



Understanding Pennsylvania School Performance

Chris Chung
Springboard Capstone Project 3
February 2022



The Problem

PENNSYLVANIA SYSTEM OF SCHOOL ASSESSMENT (PSSA)

Every PA student in grades 3-8 attending district, charter, and cyber schools is assessed in English Language Arts and Math.

Need to understand the factors that affect school performance on the PSSAs.

- School Performance Profile
- Turnaround interventions
- School closures
- Charter school renewals

Goals

Investigate the magnitude of gaps in school performance and understand what factors are associated with school performance.

Create two models (ELA and Math) to establish relationship between the features of a school and the percentage of students that are proficient or advanced in that school.

Select one school and use the model to produce counterfactual explanations.



Data Preparation

Data Acquisition

Data Wrangling

Data Cleaning

Data Preparation

Data Acquisition

Data Wrangling

Data Cleaning

Two datasets from [Future Ready PA Index](#).

SCHOOL PERFORMANCE DATA FROM 2018-2019 SCHOOL YEAR

- School level
- General school information
- Demographics
- Learning profile
- Socioeconomic makeup
- Target: Percent Proficient or Advanced in ELA
- Target: Percent Proficient or Advanced in Math

SCHOOL FISCAL DATA FROM 2018-2019 SCHOOL YEAR

- School level
- Personnel and non-personnel expenditures per student from different sources (Local, State, Federal)

Data Preparation

Data Acquisition

Data Wrangling

Data Cleaning

School Name	Data Element	Value
Abington Heights MS	School Enrollment	1119
Abington Heights MS	Male	52.28
Abington Heights MS	% Regular Attendance	90.9
...
Abington Heights MS	% Proficient or Adv. in ELA	76.0
Abington Heights MS	% Proficient or Adv. in Math	53.9
...

Long Format

School Name	Data Element	Value
Abington Heights MS	Local Non-Personnel	2722.02
Abington Heights MS	Local Personnel	7036.48
Abington Heights MS	State Non-Personnel	1233.26
Abington Heights MS	State Personnel	3188.01
Abington Heights MS	Federal Non-Personnel	78.66
Abington Heights MS	Federal Personnel	203.33
...

Data Preparation

Data Acquisition

Data Wrangling

Data Cleaning

Wide format and join

School Name	School Enrollment	Male	% Regular Attendance	...	Federal Personnel	% Proficient or Adv. ELA	% Proficient or Adv. Math
Abington Heights MS	1119	52.28	90.9	...	203.33	76.0	53.9
ASPIRA Bilingual Charter School	365	54.79	68.1	...	1467.44	11.1	4.9
Barrett Elementary School	200	50.00	71.8	...	3259.03	42.2	22.7
Clarks Summit Elementary School	339	57.52	96.2	...	206.43	75.2	61.0
...

Data Preparation

Data Acquisition

Data Wrangling

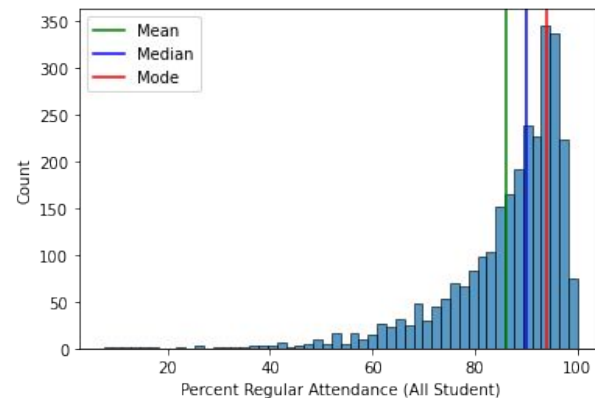
Data Cleaning

- Checked for duplicate rows and duplicate schools
- Dropped 5 rows with null values in almost every feature
- Dropped 126 rows with missing target values
- Imputed missing Percent Regular Attendance with mode
- 4 rows with negative expenditure values left alone
- Created binary variable for Title I status

Cleaned dataset:

2673 rows

32 columns

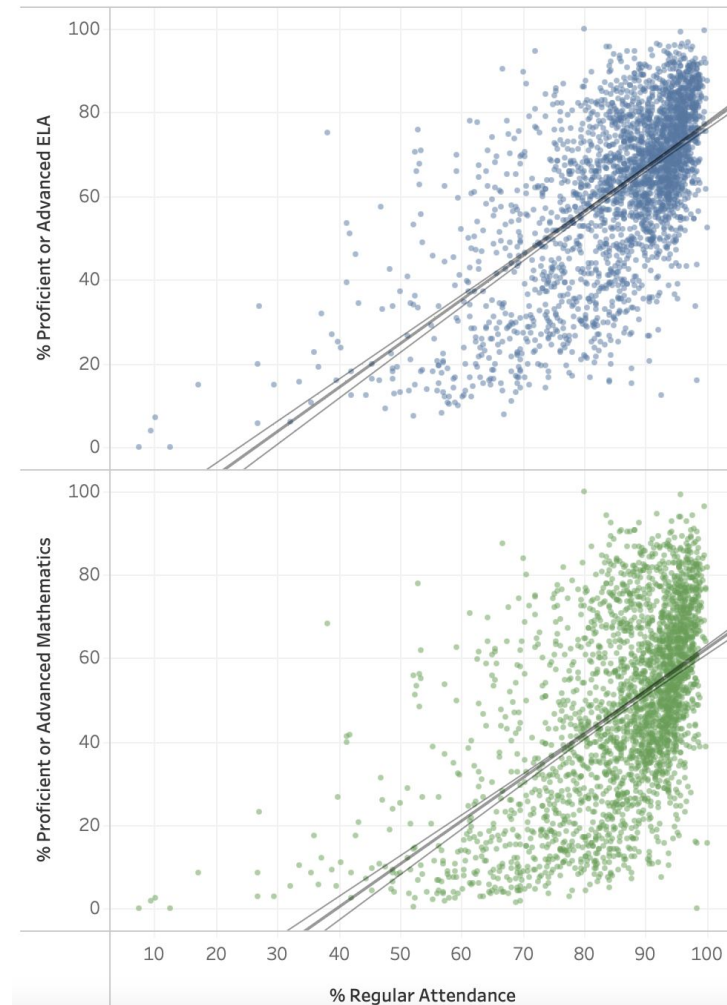




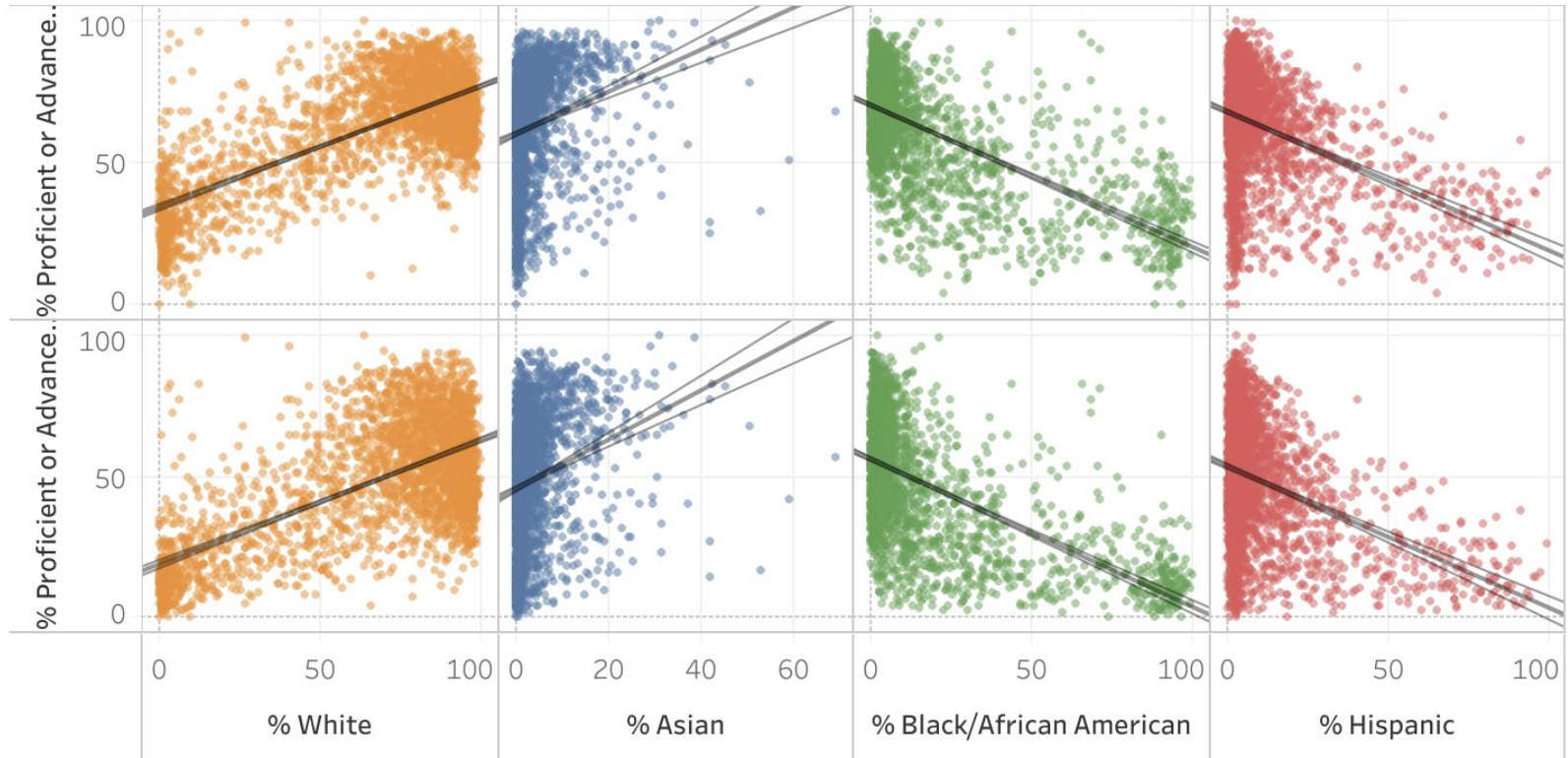
Exploratory Data Analysis

Percent Regular Attendance

