CA Elementary Test Score Case Study

Rishab Mohandoss

Executive Summary

There is a belief that increasing California Elementary Schools staffing can have a positive effect on test scores. This report investigates the effect of various factors on standardized test scores across 420 elementary schools in California. The primary focus is to determine whether hiring more teachers, effectively reducing the Student-Teacher Ratio (STR), can significantly improve academic performance. We compared numerous metrics , influencing factors such as computer access, income disparities, and other test scores in our research. From this we fine tuned a range of variables that have the most pressing effect on test scores. We believe this analysis as well as deeper research into these variables can help effectively increase test scores in California elementary schools.

Objective

This analysis investigates the relationship between student-teacher ratios (STR) and standardized test scores across 419 California elementary school districts. The primary aim is to determine whether decreasing STR through increased staffing significantly improves academic performance. By identifying key factors that influence test outcomes, this study seeks to inform decisions on staffing and resource allocation in California schools.

Simple Linear Regression

In our simple Linear Regression we will be looking at test scores relative to the Student Teacher Ratio (STR). The STR measures the number of students in a classroom relative to the number of teachers. This will directly let us know how effective the proposed hypothesis of increasing the number of teachers would be on improving test scores. It is important to note that increasing the number of teachers would decrease the STR, as there are less students per teacher.

The regression line derived from the data set has an equation of:

Test Score = 698.10 - 2.24 × STR

The R Squared value from the regression analysis is:

R-squared = 0.04987

The regression equation, Test Score = 698.10 - 2.24 (STR), evaluates the relationship between student test scores and the STR. The intercept value of 698.10 represents the estimated test score when STR is zero, which, while not realistic, serves as a base for the regression line. The slope of 2.24 indicates that for every one unit increase in STR, test scores are expected to decrease by approximately 2.24 points. In other words, with one student increase per teacher on average, resulting in a 2.24 point decrease in test score. This negative relationship suggests that as classrooms become more crowded with more students per teacher, academic performance declines. Therefore, the equation supports the hypothesis that increasing the number of teachers, which would reduce the STR, is likely to improve student test scores.

An R-squared value of 0.04987 means that approximately 4.99 percent of the variation in test scores can be explained by changes in STR using this regression model. This is a relatively low R-squared, suggesting that while there may be a statistically significant relationship between STR and test scores, STR alone does not strongly predict test performance. This does not completely rule out a correlation between the factors, but suggests that other factors not included in the model likely play a larger role in determining test outcomes.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.223324
R Square	0.049873
Adjusted R Square	0.047595
Standard Error	18.53408
Observations	419

ANOVA

	df	SS	MS	F	Significance F
Regression	1	7519.094	7519.094	21.88887629	3.91039E-06
Residual	417	143244.6	343.5121		
Total	418	150763.6			

	Coefficients	andard Err	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	698.1029	9.455295	73.83196	1.4903E-241	679.5169295	716.6888912	679.5169295	716.6888912
17.88990888	-2.24152	0.479106	-4.67855	3.91039E-06	-3.183285631	-1.299759042	-3.183285631	-1.299759042

