Analysis Result Presentation

Education Inequality

Exploring what socioeconomic factors are responsible for this inequality

Introduction

- The issue of educational inequality in U.S. high schools is a pressing concern.

- To investigate this, I analyzed ACT/SAT scores as indicators of students' academic performance, examining how various socioeconomic factors influence these scores.

- The socioeconomic variables considered in this project include: unemployment rate, median household income, percent of children in married couple families, percent of adults with college degrees and percent of students at the school who received free or freduced price lunch.

About the Data

Made use of two datasets:

1. Edgap dataset

- This is the primary data set from Edgap.org
- From 2016
- Includes information about average ACT scores for schools and several socioeconomic characteristics of the school district.
- Socioeconomic factors such as household income, unemployment, adult educational attainment, and family structure

2. National Center for Education Statistics

- The secondary data set is basic information about each school from the National Center for Education Statistics.
- Consists of basic identifying information about schools and can be assumed to be of reasonably high quality.

Analysis Question: How are average school ACT scores influenced by socioeconomic factors?

This question could further be broken down into more questions:

1. How does Average ACT scores vary by state
2. How well can we predict students' ACT scores based on socioeconomic factors?
3. After controlling for mean income, do other socioeconomic factors still significantly affect ACT scores?
4. What factor contributes the most to ACT scores

1. How does Average ACT scores vary by state

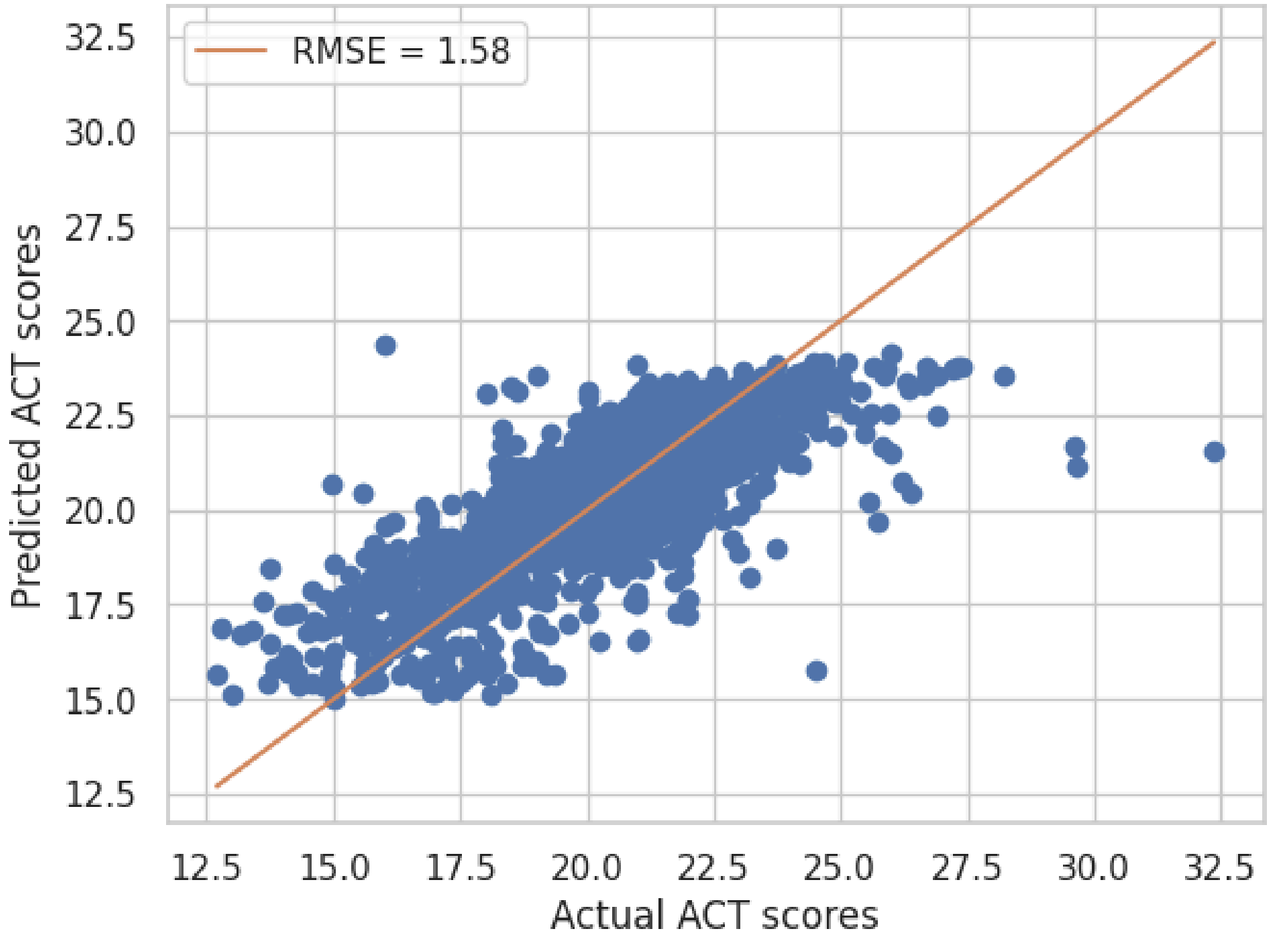
```
Top three states with the highest ACT scores and their median income:
       average_act  median_income
state
WI       22.009091      -0.126008
MA       21.746868       0.831553
WA       21.395333       0.002725

Bottom three states with the lowest ACT scores and their median income:
       average_act  median_income
state
LA       18.948428      -0.547890
NY       17.650441      -0.234813
DE       17.547823       0.184715
```

