The Impact of Socioeconomic Factors on Grade 6 Reading Achievement in Ontario

Manal Saleh Al Adlouni

June 2025

Introduction

This report investigates how socioeconomic factors influence Grade 6 reading achievement across Ontario municipalities. Drawing on the **2023–2024 school performance dataset**, which provides the most recent and comprehensive measures of student achievement at the school level, alongside **2021 Canada Census data**, I built a reproducible analysis pipeline to explore the relationships between population density, low-income rates, parental education, and student performance. The goal is to provide actionable, data-driven recommendations to educators and policymakers aiming to improve educational equity and outcomes for all students in Ontario.

Summary of Key Findings (EDA & Hypothesis Testing)

Exploratory data Analysis (EDA) revealed that Grade 6 reading achievement scores are generally high across Ontario, with most municipalities clustered between **70%** and **95%**.

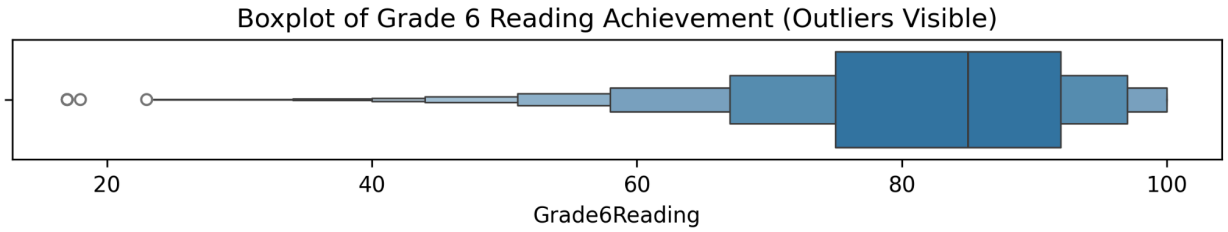


Figure 1. Distribution of Grade 6 Reading Achievement Across Ontario Municipalities

The boxenplot above displays the distribution of Grade 6 achievement rates by municipality. The central boxes represent the interquartile ranges, while the individual points on the left indicate municipalities with usually low achievement scores (outliers). The majority of municipalities cluster between **70%** and **95%**, with a median near **85-90%.** Outliers were flagged but retained for analysis, as they may reflect genuine local challenges.

However, a few municipalities stand out as outliers with significantly lower scores, as confirmed by both z-score analysis and boxenplots. These outliers may reflect unique local challenges or special circumstances and were retained for further analysis. To further explore potential predictors, I examined the relationship between population density and reading achievement. The scatterplot below shows no clear trend, a finding confirmed by a near-zero correlation coefficient.