Table 1 - Linear regression between Test score and STR

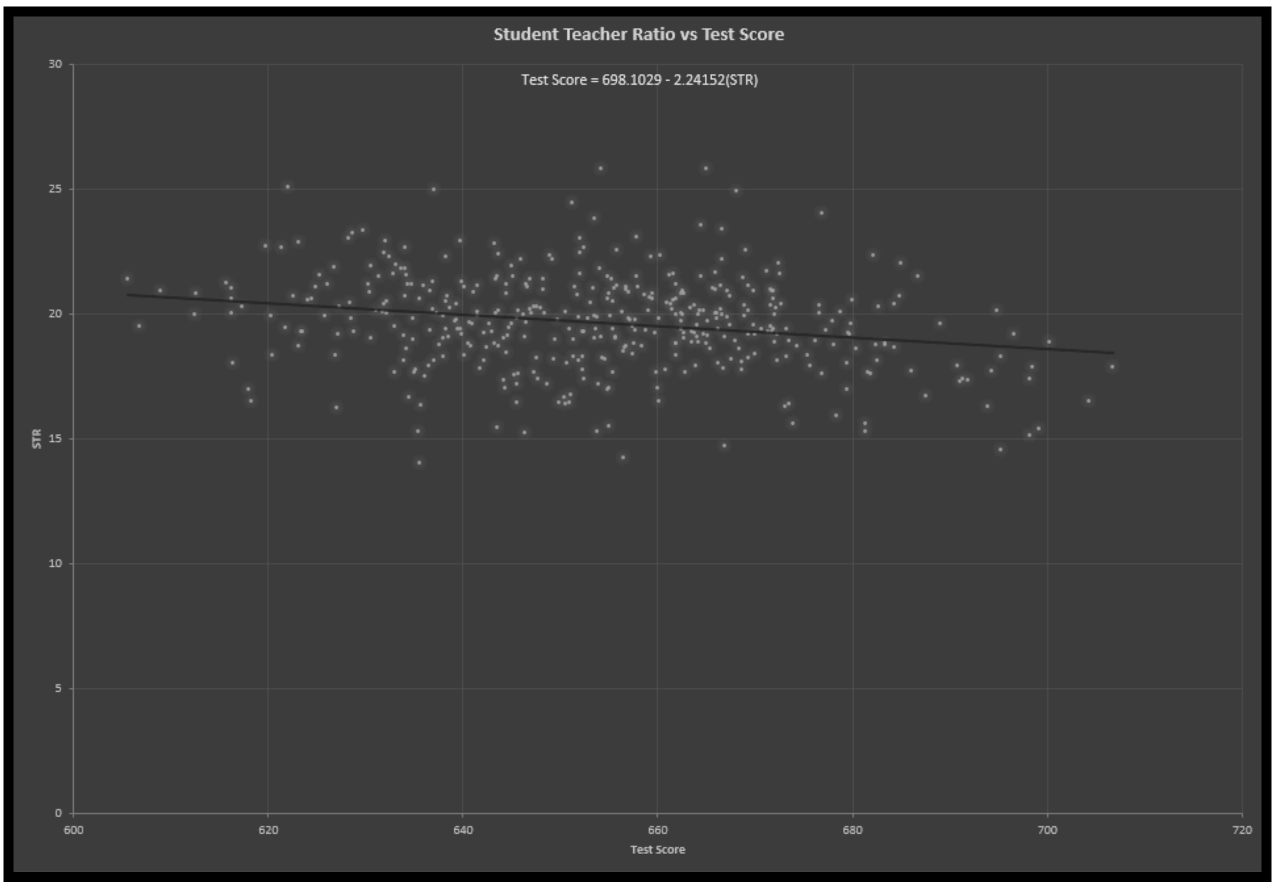

Fig 1 - Linear regression model between Test score and STR

Multi Linear Regression

In our multiple linear regression, we will be examining test scores relative to two variables: the Student Teacher Ratio (STR) and ELPCT, which represents the percent of English Learners at a school. ELPCT provides additional insight into how language barriers may affect student achievement. A higher percentage of English Learners could influence overall test performance. By including both STR and ELPCT in the model, we can better understand how class size and language proficiency together impact student outcomes for the Test Scores.

The regression line derived from the data set has an equation of:

Test Score = 689.455 - 1.0767 × STR - 0.66 × EL%

The R-squared value from the regression analysis is:

R-squared = 0.4253

The regression equation, Test Score = 689.455 - 1.0767 × STR - 0.66 × EL%, evaluates the relationship between student test scores and two key variables, the Student Teacher Ratio (STR) and the percentage of English Learner students (ELPCT). The intercept value of 689.455 represents the estimated test score when both STR and EL% are zero, which is not realistic but serves as a base value for the regression line. The slope of 1.0767 for STR indicates that for every one unit increase in STR, test scores are expected to decrease by approximately 1.08 points. Similarly, for every one percent increase in English Learner students, test scores are predicted to drop by about 0.66 points. These negative relationships suggest that both larger class sizes and higher proportions of English Learners can contribute to lower average test performance. The statistical significance of both variables (STR with $p < 0.01$ and EL% with $p < 0.05$) supports their strong predictive influence in this analysis.

