Analysis of parental involvement and demographics in relation to academic progress

By Maclaryn Leonard and Satirtha Saha

Introduction of the Case Study

The purpose of this case study within our senior seminar classroom setting is to analyze and interpret the data set we were given in order to better understand how real life situations can be impacted by the data career field.

By seeing different scenarios we are able to acknowledge the ways in which data can be implemented in the world outside of a classroom setting and to add to our data portfolios to be used in our future endeavors following graduation.

Through the analysis of this case study we will be able to introduce potential policy implementations that could be introduced to the school district as well as the community surrounding it.

Motivation on the importance of the case study

By implementing systems to not only the target demographic but the entire school system there is a greater capability to strengthen the community along with increasing test scores.

By creating a situation in which all students can receive the best possible education not only will the school grow but also the community as a whole.

Ensuring that the school, families, and students are all involved in the education process there is more availability for growth both academically and within the surrounding community.

Introduction of the dataset

The dataset we were given for this case study was collected from the local Beloit School district from the third grade students testing and along with the data collected via a survey given to parents to fill out during parent teacher conferences to ensure the most potential involvement.

Due to the fact that they were only able to survey the parents who were involved in the parent teacher conferences and therefore the data has a bias. The biases in this data set in the sense that only the parents who came to the conferences along with the ones willing to participate

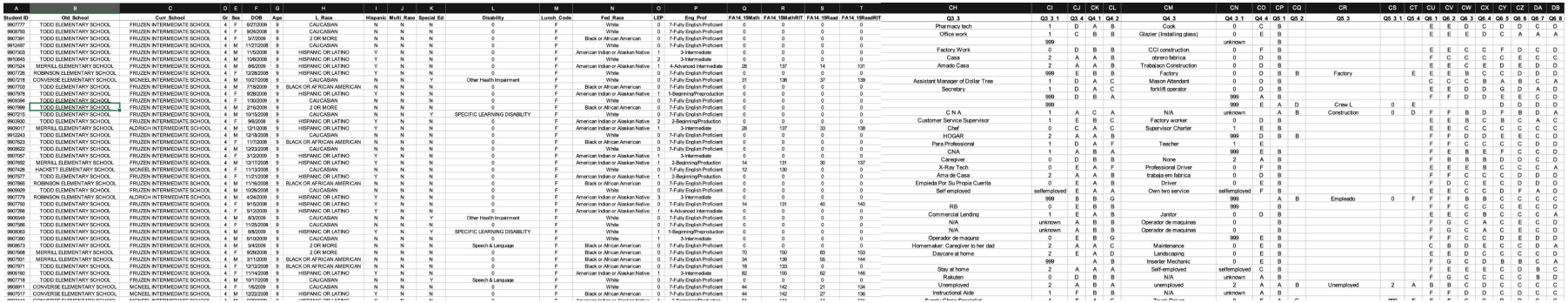
While discussing the provided data set we wanted to ensure that the data used in our regression model and visualizations were:

- Concise and Understandable.
- Necessary for the implementation of our policy.
- Able to be visualized in an effective manner.

Data Cleaning Process

For the data cleaning:

- We began by removing unnecessary variable columns that were not relevant to our hypothesis or model building.
- Next, we developed a script to summarize each column, which helped us understand the diversity within the dataset. Based on this, we removed values with very low frequencies, as they contributed little to the analysis.
- We also created a script to convert string values to numerical ones. In addition, we derived new variables such as **Per_Person_Budget**, calculated by dividing the household's yearly income by the total number of household members. Another new variable we created was the **Parental_Involvement_Score**, which is the sum of the numerical values assigned to answers from the parental involvement questions in the survey.
- Finally, we excluded students from the dataset if more than one parental involvement question was left unanswered by their parents.