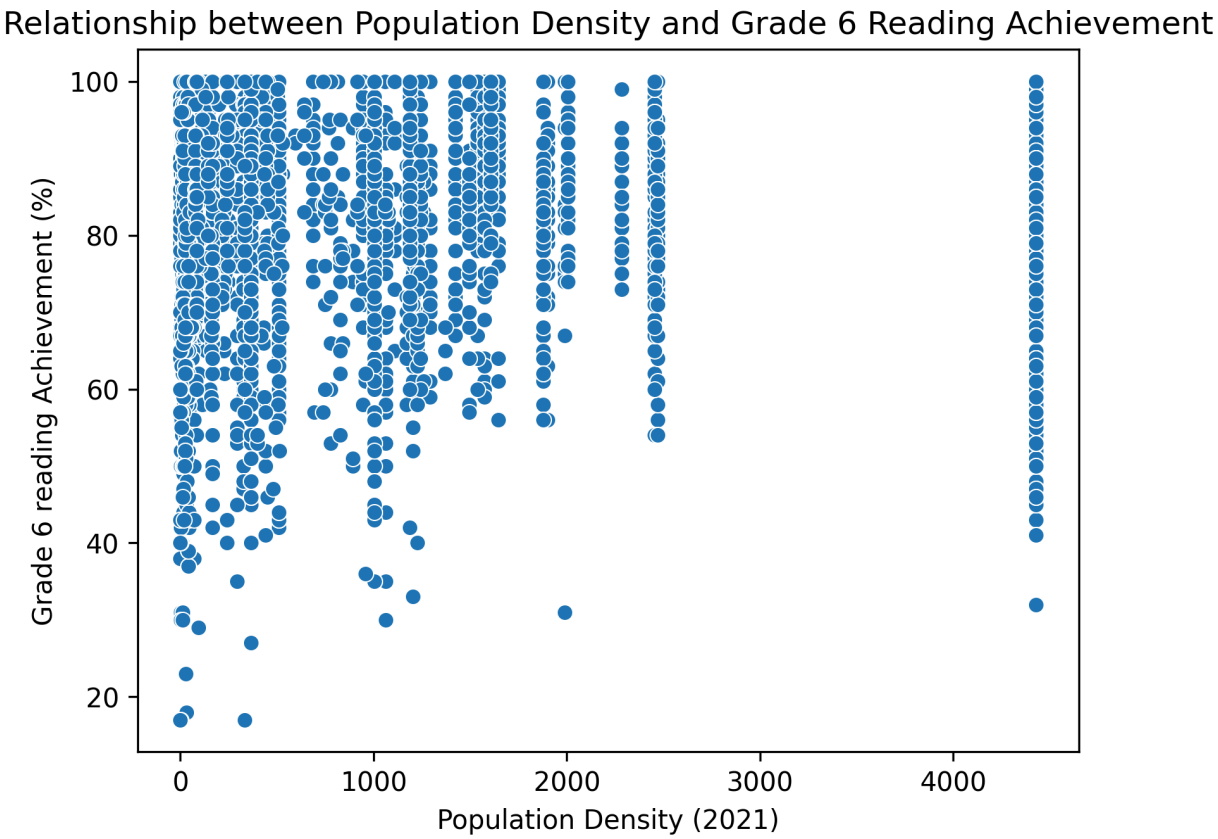


Figure 2. Relationship between Population Density and Grade 6 Reading Achievement

Each point represents a municipality. The scatterplot reveals no clear trend between population density and Grade 6 reading achievement, a finding confirmed by a near-zero correlation coefficient (r=0.049).

Hypothesis testing focused on the relationship between population density and reading achievement. The Pearson correlation coefficient was **0.049**, indicating almost no linear

relationship. A scatterplot confirmed this lack of association, showing a random cloud of points with no clear trend. Linear regression further supported this finding: population density explained only **0.2%** of the variance in reading achievement, and the effect size was negligible.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:          Grade6Reading   R-squared:                       0.002
Model:                            OLS   Adj. R-squared:                  0.002
Method:                 Least Squares   F-statistic:                     10.23
Date:                Sun, 22 Jun 2025   Prob (F-statistic):            0.00139
Time:                        00:22:06   Log-Likelihood:                -17004.
No. Observations:                4227   AIC:                         3.401e+04
Df Residuals:                    4225   BIC:                         3.402e+04
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          81.6187      0.284    287.626      0.000      81.062      82.175
Density_2021    0.0004      0.000      3.199      0.001       0.000       0.001
==============================================================================
Omnibus:                      631.314   Durbin-Watson:                   1.113
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1004.314
Skew:                          -1.026   Prob(JB):                    8.24e-219
Kurtosis:                       4.220   Cond. No.                     3.15e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 3.15e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Table 1. OLS Regression Results: Predicting Grade 6 Reading Achievement from Population Density

Population density is not a meaningful predictor of Grade 6 reading achievement. The model explains only **0.2%** of the variance in achievement, and the effect size is negligible: for every unit increase in density, reading achievement increases by just **0.0004** percentage points.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:          Grade6Reading   R-squared:                       0.157
Model:                            OLS   Adj. R-squared:                  0.156
Method:                 Least Squares   F-statistic:                     196.0
Date:                Sun, 22 Jun 2025   Prob (F-statistic):          2.26e-154
Time:                        00:22:32   Log-Likelihood:                -16644.
No. Observations:                4226   AIC:                         3.330e+04
Df Residuals:                    4221   BIC:                         3.333e+04
Df Model:                           4
Covariance Type:            nonrobust
==============================================================================
                         coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const                 86.8284      0.437    198.515      0.000      85.971      87.686
Density_2021           0.0019      0.000      8.076      0.000       0.001       0.002
LowIncome_imputed     -0.4243      0.043     -9.874      0.000      -0.509      -0.340
NoDegree              -0.4483      0.035    -12.679      0.000      -0.518      -0.379
Density_LowIncome -7.942e-05   1.91e-05     -4.163      0.000      -0.000    -4.2e-05
==============================================================================
Omnibus:                      429.074   Durbin-Watson:                   1.136
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              623.186
Skew:                          -0.779   Prob(JB):                    4.75e-136
Kurtosis:                       4.056   Cond. No.                     6.74e+04
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 6.74e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Table 2. Multiple Regression Results: Socioeconomic Predictors of Grade 6 Reading Achievement

