An Inquiry Into Attendance Rates

By Kevin Macias-Matsuura

What and How?

Variable: **Student Attendance Rate**

- Variables in percentages (0-1)

Data: NYC School Quality Report (SQR)

All public middle/elementary (non-charter) schools in NYC

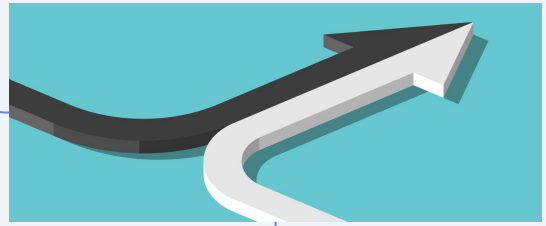
- Charters can choose their evaluation metrics (with approval)

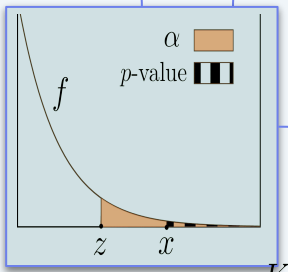
The Process

Model / Test

Gather / Clean

Feature Engineering

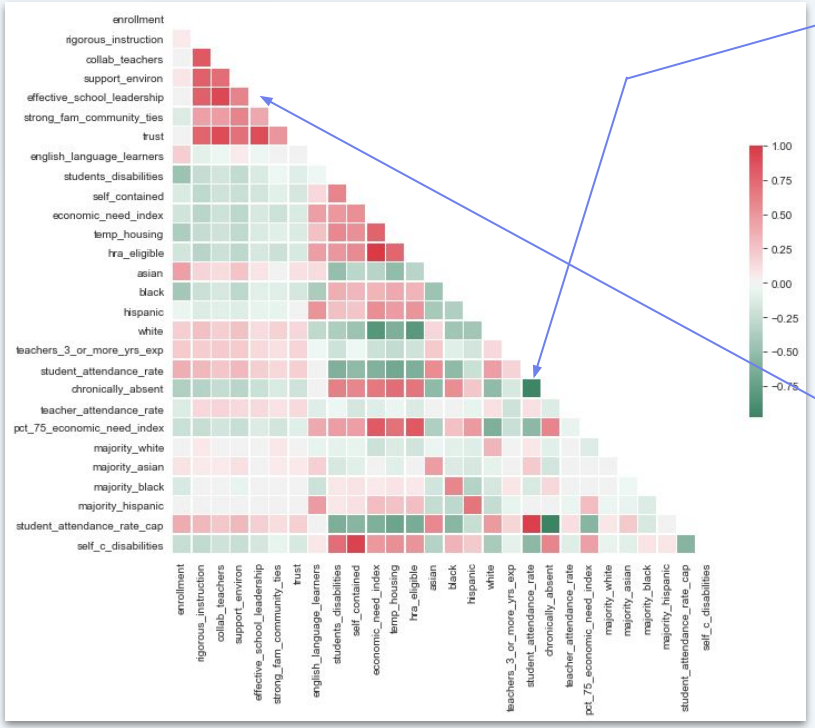




Stat - Testing

EDA

Statistical Testing



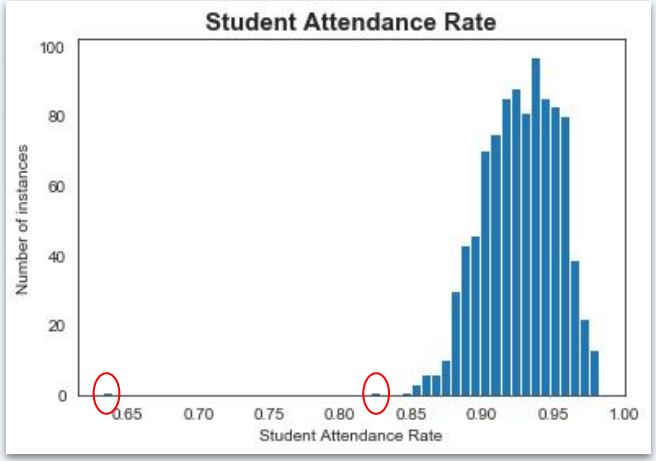
chronically_absent
student_attendance_rate

High negative correlation

High positive correlation

Came from same section of the questionnaire.