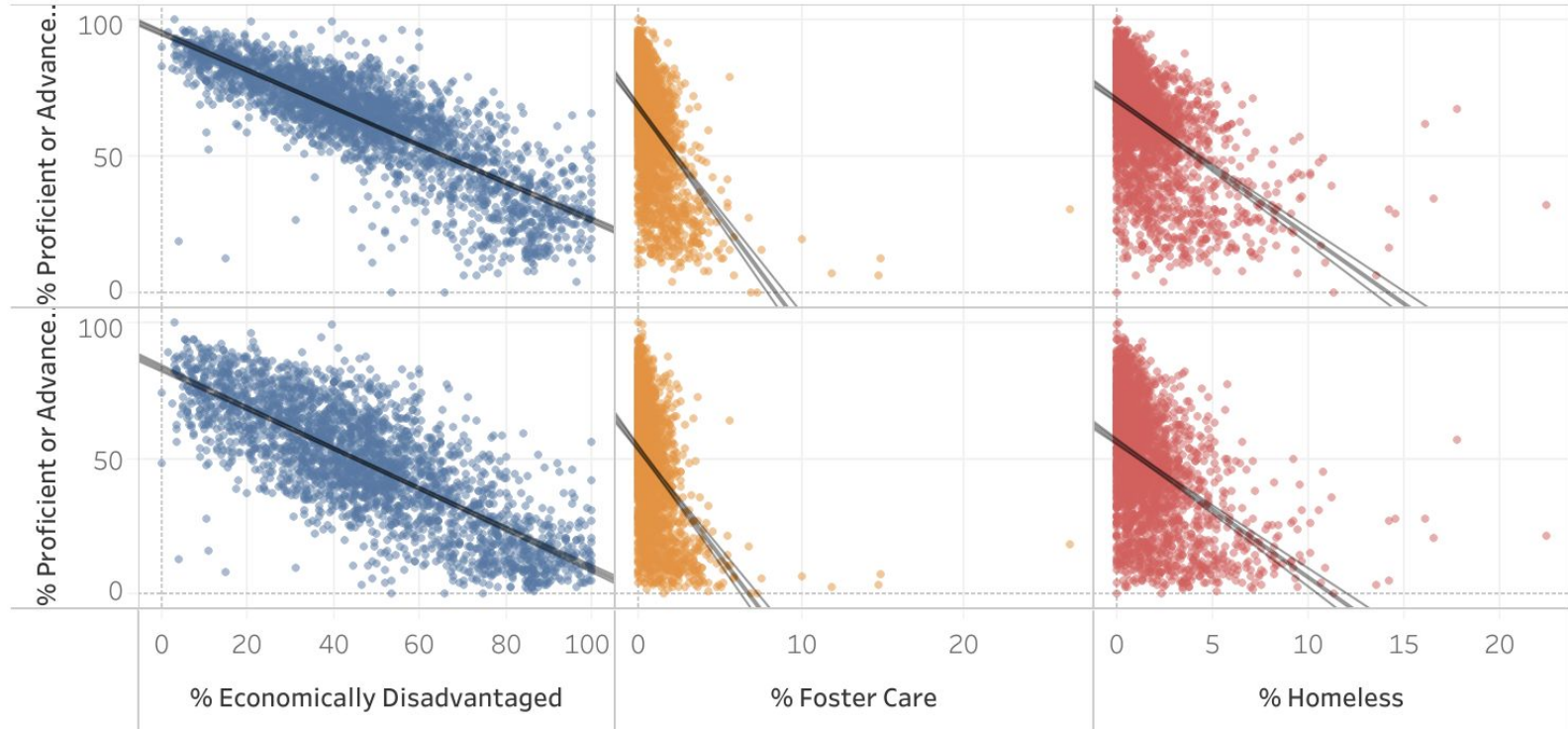
The percentage of students enrolled in a school for 90 or more school days and present 90 percent or more of these school days.