Both the percentage of low-income households (**LowIncome_imputed**) and the percentage of parents without a degree (**NoDegree**) are strong, statistically significant predictors of lower Grade 6 reading achievement. Each 10 percentage point increase in these factors is associated with a 4.2-4.5 point decrease in achievement, holding other variables constant.

Population density (**Density_2021**) shows a small but statistically significant positive effect, though the practical impact is minimal. The interaction term (**Density_LowIncome**) is also significant and negative, suggesting that the negative effect of low income on achievement is slightly stronger in more densely populated areas.

The model explains about **16%** of the variation on grade 6 reading achievement.

Outlier Municipalities

While most municipalities in Ontario cluster within a relatively narrow range of Grade 6 reading achievement (typically 70-95%), a handful of municipalities stand out as outliers with notably lower of higher scores. For example, **rainy river** recorded a reading achievement rate of **41.5%**, which is more than three standard deviations below the provincial median. Other notable outliers include **magnetawan (42.0%), cobalt (47.0%),** and **duttondunwich (47.0%).** These outliers may reflect unique local challenges (such as recent demographic changes, funding issues, or community-specific factors) or , conversely, may highlight areas where further support is needed. Further qualitative research could help uncover the drivers behind these exceptional cases.

Model Diagnostics

To ensure the validity of the regression results, model diagnostics were performed. Residual plots were examined and indicated approximate normality, with some mild skewness. The Durbin-Watson statistic was **1.14**, suggesting some positive autocorrelation, though this is not unexpected in cross-sectional municipal data. The model's R-squared value of **0.157** indicates that about **16%** of the variance in reading achievement is explained by the included predictors . While this leaves substantial unexplained variance, the model's predictors are statistically robust.

Visual Summary Table

Statistic	Value
Median Reading Achievement	85 %
Mean Low Income (%)	9.7 %
Mean No Degree (%)	5.1 %
Number of Schools	6513
Number of Municipalities	321

Table 3. Visual summary table

Interpretation of Visual Summary Table:

The summary statistics provide important context for understanding the overall landscape of Grade 6 reading achievement and its potential drivers across Ontario municipalities:

- **Median reading achievement (85%):** The typical municipality has a Grade 6 reading achievement rate of 85%, indicating that most communities are performing well, with the majority of students meeting the provincial standard.
- **Mean low income (9.7%):** On average, about 1 in 10 school-aged children live in low-income households. This relatively low mean suggests that, while economic disadvantage is present, it is not the norm for most municipalities. However, even small increases in this percentage are associated with notable declines in achievement, as shown in our regression analysis.
- **Mean No Degree (5.1%):** The average municipality has just over 5% of parents without a degree, diploma, or certificate. This low figure suggests that most parents have some form of post-secondary education, which is generally associated with higher student achievement.

- **Number of Municipalities (321):** The analysis covers a large and diverse set of municipalities, lending robustness to the findings and ensuring that results are representive across the province.
- **Number of Schools (6513)**: The dataset includes 6513 schools, allowing for detailed analysis within and across municipalities.

These summary statistics highlight that while the overall achievement rate is high, there is meaningful variation in socioeconomic factors across municipalities. Even modest differences in low-income rates or parental education can have a significant impact on educational outcomes, reinforcing the importance of targeted, data-driven interventions.