EDA

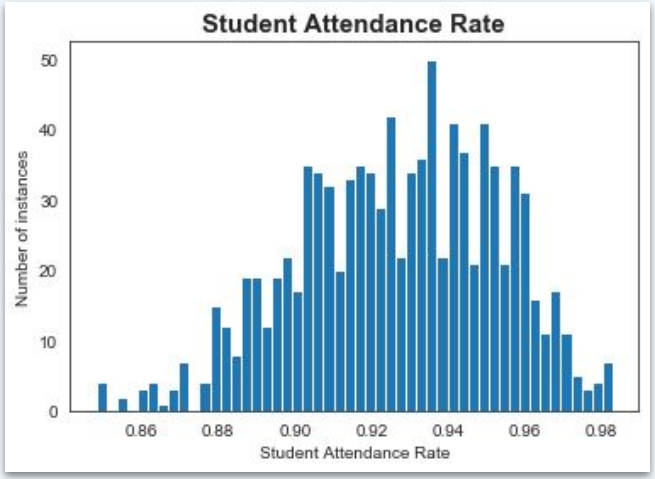


2 outliers

Negative skew

0 outliers

Approximately normal distribution



Models

Model #1

Coef	Result	Measurement	Result
Intercept	0.3621	Train RMSE	0.000908
economic_need_index_cat[T.3]	-0.0005	Test RMSE	0.015365
economic_need_index_cat[T.2]	-0.0004	Adj R^2	0.945
economic_need_index_cat[T.1]	-0.0024		
rigorous_instruction	0.0029		
collab_teachers	0.0155		
support_environ	-0.0029		
effective_school_leadership	-0.0192		
strong_fam_community_ties	0.0350		
trust	-0.0034		
english_language_learners	-0.0090		
students_disabilities	-0.0056		
self_contained	0.0221		
economic_need_index	0.0171		
temp_housing	0.0012		
hra_eligible	-0.0094		
asian	0.0220		
black	0.0212		
hispanic	0.0191		
white	0.0207		
teachers_3_or_more_yrs_exp	0.0023		
chronically_absent	-0.1024		
teacher_attendance_rate	-0.0092		
majority_white	-0.0020		
majority_asian	0.0005		
majority_black	-0.0013		
majority_hispanic	0.0001		
pct_75_economic_need_index	0.0031		

Model #2

Coef	Result	P-Value	Measurement	Result
Intercept	0.9545	0.000	Train RMSE	0.004675
collab_teachers	-0.0016	0.398	Test RMSE	0.005376
strong_fam_community_ties	0.0058	0.255	Adj R^2	0.966
english_language_learners	-0.0026	0.202		
students_disabilities	-0.0088	0.006		
self_contained	0.0073	0.092		
economic_need_index	0.0021	0.280		
asian	0.0248	0.000		
black	0.0119	0.076		
hispanic	0.0156	0.017		
white	0.0134	0.039		
chronically_absent	-0.1885	0.000		
majority_white	-0.0022	0.228		
majority_asian	0.0038	0.011		
majority_black	0.0016	0.040		

Model #3

Coef	Result	P-Value	Measurement	Result
Intercept	0.9557	0.000	Train RMSE	0.004673
strong_fam_community_ties	0.0038	0.406	Test RMSE	0.00539
english_language_learners	-0.0021	0.319	Adj R^2	0.966
students_disabilities	-0.0109	0.006		
self_contained	-0.0042	0.737		
economic_need_index	0.0023	0.379		
asian	0.0245	0.000		
black	0.0118	0.084		
hispanic	0.0155	0.021		
white	0.0130	0.047		
chronically_absent	-0.1879	0.000		
majority_white	-0.0021	0.253		
majority_asian	0.0037	0.018		
majority_black	0.0016	0.047		
self_c_disabilities	0.0392	0.319		
hra_temp	-0.0016	0.631		
asian_econ	-0.0002	0.971		

Final Model

Coef	Result	P-Value
Intercept	0.9545	0.000
strong_fam_community_ties	0.0058	0.255
english_language_learners	-0.0026	0.202
students_disabilities	-0.0088	0.006
self_contained	0.0073	0.092
economic_need_index	0.0021	0.280
asian	0.0248	0.000
black	0.0119	0.076
hispanic	0.0156	0.017
white	0.0134	0.039
chronically_absent	-0.1885	0.000
majority_white	-0.0022	0.228
majority_asian	0.0038	0.011
majority_black	0.0016	0.040

Measurement	Result
Train RMSE	0.004676
Test RMSE	0.005379
Adj R^2	0.966

High Fit / Low Interpretability

Top 3
chronically_absent
asian
hispanic

Takeaways

- **Chronic absences** are a strong factor in driving attendance down

- Factors one might assume **not always statistically impactful**

  Ex. temporary housing, teacher collaboration, etc

- Because money allocated by student attendance, recommend **possible implementation of additional questions** on the SQR

  Ex. How many times a month are you called by the school and/or teachers