This indicates that average ACT scores are not solely influenced by median income, as might be expected. For example, Wisconsin, the state with the highest ACT scores, does not have the highest median income. Therefore, it is crucial to investigate what other factors might be contributing to these variations in academic performance.

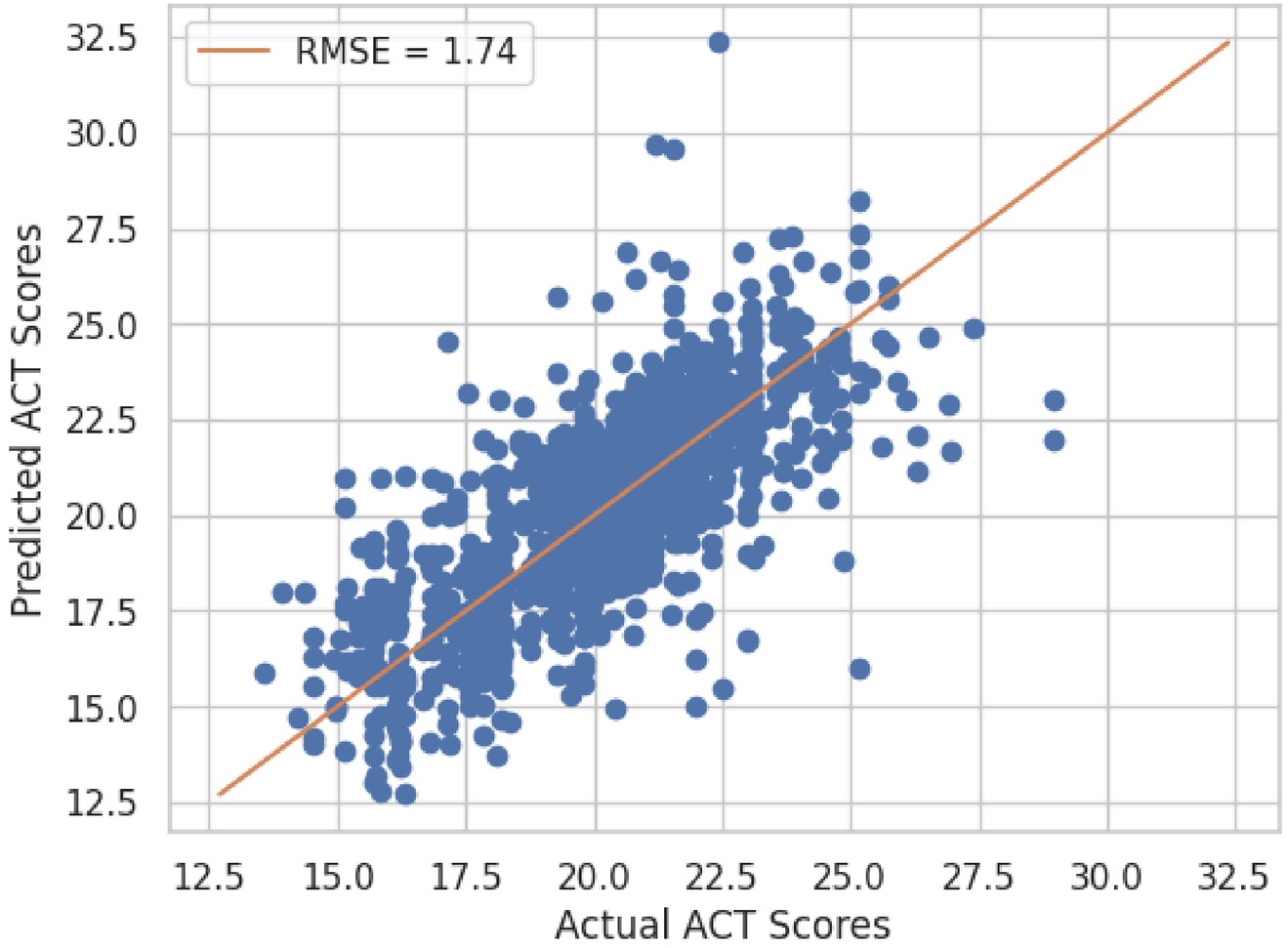
2. How well can we predict students' ACT scores based on socioeconomic factors?