	Grade6Reading	Density_2021	LowIncome_imputed	NoDegree
Grade6Reading	1.000000	0.049157	-0.326933	-0.310797
Density_2021	0.049157	1.000000	0.228898	0.177245
LowIncome_imputed	-0.326933	0.228898	1.000000	0.493462
NoDegree	-0.310797	0.177245	0.493462	1.000000

Table 4. Correlation matrix for key variables

Key Findings from the Correlation Matrix

- **Grade 6 reading Achievement** is negatively correlated with both the percentage of low-income households (**r= -0.33**) and the percentage of parents without a degree (**r= -0.31**). By that being said , Municipalities with higher rates of low-income families or lower parental education tend to have lower reading achievement among Grade 6 students.
- **Population Density** shows a very weak positive correlation with reading achievement (**r= 0.05**), indicating almost no linear relationship between how densely populated a municipality is and student achievement.

- **Low Income and No Degree** are moderately positively correlated (**r= 0.49**), meaning municipalities with more low-income families also tend to have more parents without a degree.
- **Density_2021** is weakly positively correlated with both LowIncome_imputed (**r=0.23**) and NoDegree (**r = 0.18)**, suggesting that denser municipalities may have slightly higher rates of socioeconomic disadvantage, but these relationships are not strong.

The most meaningful relationships are the negative associations between socioeconomic disadvantage (both low income and low parental education) and Grade 6 reading achievement. Population density, by contrast, shows almost no association with achievement, reinforcing the conclusion that targeted interventions should focus on socioeconomic factors rather than municipality size or density.

Key Insights on Socioeconomic Impact

1. **Low-Income Rates Are Strongly Linked to Achievement:**

   Multiple regression analysis showed that for every 10 percentage point increase in the proportion of low-income households, Grade 6 reading achievement drops by about **4.2** points. This underscores the powerful influence of economic disadvantage on educational outcomes.

2. **Parental Education Is a Significant Predictor:**

   Municipalities with higher percentages of parents without a degree also have lower student achievement, with a similar effect size to low-income rates. This suggests that both economic and educational capital in the home environment play critical roles in shaping student success.

3. **Population Density Has Minimal Impact:**

   While statistically significant due to the large sample size, the effect of population density on achievement is so small as to be practically irrelevant.

Limitations

This analysis is subject to several limitations. First, the data is cross-sectional, so while associations can be identified, causality cannot be established. Second, some missing values were imputed using the median, which, while common, may introduce bias or mask underlying patterns. Third, all data is aggregated at the municipal level, potentially obscuring important within-community variation. Finally, other potentially influential factors - such as school funding, teacher-student ratios, or access to extracurricular programs - were not included in this model due to data availability.

Actionable Recommendations

1. **Target Resources to High-Need Areas:**

   Direct additional funding, literacy programs, and tutoring support to municipalities with high proportions of low-income families and parents without degrees. This could include after-school reading clubs, summer literacy camps, or partnerships with local libraries.

2. **Enhance Parental Engagement in Low-Education Communities:**

   Launch outreach initiatives that equip parents with strategies to support learning at home, such as family literacy nights , workshops, or take-home reading materials. Tailoring these efforts to communities with lower parental education can build capacity and foster a culture of learning .

Stakeholder Impact

- **Students:** Will benefit from more equitable access to resources and support, potentially narrowing achievement gaps.
- **Schools:** May receive targeted funding and programming, but could also face increased scrutiny or accountability measures.
- **Policymakers:** Gain a data-driven framework for allocating resources where they are most needed.
- **Parents:** Are empowered to support their children, but care must be taken to avoid stigmatizing families or communities.

Ethical Considerations

All data used in this analysis is aggregated at the municipal level, minimizing privacy risks. However, it is important to avoid reinforcing stereotypes or stigmatizing disadvantaged communities.

Recommendations should be implemented with sensitivity, ensuring that additional support is framed as an opportunity rather than a deficit. Transparency about data limitations (such as imputation of missing values and the presence of outliers) is also essential for responsible interpretation.

Nuanced Recommendations & Trade-Offs

While socioeconomic factors are strong predictors of achievement , they are not the only influences. Interventions must be holistic, addressing not just economic and educational barriers but also factors like school climate , teacher quality, and community resources. Targeting resources may require difficult trade-offs, such as reallocating funds from higher-performing areas.

Outlier municipalities may need bespoke solutions rather than one-size-fits-all programs. All recommendations are grounded in the data, but should be regularly reviewed as new evidence emerges.

