Tools The analysis part of the project will be conducted in R. Visualizations will be done in R and Tableau.

Impact The overall goal of the project is to provide a tool for developing an action plan that can effectively increase high school performance by identifying areas of growth and opportunity across the city.

Sources

- <https://www.kaggle.com/nycopendata/high-schools>
- https://www.nycenet.edu/offices/d_chanc_oper/budget/exp01/y2014_2015/guide.asp
- <https://data.cityofnewyork.us/browse?category=Education>