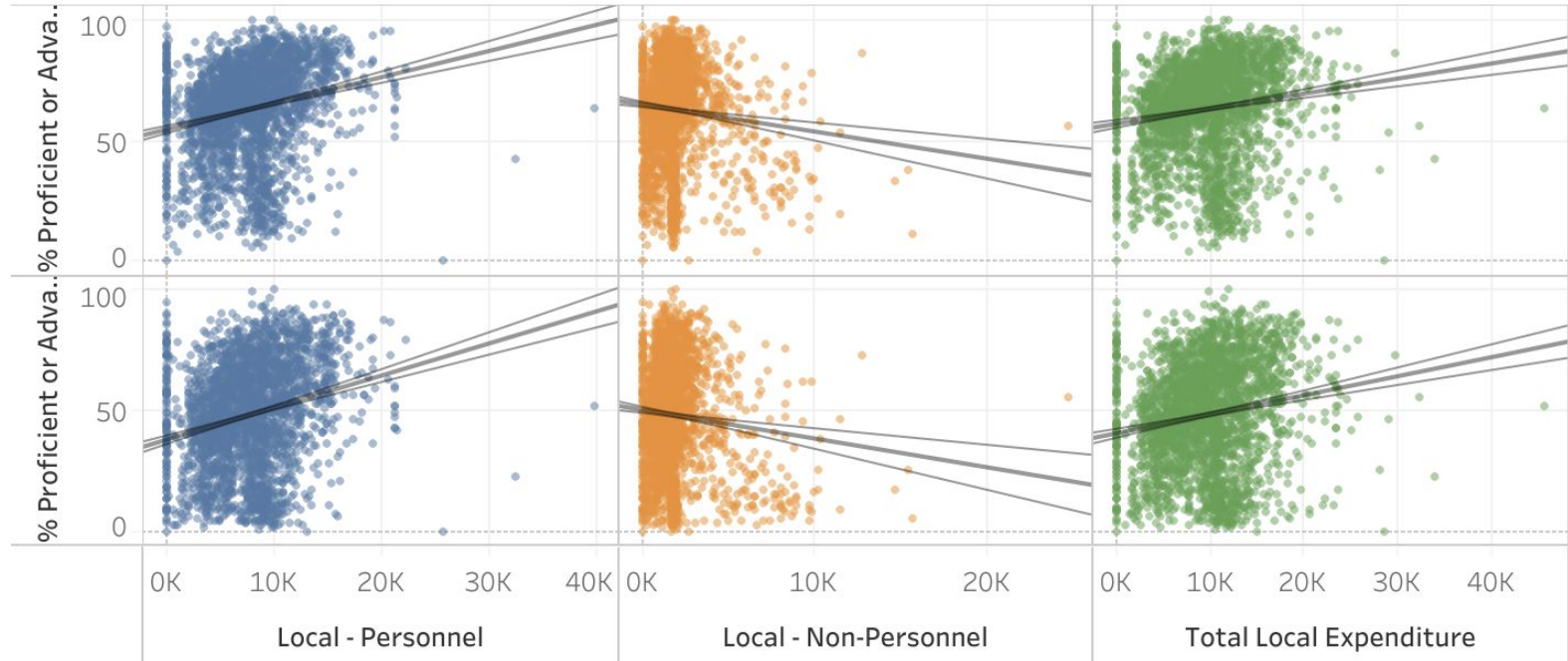
Racial/Ethnic Demographics



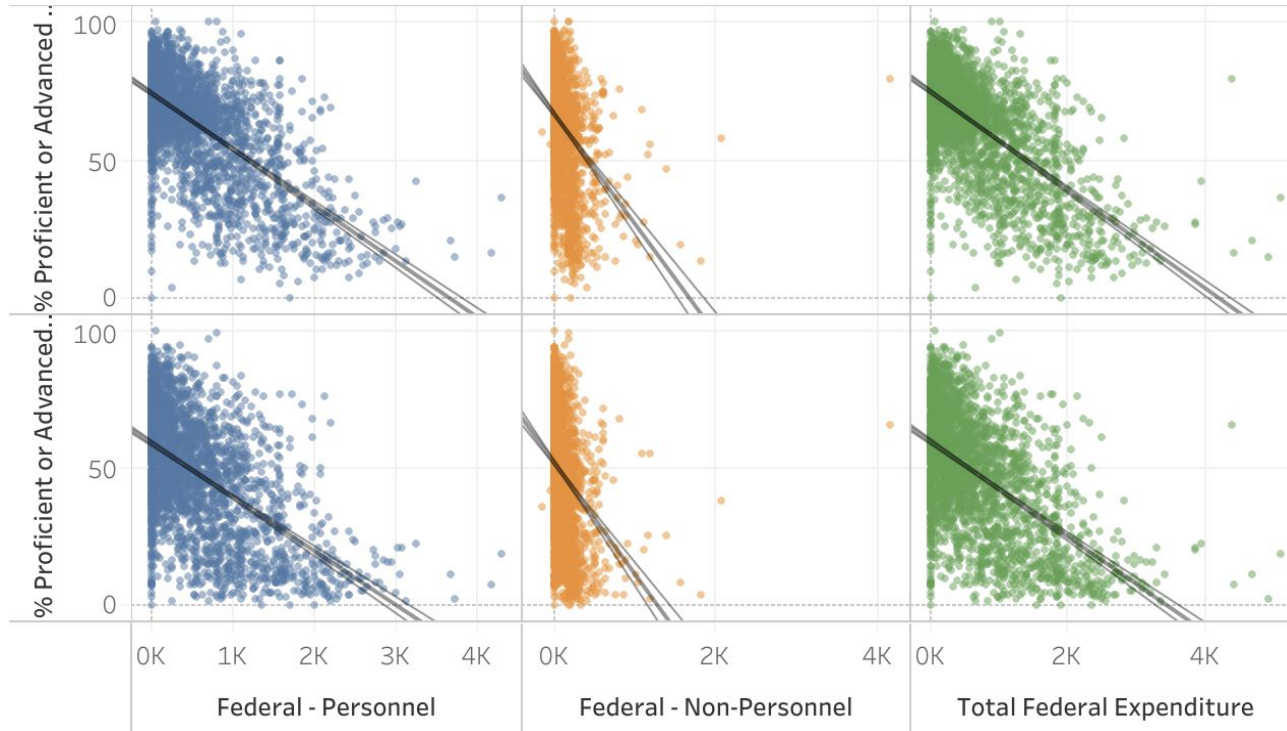
Socioeconomic Demographics



Local Expenditures



Federal Expenditures





Modeling

Baseline Model

Extended Modeling

Model Selection

Feature Impact

Modeling

Baseline Model

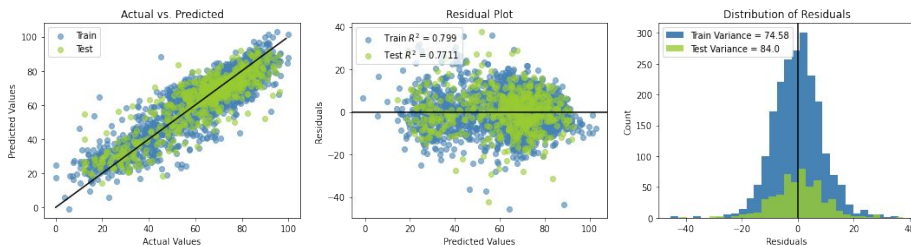
Extended Modeling

Model Selection

Feature Impact

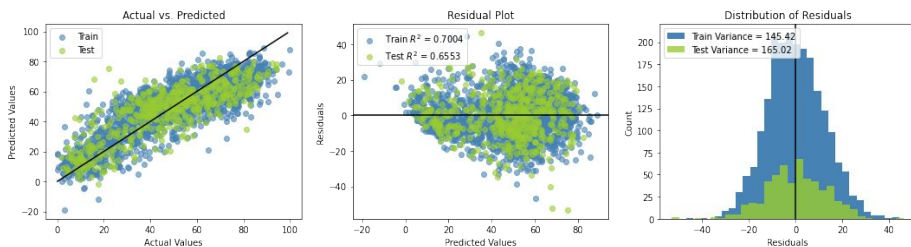
Baseline Model: Linear Regression