Data-Driven Framework for Decision-Making

By integrating census and school achievement data, this analysis offers a replicable framework for predicting which municipalities are at greatest risk for low achievement. Policymakers can use this approach to optimize the allocation of resources, track progress over time, and adapt interventions as new data becomes available. This continuous, evidence-based process supports more effective and equitable educational policy.

Conclusion

This analysis demonstrates that socioeconomic disadvantage and parental education are the strongest predictors of Grade 6 reading achievement in Ontario, while population density has little practical impact. Targeted, data-driven interventions - implemented with care and sensitivity - can help close achievement gaps and promote educational equity across the province.

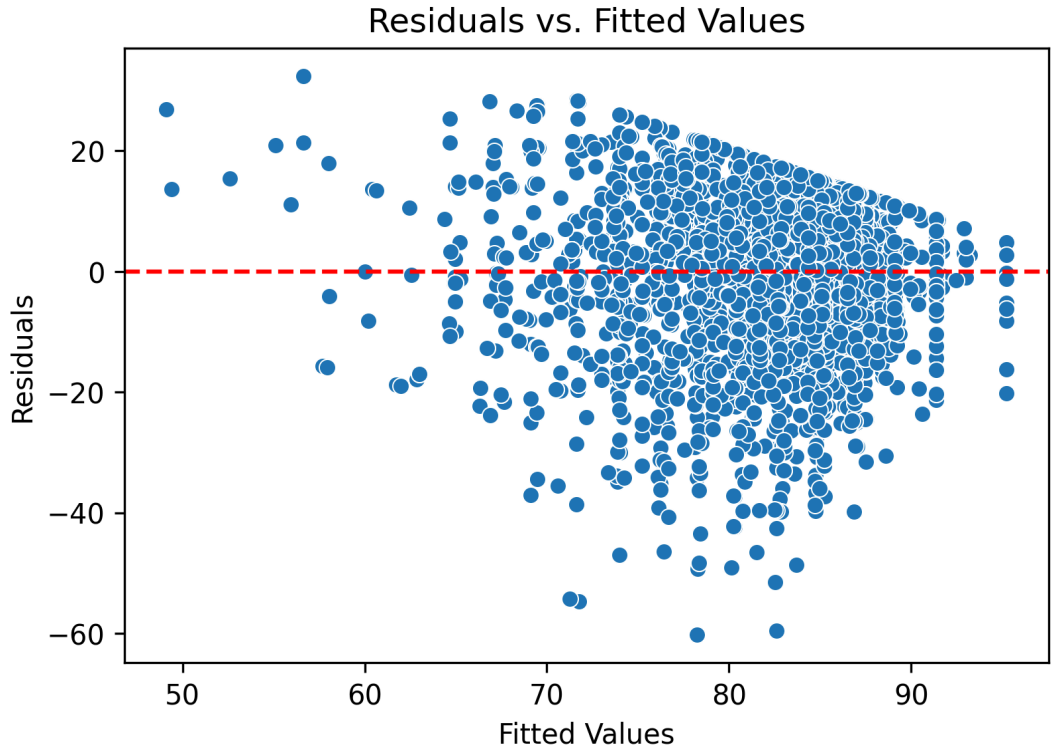
References

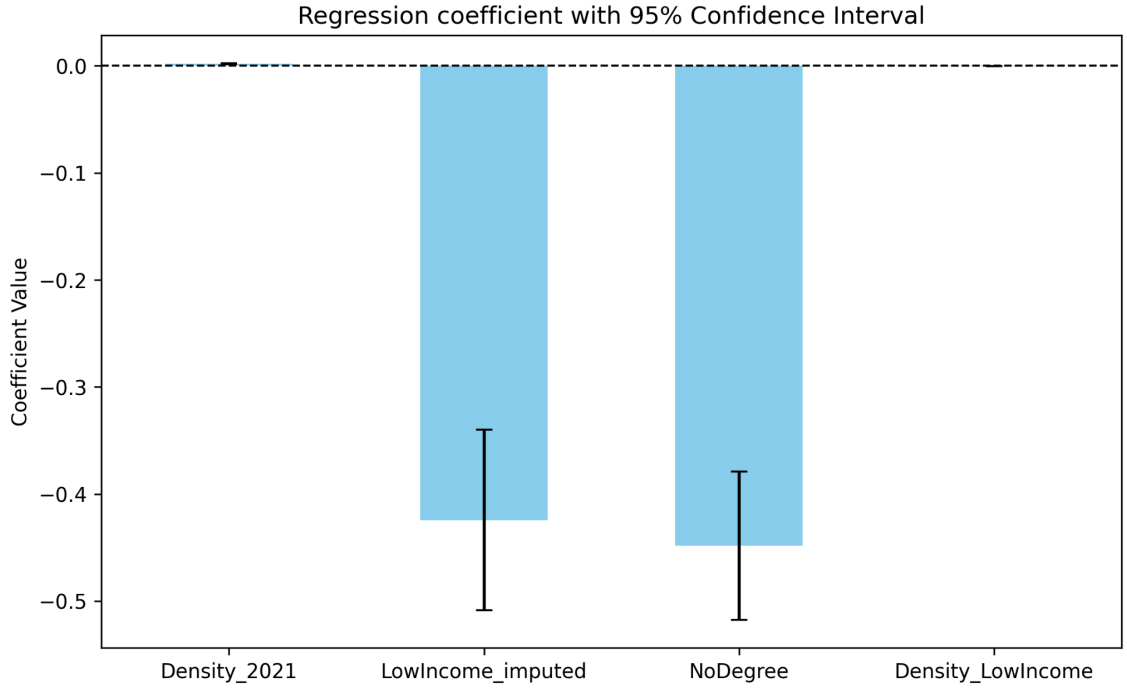
- Ontario Preliminary 2023-2024
- Population and dwelling counts
- EQAO
- Stack overflow
- Scikit Learn Documentation
- Stack Exchange
- Plotly
- OLS Regression
- Seaborn Documentation