BUILDING A MODEL USING LINEAR REGRESSION



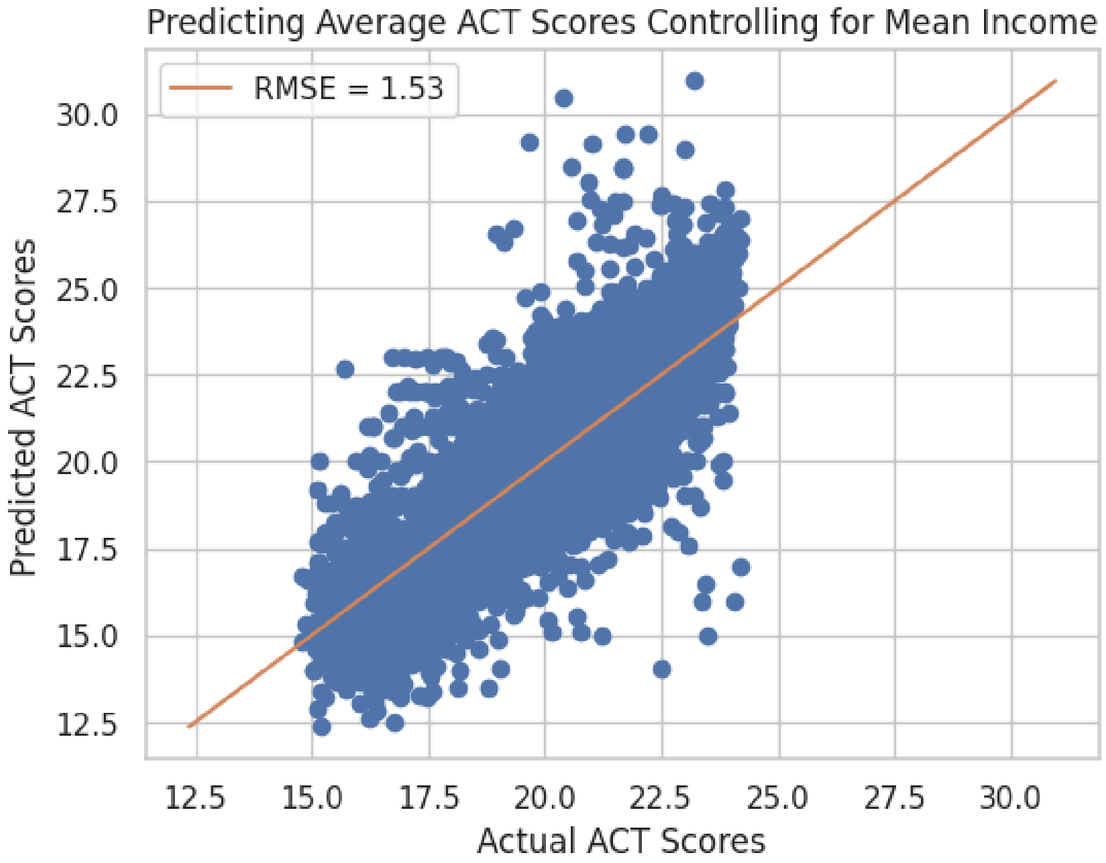
After training the model using the train data, the model performs well on the test data, with the RMSE difference being just 0.05. However, an RMSE of 1.58 is still somewhat high. Could a different type of model yield more accurate predictions?

BUILDING A MODEL USING STANDARD DECISION TREE



With and R squared score of 0.519, this model is able to predict about 51% of the ACT score's variability. This is a poorer prediction, therefore, the linear regression is a better model.