PREDICTING ELA PERFORMANCE



r^2	0.77
MAE	6.95
95% Worst Residuals	-18.04 17.50

PREDICTING MATH PERFORMANCE



r^2	0.66
MAE	10.12
95% Worst Residuals	-24.41 23.48

Modeling

Baseline Model

Extended Modeling

Model Selection

Feature Impact

Search space for extended modeling



6 models to compare for ELA and Math

Modeling

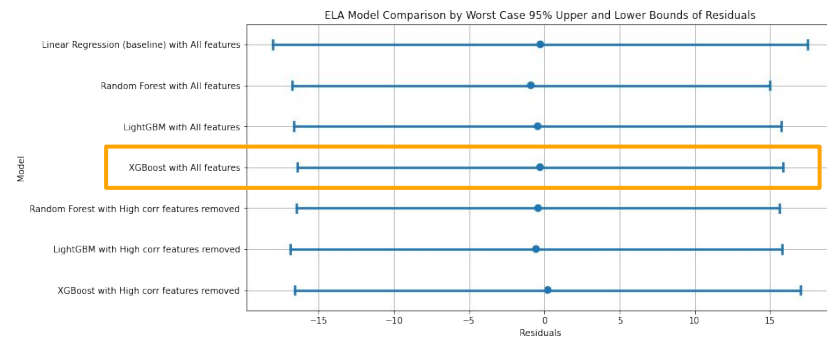
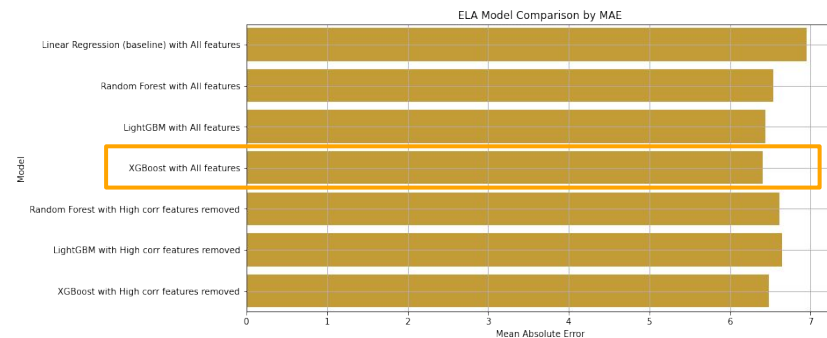
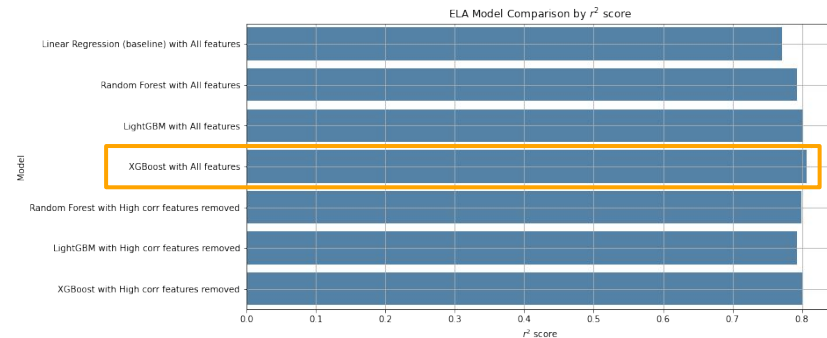
Baseline Model