An R-squared value of 0.4253 means that approximately 42.53 percent of the variation in test scores can be explained by changes in STR and ELPCT using this regression model. This is a moderate R-squared, indicating a meaningful relationship between the included variables and test performance. While the model does not account for all possible influences, it demonstrates that both class size and language proficiency when grouped together are important factors in understanding differences in student achievement.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.652153943
R Square	0.425304765
Adjusted R Squ	0.422541808
Standard Error	14.4317983
Observations	419

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	64120.5	32060.25	153.9309636	9.18375E-51
Residual	416	86643.15	208.2768		
Total	418	150763.6			

	Coefficients	*andard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	685.4549993	7.402354	92.5996	9.6577E-280	670.9043189	700.00568	670.9043189	700.0056798
17.88990888	-1.076728691	0.379694	-2.83578	0.00479452	-1.82308665	-0.33037073	-1.823086654	-0.330370727
0	-0.647484198	0.039277	-16.4852	2.38042E-47	-0.72468992	-0.57027848	-0.724689919	-0.570278477

Table 2 - Multi regression between Test score and STR and ELPCT

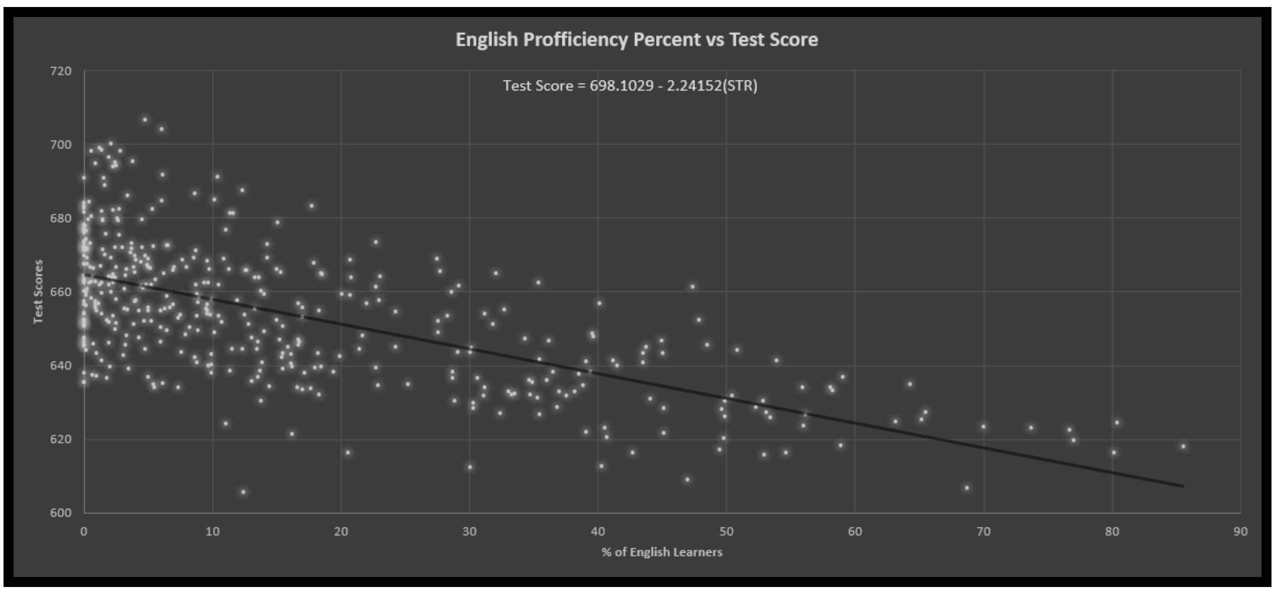

Fig 2 - Regression model between Test score and ELPCT

Multi Regression STR + ELPCT + MEALPCT

In our three variable multi-regression we examine the correlation between Test scores and Student teacher ratio (STR), Percent of English learners (ELPCT), and percent qualifying for reduced-price lunch (MEALPCT). MEALPCT, the percent qualifying for reduced-price lunch represents the socioeconomic background of the students, providing insight on how the student's economical background affects their ability to perform on tests. By including STD, and ELPCT with the addition of MEALPCT in the model, we can better understand how class size, language proficiency and also income barrier together impact student outcomes for the Test Scores.