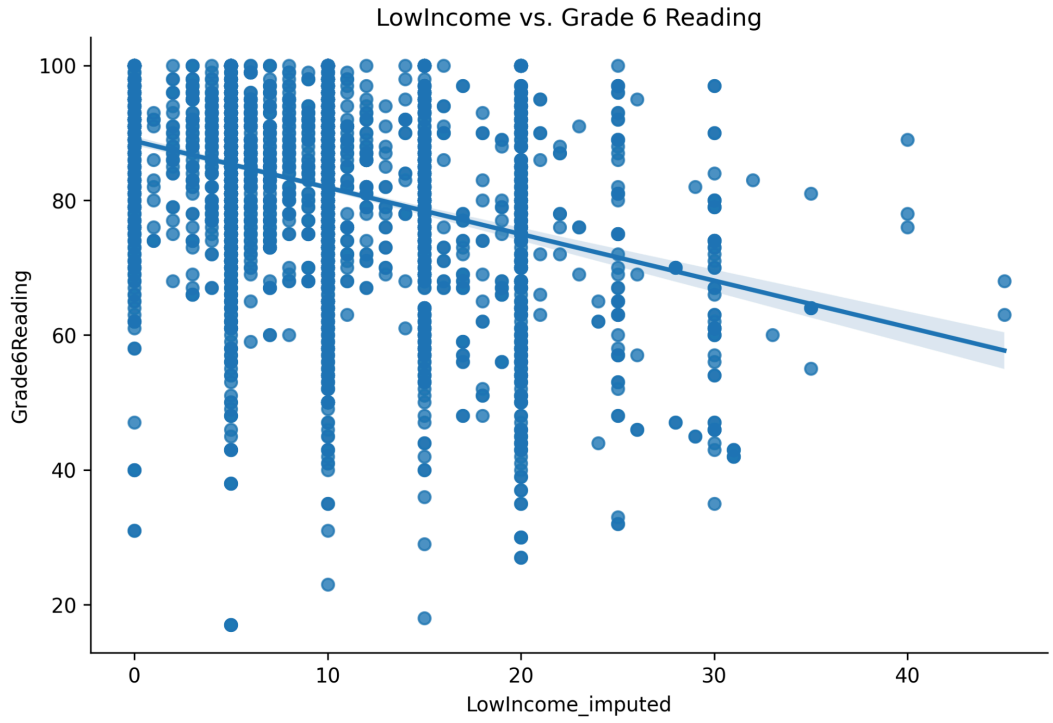
Appendix

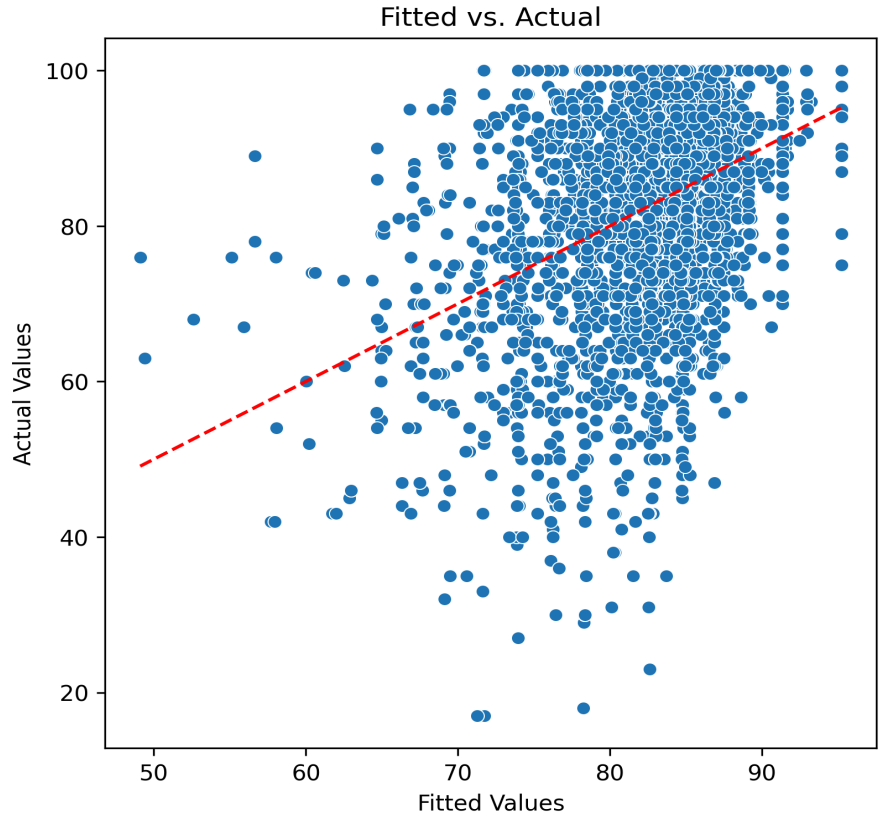
```
                            OLS Regression Results
==============================================================================
Dep. Variable:          Grade6Reading   R-squared:                       0.157
Model:                            OLS   Adj. R-squared:                  0.156
Method:                 Least Squares   F-statistic:                     196.0
Date:                Tue, 24 Jun 2025   Prob (F-statistic):          2.26e-154
Time:                        01:28:51   Log-Likelihood:                -16644.
No. Observations:                4226   AIC:                         3.330e+04
Df Residuals:                    4221   BIC:                         3.333e+04
Df Model:                           4
Covariance Type:            nonrobust
==============================================================================================
                                 coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------------
const                         86.8284      0.437    198.515      0.000      85.971      87.686
Density_2021                   0.0019      0.000      8.076      0.000       0.001       0.002
LowIncome_imputed             -0.4243      0.043     -9.874      0.000      -0.509      -0.340
NoDegree                      -0.4483      0.035    -12.679      0.000      -0.518      -0.379
Density_LowIncome -7.942e-05   1.91e-05     -4.163      0.000      -0.000    -4.2e-05
==============================================================================
Omnibus:                      429.074   Durbin-Watson:                   1.136
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              623.186
Skew:                          -0.779   Prob(JB):                    4.75e-136
Kurtosis:                       4.056   Cond. No.                     6.74e+04
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 6.74e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
```


Residuals vs. Fitted Values
Residuals
Fitted Values
20
0
−20
−40
−60
50
60
70
80
90


Regression coefficient with 95% Confidence Interval
Coefficient Value
0.0
−0.1
−0.2
−0.3
−0.4
−0.5
Density_2021
LowIncome_imputed
NoDegree
Density_LowIncome


LowIncome vs. Grade 6 Reading
Grade6Reading
LowIncome_imputed
100
80
60
40
20
0
10
20
30
40


Fitted vs. Actual
Actual Values
Fitted Values
100
80
60
40
20
50
60
70
80
90