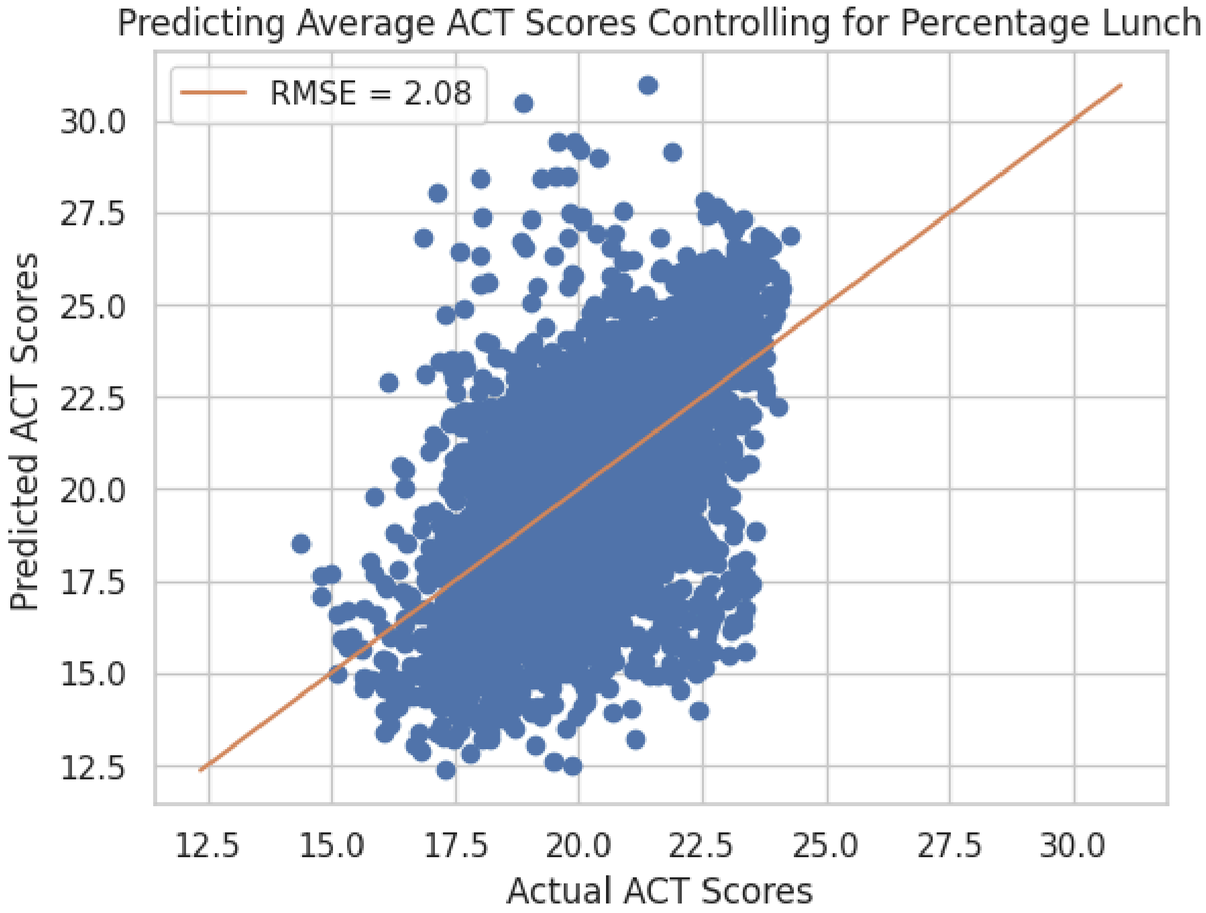
3. After controlling for mean income, do other socioeconomic factors still significantly affect ACT scores?



After controlling for mean income, the model remains similar to the original. This provides further evidence that another factor is more influential in predicting ACT scores.