The regression line derived from the data set has an equation of:
Test Score = 781.14 - 4.54 × STR - 0.025 × EL% - 0.40 × Lunch%

The R-squared value from the regression analysis is:
R-squared = 0.773

The regression equation, Test Score = 781.14 - 4.54 × STR - 0.025 × ELPCT - 0.40 × LunchPCT, evaluates the relationship between student test scores and three variables, the Student Teacher Ratio (STR), the percentage of English Learners (ELPCT), and the percentage of students receiving free lunch (MEALPCT). The intercept value of 781.14 represents the estimated test score when all three variables are zero, which serves as the base for the regression line. The slope of 4.54 for STR indicates that for every additional student per teacher, test scores are expected to decrease by 4.54 points. For every one percent increase in English Learners, test scores drop by 0.025 points, and for every one percent increase in students on free lunch programs, test scores decrease by 0.40 points. These results suggest that larger class sizes, higher proportions of English Learners, and greater economic disadvantage are all associated with lower academic performance.

An R-squared value of 0.773 means that approximately 77.3 percent of the variation in test scores can be explained by changes in STR, ELPCT, and MEALPCT using this model. This is a high R-squared, indicating a strong explanatory relationship between the predictors and test outcomes. All variables are statistically significant, showing high significance ($p < 0.01$) This asserts that class size, language proficiency, and socioeconomic status are all influential in shaping student academic achievement.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.879274475
R Square	0.773123602
Adjusted R Squai	0.771483532
Standard Error	9.078600984
Observations	419

ANOVA

	df	SS	MS	F	Significance F
Regression	3	116558.9	38852.98	471.39661	3.0371E-133
Residual	415	34204.71	82.421		
Total	418	150763.6			

	Coefficients	andard Erri	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	699.88344	4.691596	149.1781	0.00000000	690.6611839	709.1056934	690.6611839	709.1056934
17.88990888	-0.98888	0.238879	-4.13968	4.21407E-05	-1.458448349	-0.5193202	-1.458448349	-0.5193202
0	-0.12211	0.032316	-3.77855	0.000180833	-0.185630366	-0.05858385	-0.185630366	-0.05858385
2.040800095	-0.54585	0.021641	-25.2235	8.79027E-86	-0.588392817	-0.50331488	-0.588392817	-0.503314881

Table 3 - Multi regression between Test score and STR, ELPCT, and LunchPCT

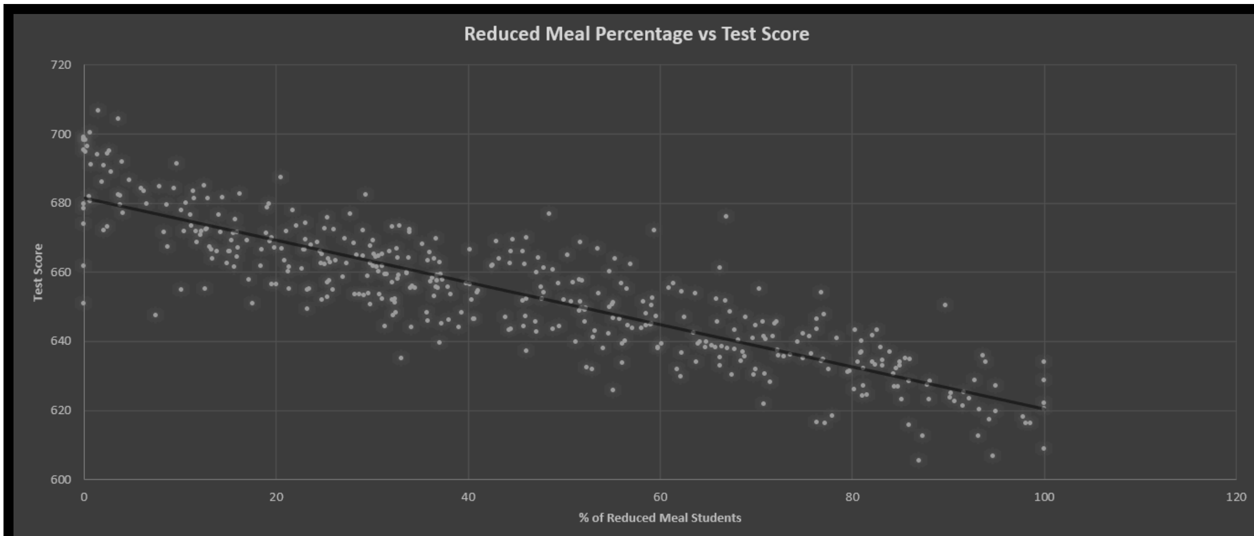

Fig 3 - Multi regression model between Test score and LunchPCT

Multi Regression STR + ELPCT + MEALPCT + MATHSCR

In our four-variable multi-regression, we look at the correlation between test scores and four key variables, Student Teacher Ratio (STR), Percent of English Learners (ELPCT), Percent qualifying for reduced-price lunch (MEALPCT), and average math score (MATHSCR). The inclusion of MATHSCR allows us to measure how prior student performance in math reflects on test scores. Math is often emphasized in better funded schools, contributing to overall test achievement. By combining these four variables, we gain a more complete understanding of how class size, language proficiency, socioeconomic status, and academic preparedness affect test scores.

