



# Predicting Academic Success of NY Public Schools

Understanding Factors that Influence School Success

Lisa Berger, Merve Keskin, Leann McLaren, Nathan Schley

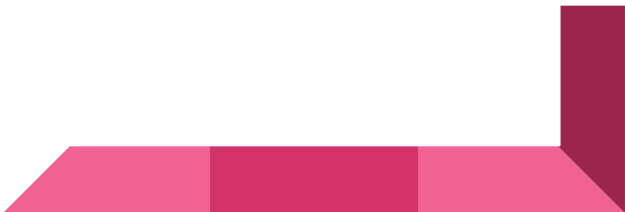
# Motivating Question

Using school-level features, can we accurately predict school-level academic success for New York public (including charter) schools?

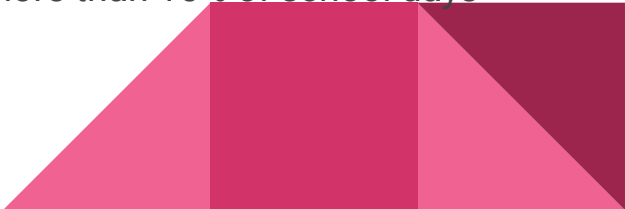
In addition to informing parents, this information is critical to informing state and local leaders who attempt to improve public schools through policy creation and its implementation.



# NYSED School Report Card Dataset

- All states in the US are required to report school-level data each year. NY State data is found at: *data.nysed.gov*
  - Proficiency for elementary assessments and high school Regents exams are published, for each school, going back to 1999.
  - Our resulting dataset includes school-level features for approximately 5000 New York public and charter schools for the calendar year 2021-2022.
- 

# Primary Features of Interest

- School funding
    - The yearly per-pupil spending
  - Teacher certification
    - The percentage of teachers who are teaching outside their field of certification
  - Teacher experience
    - The percentage of teachers who have fewer than 4 years of experience
  - Chronic absentee rate.
    - Percentage of students, out of total enrollment, who missed more than 10% of school days
- 

# Primary Outcome: Core Subject Performance Index

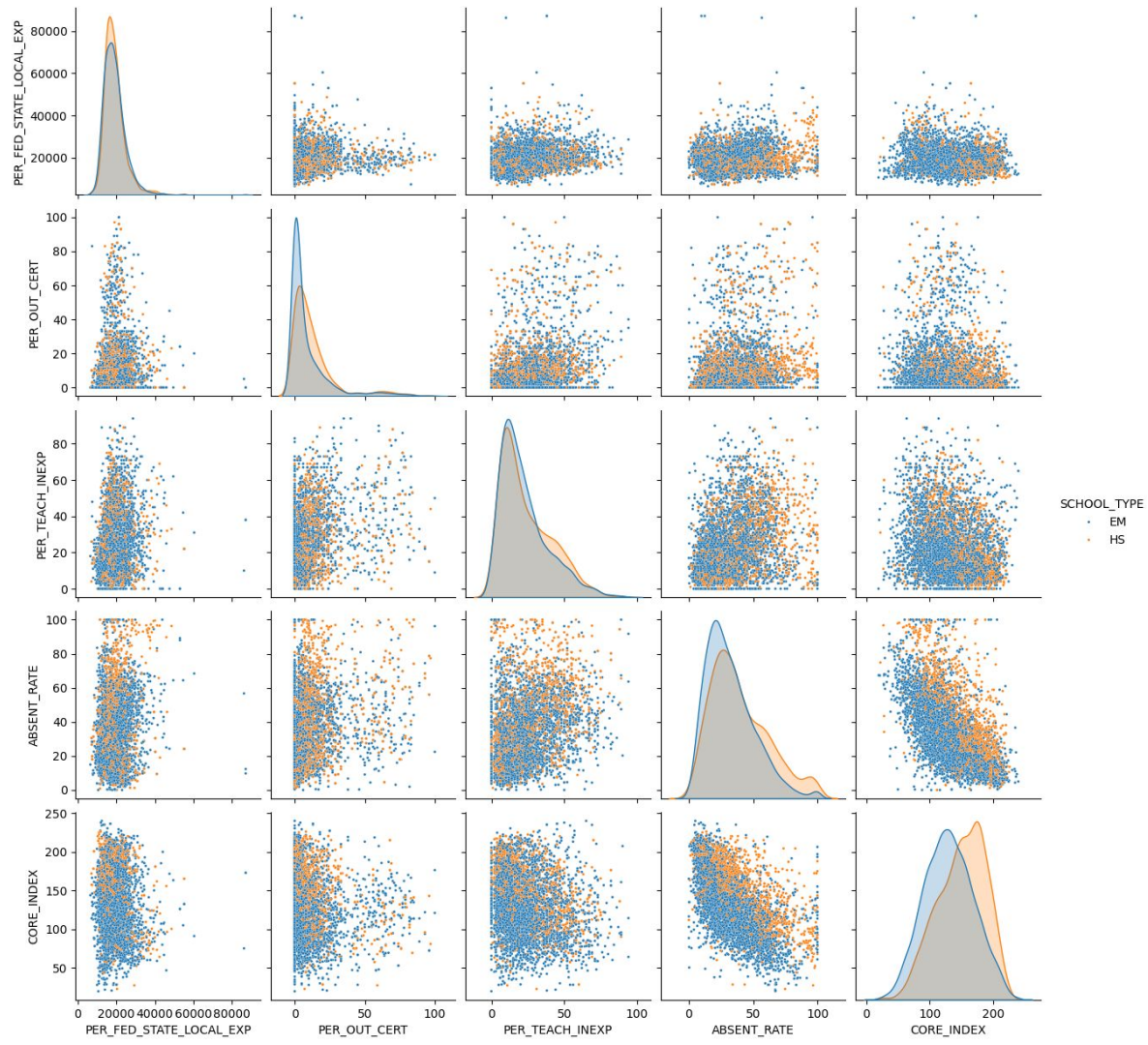
**Elementary:** A composite measure of elementary academic success at the school level, the *elementary core subject performance index* is determined from assessment scores in English language arts, mathematics, and science.