Extended Modeling

Model Selection

Feature Impact

ELA Model Comparison



Modeling

Baseline Model

Extended Modeling

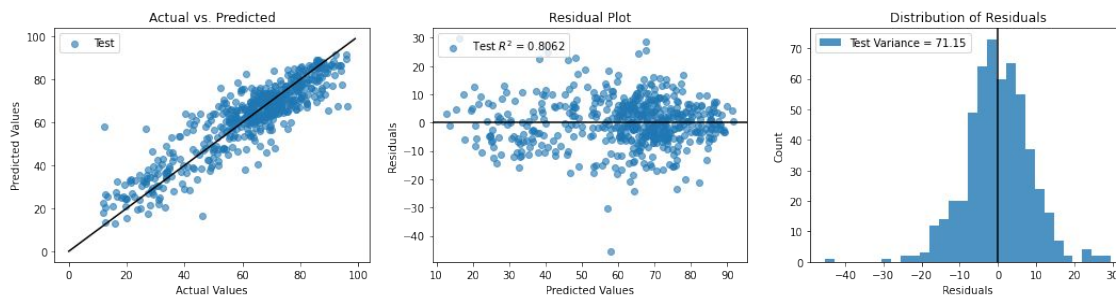
Model Selection

Feature Impact

ELA Model Selection

XGBoost with all features included

- r^2 score = 0.81
- MAE = 6.41
- Worst case residuals: $[-16.45, 15.87]$



Modeling

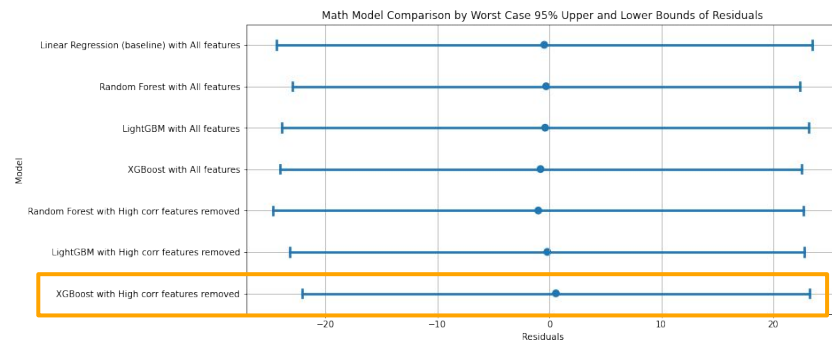
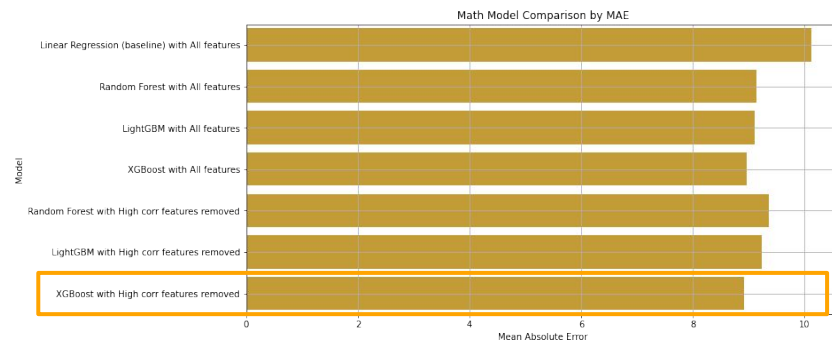
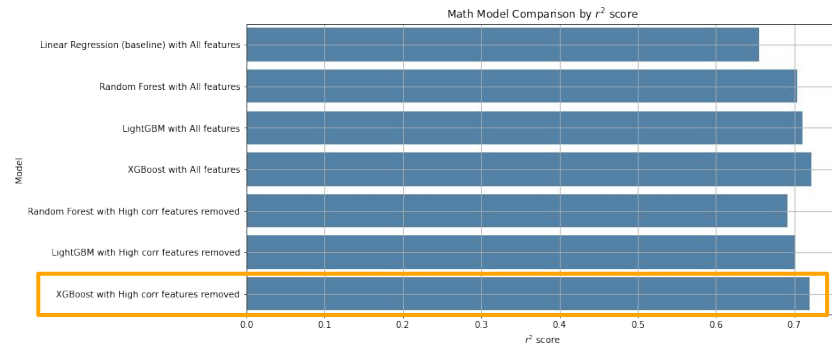
Baseline Model