4. What factor contributes the most to ACT scores

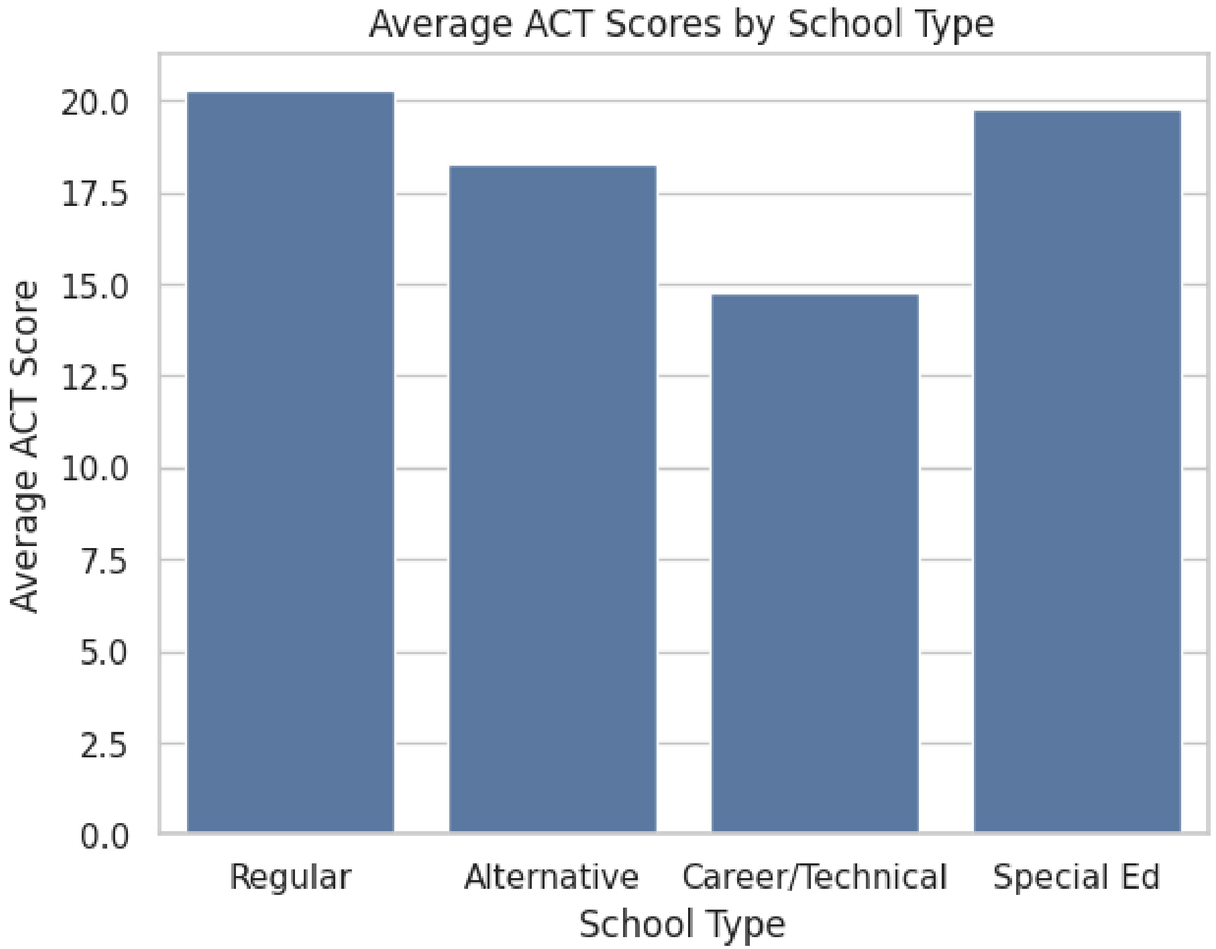
THIS IS A POORER MODEL WITH A HIGHER RMSE VALUE



This reveals that the socioeconomic factor with the most significant impact on predicting ACT scores is the percentage of students receiving free or reduced-price lunch. The R-squared value is 0.31, indicating that without this feature, only approximately 31% of the variability in ACT scores can be explained by this model.

Additional Analysis

WHAT INFORMATION CAN WE GATHER FROM THE CATEGORICAL DATA THAT WASN'T USED IN THE REGRESSIONS?



This reveals that Kids in special Ed did almost just as good as those in regular schools.

Conclusion

THERE ARE LOTS OF TAKEAWAYS FROM THIS ANALYSIS:

- THE AVERAGE ACT SCORE ISN'T NECCESSARILY INFLUENCED JUST BY MEDIAN INCOME AS WOULD BE EXPECTED.

- THE LINEAR REGRESSION MODEL WAS BETTER AT PREDICTING THE ACT SCORES COMPARED TO THE STANDARD DECISION TREE IN THIS CASE.

- THE MOST IMPORTANT PREDICTOR OF ACT SCORES IN THIS DATASET WAS THE PERCENT OF STUDENTS AT THE SCHOOL WHO RECEIVE FREE OR REDUCED PRICE LUNCH

- THE TYPE OF SCHOOL STUDENTS ATTENDED COULD BE AN INDICATOR OF ACADEMIC PERFORMANCE. THOSE WHO WENT TO REGULAR SCHOOLS PERFORMED THE BEST BUT THOSE THAT RECEIVED SPECIAL EDUCATION PERFORMED ALMOST AS GOOD. THOSE THAT ATTENDED CAREER OR TECHNICAL SCHOOLS PERFORMED THE POOREST.