**High School:** The school-level *secondary core subject performance index* is determined from Regents exam scores in English language arts, mathematics, science, and social studies.

Composite scored Ranged from 0 - 250



# Exploratory Data Analysis



# Modeling the Data

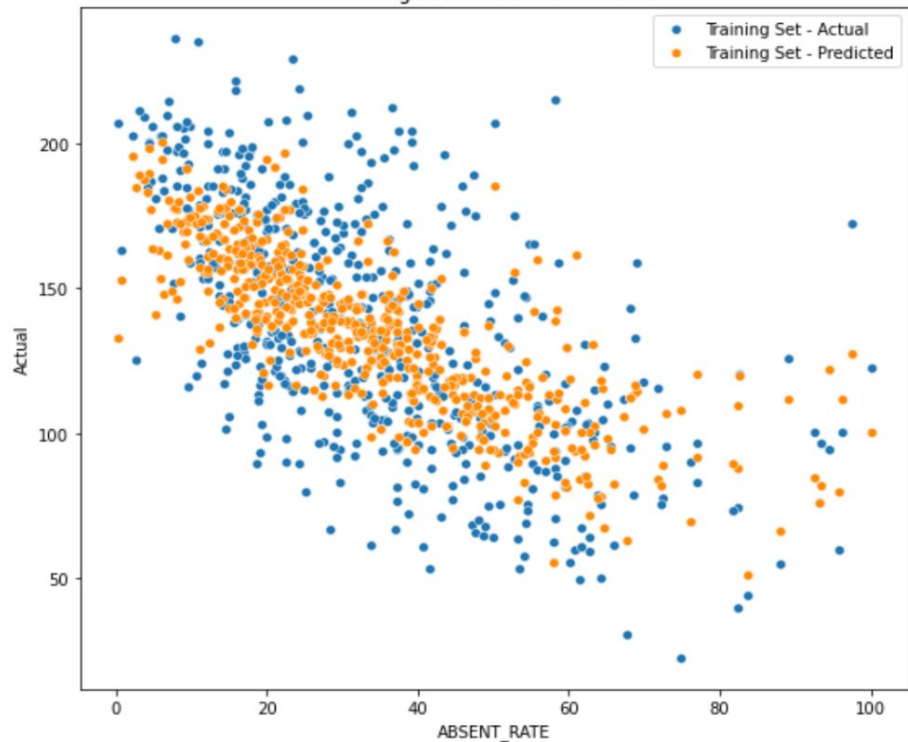
- Multivariate Linear Regression
- Polynomial Regression
- K-Nearest Neighbors
- Random Forest