Extended Modeling

Model Selection

Feature Impact

Math Model Comparison



Modeling

Baseline Model

Extended Modeling

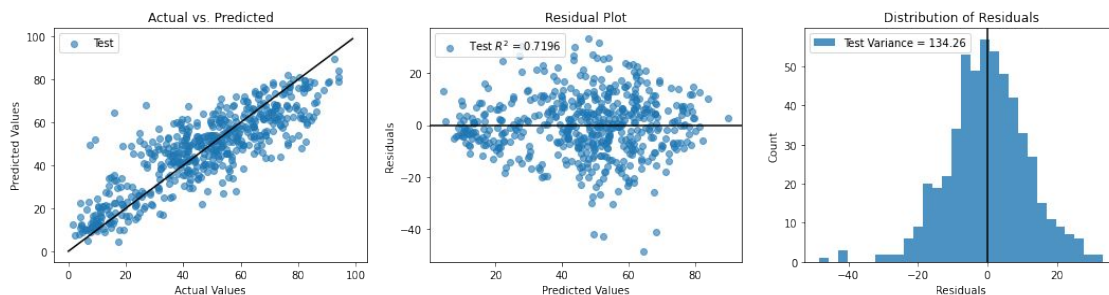
Model Selection

Feature Impact

Math Model Selection

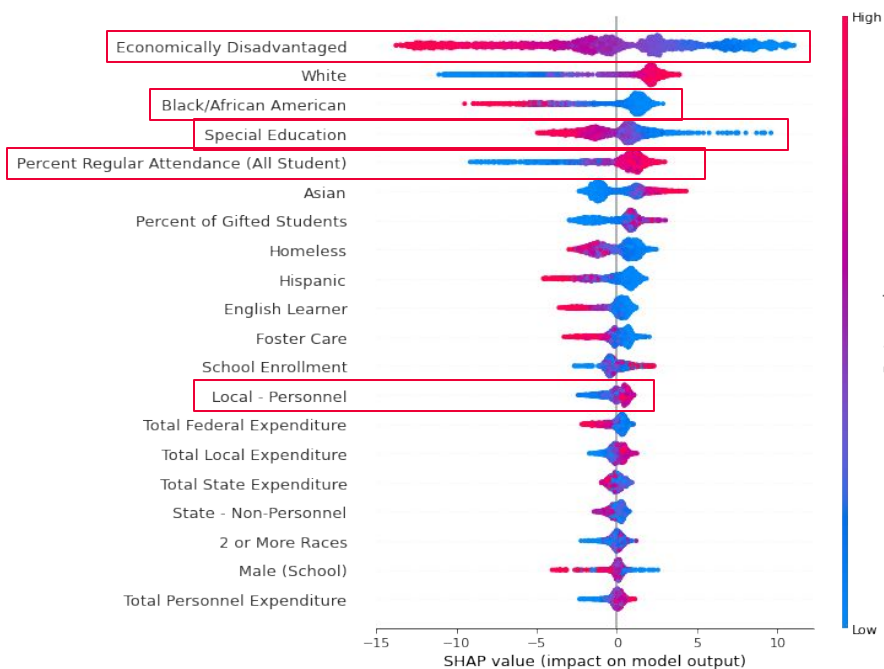
XGBoost with high correlation features removed

- r^2 score = 0.72
- MAE = 8.91
- Worst case residuals: [-22.09, 23.24]

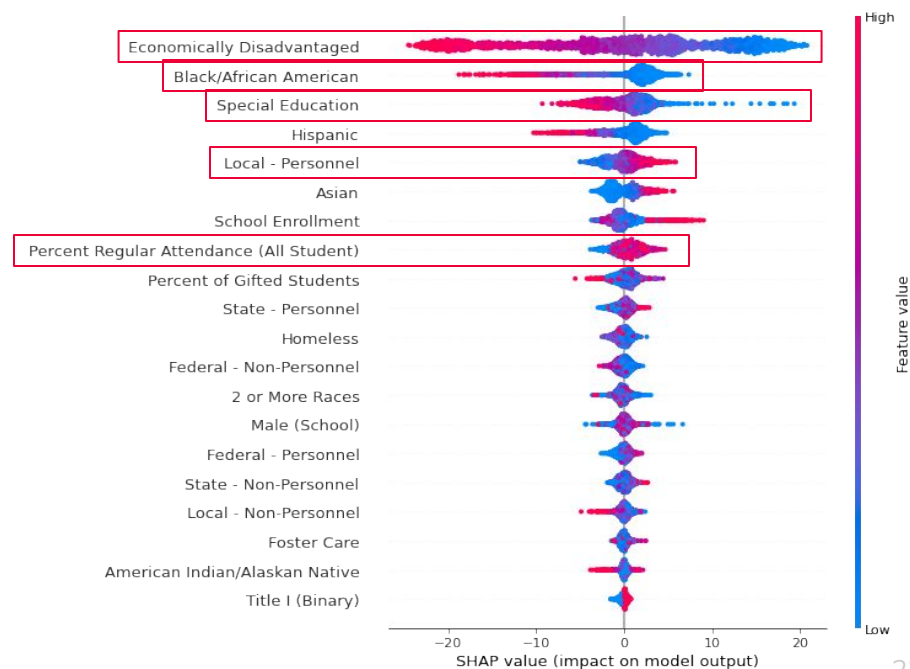


Feature Impact

ELA Model



Math Model





Counterfactual Explanations

Counterfactual Explanations

Approach

Observation:

0% proficient or advanced in both ELA and Math

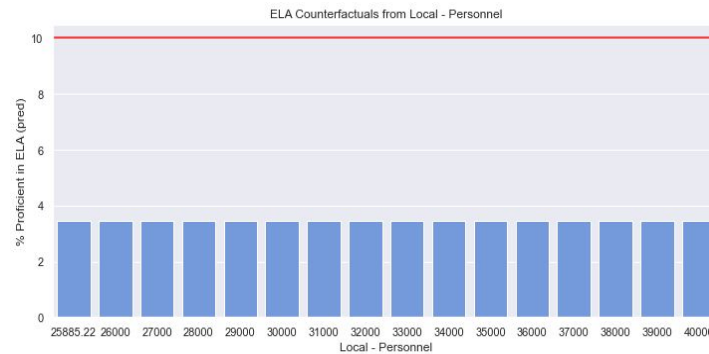
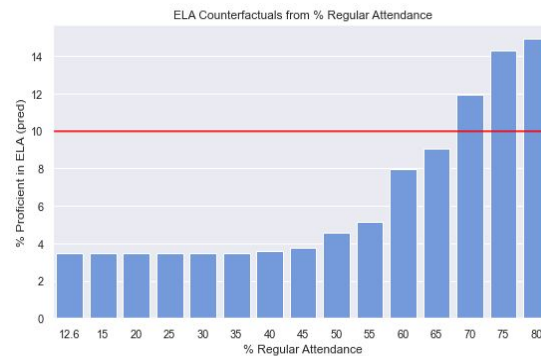
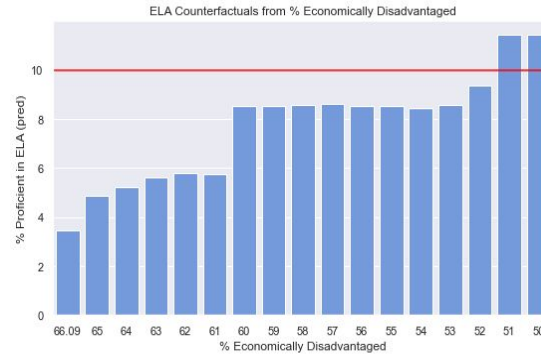
Goal: Can we find alternative scenarios so that this school can reach 10% (or as close to 10% as possible) proficiency in both ELA and Math?