The regression line derived from the data set has an equation of:

Test Score = 134.78 - 0.3378 × STR - 0.088 × ELPCT - 0.105 × MEALPCT + 0.8144 × MATHSCR

The R-squared value from the regression analysis is:

R-squared = 0.9764

The regression equation, Test Score = 134.78 - 0.3378 × STR - 0.088 × ELPCT - 0.105 × MEALPCT + 0.8144 × MATHSCR, evaluates the relationship between test scores and four predictors. The intercept value of 134.78 represents the base test score when all variables are zero. The slope of 0.3378 for STR means that for every additional student per teacher, test scores are expected to decrease by 0.34 points. For every one percent increase in English Learners, test scores decrease by approximately 0.088 points. For every one percent increase in students qualifying for free lunch, test scores fall by about 0.105 points. Most notably, for every one-point increase in the average math score, the overall test score is expected to increase by 0.8144 points. This highlights the strong impact of prior academic performance, particularly in math.

An R-squared value of 0.9764 indicates that approximately 97.64 percent of the variation in test scores can be explained by the model. This is a very high R-squared, showing the model fits the data extremely well. All variables in the model are statistically significant, with p-values well below 0.01. This confirms that student-teacher ratio, English language proficiency, socioeconomic status, and math achievement all play a substantial role in test scores. Schools with better funding, smaller class sizes, more English proficient students, and stronger math programs are shown to achieve significantly higher test scores.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.988153
R Square	0.976447
Adjusted I	0.976219
Standard I	2.928675
Observatic	419

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regressior	4	147212.7	36803.18	4290.846562	0.000
Residual	414	3550.935	8.577137		
Total	418	150763.6			

	Coefficients	*andard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	134.7758	9.57319	14.07846	4.65184E-37	115.9576458	153.593888	115.957646	153.593888
17.88991	-0.33776	0.077826	-4.33993	1.79346E-05	-0.4907437	-0.184776195	-0.4907437	-0.184776195
0	-0.08798	0.01044	-8.42676	5.93105E-16	-0.108501931	-0.067456174	-0.10850193	-0.067456174
2.0408	-0.10487	0.010156	-10.3252	2.15994E-22	-0.124830134	-0.084901594	-0.12483013	-0.084901594
690	0.81441	0.013623	59.78205	9.594E-206	0.787631577	0.841189296	0.78763158	0.841189296

Table 4 - Multi regression between Test score, with STR, ELPCT,LunchPCT and MATHSCR