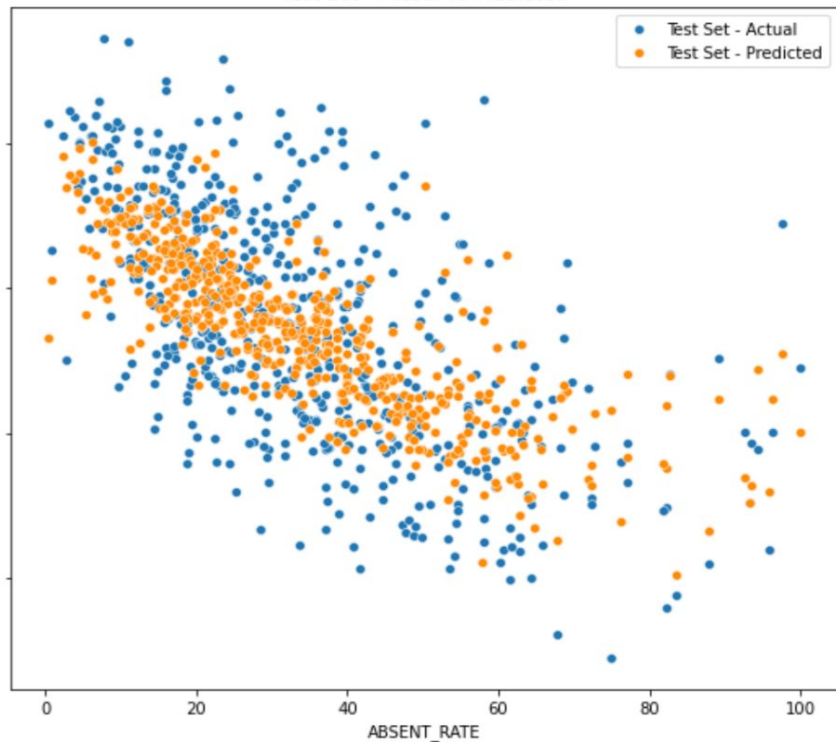
# Modeling the Data

## Absent Rate

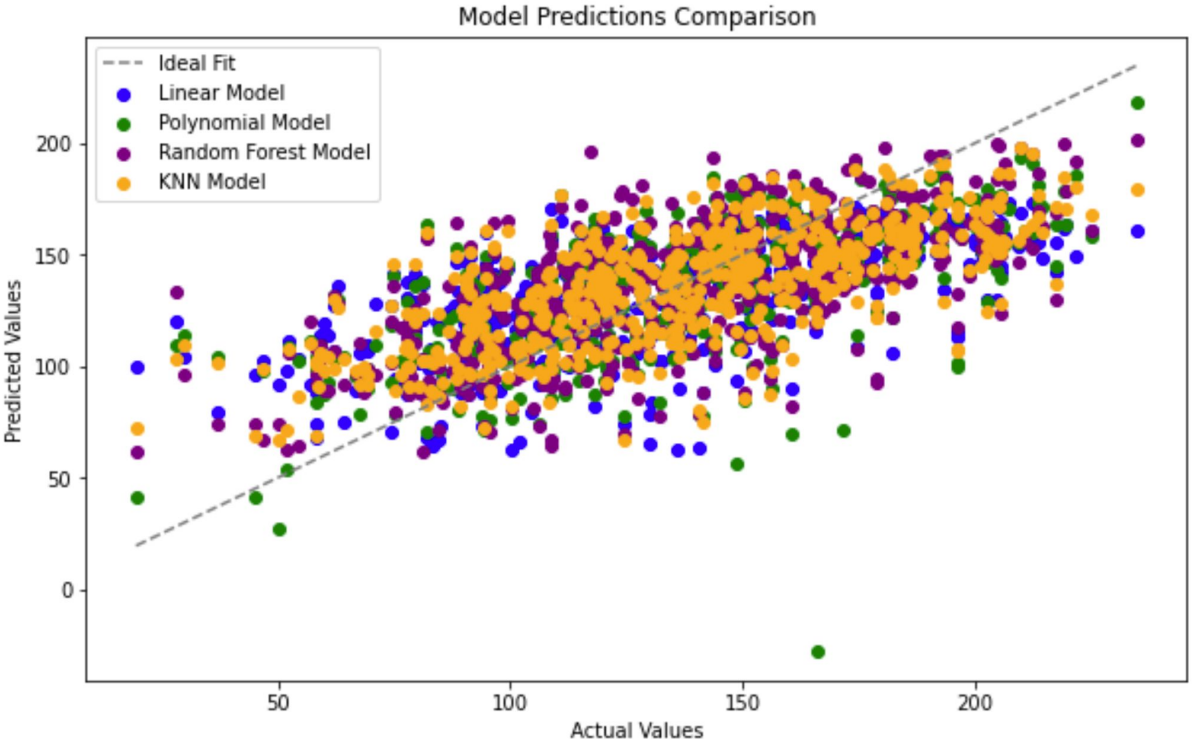
Training Set - Actual vs Predicted



Test Set - Actual vs Predicted

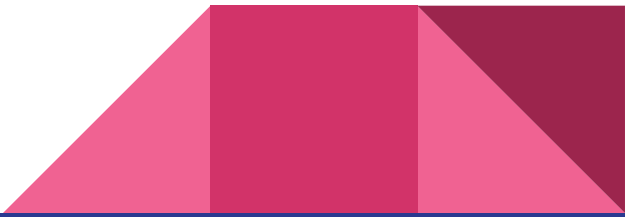


# Model Performance



R-squared Comparison	
Linear	0.382
Polynomial	0.409
Random Forest	0.416
KNN	0.465

R-squared



# Conclusions and Next Steps

Of the several features studied:

- The most impactful feature, based on the regression coefficients and on the correlation coefficients, was chronic absentee rate.
- Increased school funding, as measured by per-pupil expenditures, emerged as the least impactful feature in our study.