Features to change:

- Economically disadvantaged (66.09)
- Percent Regular Attendance (12.60)
- Local - Personnel (25885.22)

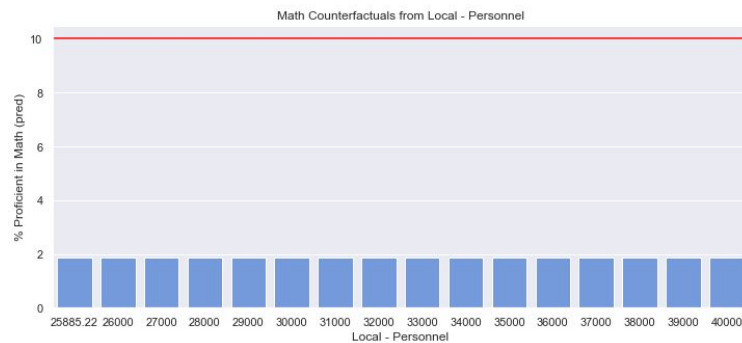
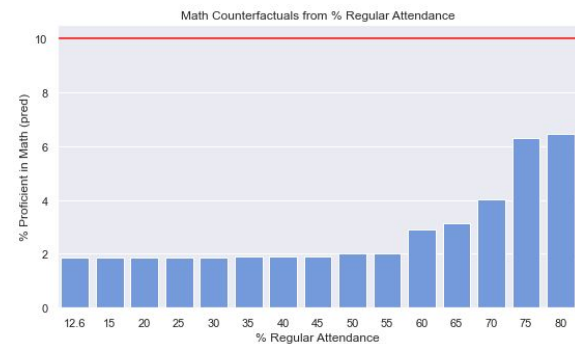
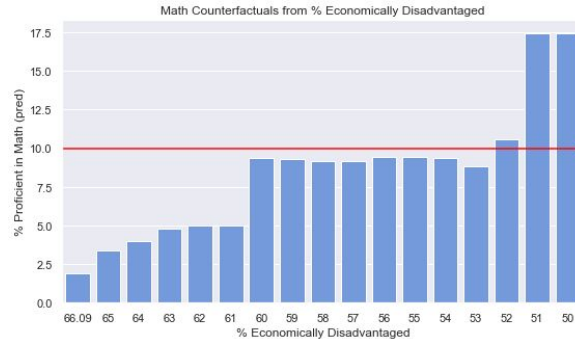
Counterfactual Explanations

ELA



Counterfactual Explanations

Math



Counterfactual Explanations

With Bayesian Optimization

ELA	Economically Disadvantaged	66.09%	↓	56.58%
	Percent Regular Attendance	12.60%	↑	21.23%
	Local - Personnel	\$25885.22	↑	\$28041.70
	% Proficient (or adv.) in ELA	0%	↑	8.59%

Math	Economically Disadvantaged	66.09%	↓	56.11%
	Percent Regular Attendance	12.60%	↑	13.68%
	Local - Personnel	\$25885.22	↑	\$27372.17
	% Proficient (or adv.) in ELA	0%	↑	9.43%

Future Work

- Remove outliers in expenditure variables
- Additional metric: adjusted r^2 score
- Counterfactuals with Bayesian Optimization, add more features to solution space
- Combine datasets from two different school years
 - Features: change in feature values from one year to another
 - Target: change in percent proficient or advanced





Recommendations

- Research and invest in ideas to improve the quality of life for impoverished neighborhoods, such as:
 - Offering social services
 - Providing funding for adults to attend trade school or other educational opportunities
 - Providing micro loans to families with small businesses to expand their ventures
- Connect with the families of local communities of underperforming schools to emphasize the significance of consistent attendance.
 - Identify any common barriers that families may face in preventing their children from attending school regularly.

Thank you!

Email: cchung106@gmail.com

LinkedIn: <https://www.linkedin.com/in/cchung106/>

GitHub: <https://github.com/cchung106/Capstone-Three>

Tableau:

https://public.tableau.com/app/profile/christopher.chung6215/viz/02_EDA_Tableau/PASchoolPerformanceEDA



Appendix

Sources

https://www.ncnewsonline.com/news/local_news/budget-report-on-pa-education-agency-provokes-spat-between-lawmakers-watchdogs/article_f332fd80-533c-5619-858d-fcecb3a8afb5.html

<https://shap.readthedocs.io/en/latest/index.html>

<https://github.com/fmfn/BayesianOptimization>

Image Sources

<https://themedialine.org/life-lines/paradise-lost-a-year-of-disrupted-education-in-the-middle-east/>

<https://elearningindustry.com/9-things-shape-future-of-education-learning-20-years>

<https://www.science.org/doi/10.1126/science.1235590>

<https://www.istockphoto.com/photo/innovative-learning-creative-educational-study-concept-for-graduation-and-school-gm1165957815-321001604>