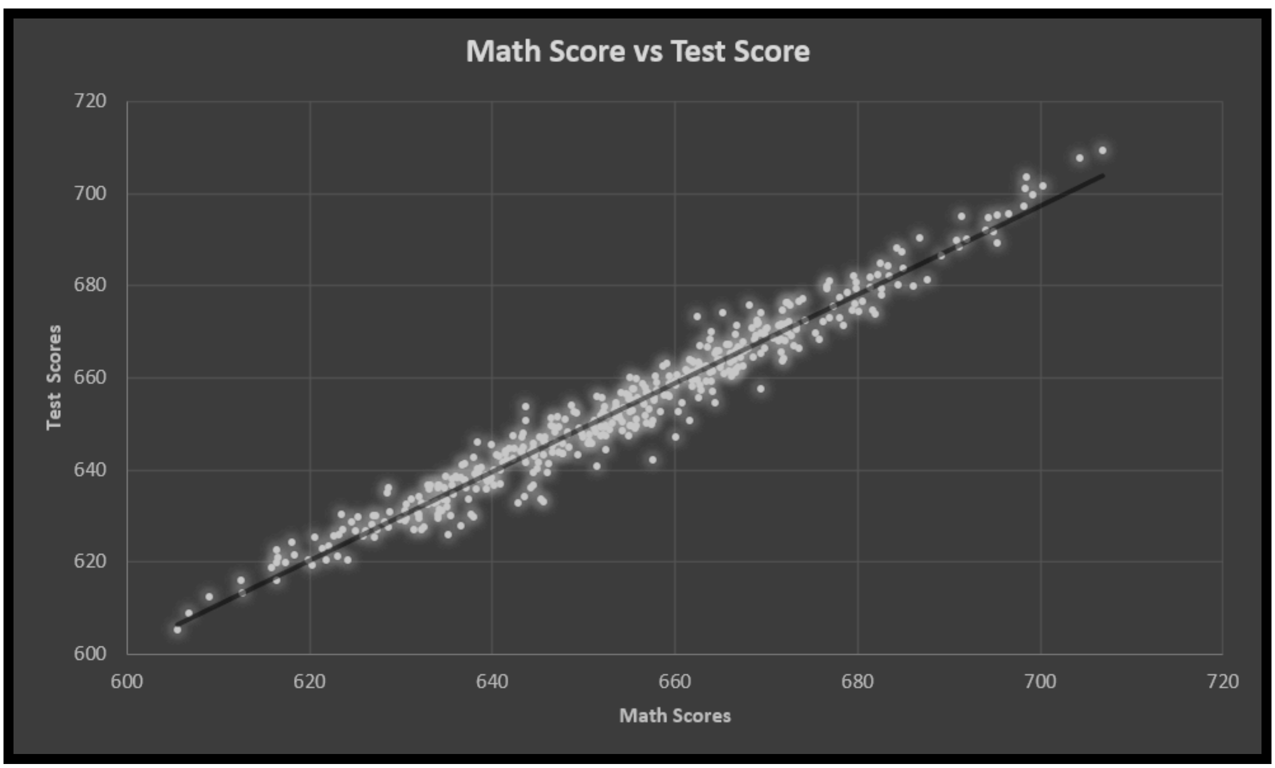

Fig 4 - Multi regression model between Test score and MATHSCR

Multicollinearity examination

While this model with 4 variables is a strong model. We do acknowledge some shortcoming between the correlations of the variables utilized, there is a strong correlation between Math score and Lunch reduced percent with a 0.82 correlation, and there is a moderate correlation between Math score and English proficiency with a 0.57 correlation. However, this shortcoming does not overshadow the fact that 97.64 percent of the variation in test scores can be explained by the model indicating that this model shows strong predictability and should be used despite its shortcoming.

Correlation between Math score and Student teacher ratio:	-0.195553424
Correlation between Math score and English proficency percent:	-0.568681849
Correlation between Math score and Lunch Percent:	-0.823014518

Conclusion

All three variables chosen are statistically significant predictors. Reduced Meal % emerges as the variable with the strongest correlation, showing the critical role of socioeconomic status in student performance in tests. However, both class size, and English proficiency are also variables with a strong correlation with a student’s performance on test scores. With student teacher being the still a strong variable, increasing student teacher ratio can significantly affect test score, with 4.54 score lost per one student teacher ratio increase, or there is on average one more student per teacher. Student teacher ratio is definitely one of the elements the school and the principal should invest money in to hire more teachers to reduce the student teacher ratio to improve test scores. Additionally, factors like math score which also play a pretty significant role in improving the student's overall test score should also be improved to give more incentive like math programs to improve on test score. However, other outside factors based on the student’s background like socioeconomic background and language (English) proficiency should also be noticed as a significant factor for future analysis. Based on this we recommend schools deviate funding to fix underlying issues that impact student outcomes. Prioritizing smaller class sizes by hiring more qualified teachers can lead to immediate academic gains. These improvements will lead to better understanding in foundational subjects like math. Simultaneously, targeting support for English learners' resources can bridge the language gap which has shown to negatively affect test performance. Furthermore, since socioeconomic disadvantage remains the strongest predictor of low achievement, schools should invest in additional services including meal programs, academic counseling, and community engagement initiatives. These strategies together can help create a more equitable learning environment, giving all students a fair chance to succeed regardless of background. All of these are simply predictions using regression analysis and we can not guarantee that they will lead to higher test scores, yet because of the data we are highly confident implementing these changes will better students test scores.