Student ID	Sex	Race	Eng_Prof	Average MathRIT_18_19	Average ReadRIT_18_19	Total_Score_18_19	Per_Person_Budget	Parental_Involvement_Score
9907777	1	1	7	199.5	207	406.5	11229.5	12
9908793	1	1	7	205	207.5	412.5	7486.333333	10
9912497	2	1	7	204.5	194.5	399	6090.2	13
9907303	2	2	3	197.5	6	203.5	3743.166667	12
9910845	2	1	3	197	86	283	6090.2	15
9907726	1	1	7	215	108	323	18142.33333	6
9907703	2	3	7	191.5	187	378.5	15225.5	15
9907999	2	3	7	213.5	194	407.5	18877.4	7
9907215	2	1	7	185.5	81	266.5	17602.75	10
9903930	1	2	2	173.5	90	263.5	3743.166667	16
9909017	2	2	3	207	204.5	411.5	10403.16667	10
9912243	2	1	7	197	211	408	10150.33333	13
9907623	1	3	7	174.5	175	349.5	5614.75	9
9908622	2	1	7	191	180	371	12483.8	10
9907057	1	2	3	218	100	318	4491.8	11
9907692	2	2	2	198	192.5	390.5	14082.2	13
9907426	1	1	7	210.5	219	429.5	6401	14
9907577	1	2	2	186.5	0	186.5	7688.6	7
9907865	2	3	7	217.5	211.5	429	13067.16667	11
9907779	2	3	3	99.5	93	192.5	4491.8	9
9907793	1	2	7	193	202	395	4491.8	9

Data Exploration

By creating new variables as well as eliminating unneeded data for our study we are able to create applicable and appropriate visualizations.

Post Cleaning: Through the use of a code script we were able to not only eliminate unnecessary variables but to also consolidate information through the knowledge we gained about the

Key variables: The variables such as english proficiency, english and math test scores, total testing scores, and parental involvement and their impact upon the other key components of this study.

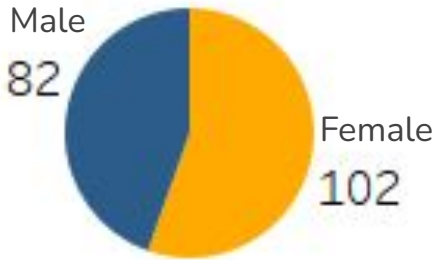
New Variables: By consolidating information we are able to have a broader understanding of these concerns within the school district.

Average MathRIT_18_19	Average ReadRIT_18_19	Total_Score_18_19	Per_Person_Budget	Parental_Involvement_Score

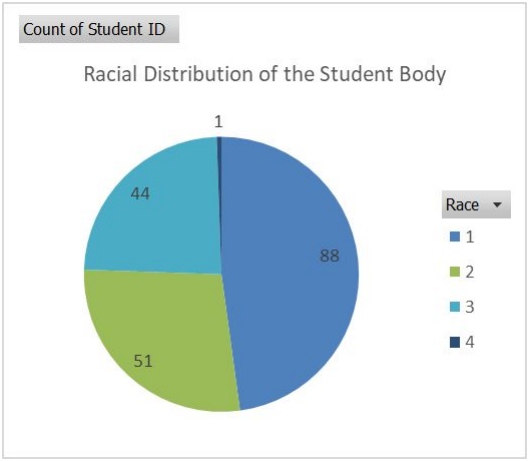
Descriptive Analysis/ Visuals

Visualizations:

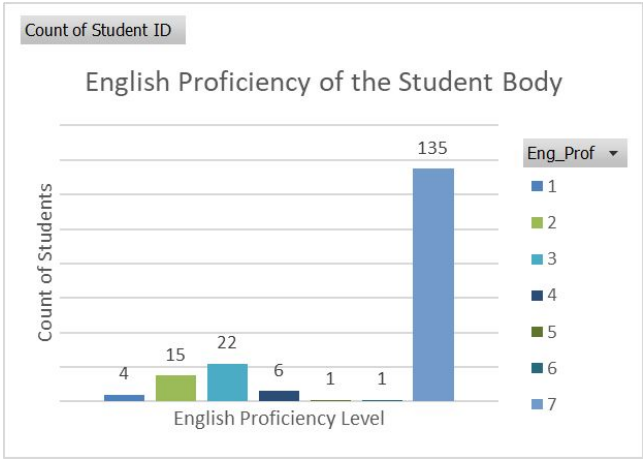
Male vs Female Student Population



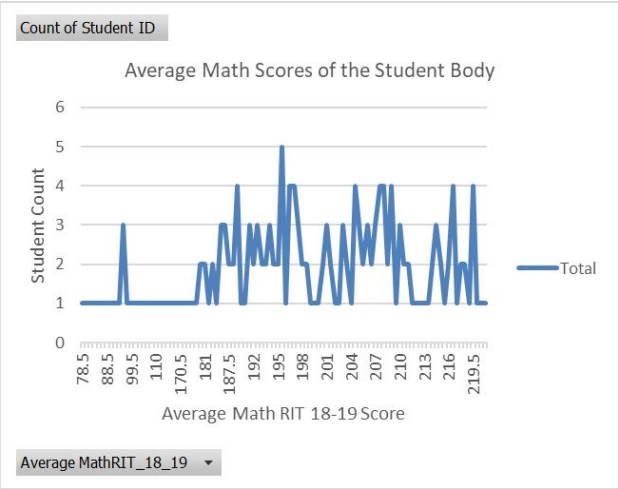
1 = White
2 = American Indian/ Alaskan Native
3 = Black or African American
4 = Asian



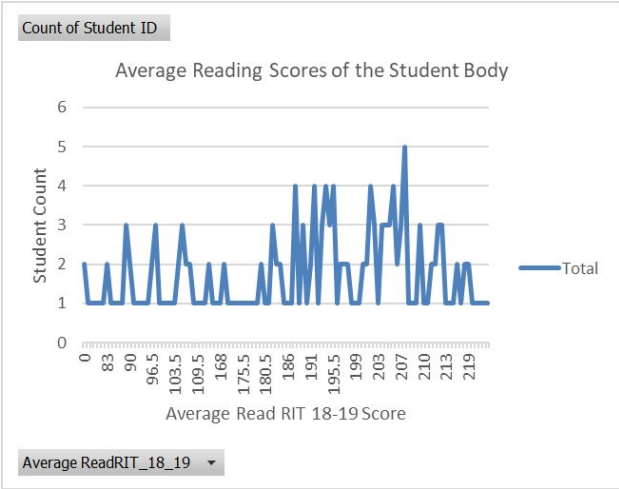
By visualizing the different demographics such as Sex, Race, and English Proficiency we were able to determine how the majority of the student body is represented but also allowing us to see areas in which students may be struggling in comparison to the majority of the total population. By identifying key demographics we are able to focus policies around struggling groups of individuals.



Math Test Scoring



Reading Test Scoring



Data Modeling

Basic Info (Applies to Both Models):

- Dataset: The dataset includes demographic variables, parental involvement scores, and household budget data to predict academic performance.
- Method: We used Ordinary Least Squares (OLS) Regression to model the relationship between independent variables (predictors) and dependent variables (academic performance measures).

Model 1: Effect of Demographics on Reading Test

- Independent Variables: Sex, Race, and English Proficiency. These factors were analyzed to see how demographic characteristics affect reading performance.
- Key Findings: The model explains the variance in reading scores (R-squared value) and shows which demographic factors have a statistically significant impact on the dependent variable.

Model 2: Effect of Parental Involvement and Economical Condition on overall Academic Performance

- Independent Variables: Parental Involvement Score, Per Person Budget, along with Sex and Race. This model assesses how parental engagement and economic conditions influence total academic performance.
- Key Findings: The model captures how household budget and parental involvement contribute to academic success and highlights which factors are most influential based on their coefficients and p-values.

Results

Model 1 Summary: Effect of demographics on Average ReadRIT_18_19

OLS Regression Results

Dep. Variable:	Average ReadRIT_18_19	R-squared (uncentered):	0.911
Model:	OLS	Adj. R-squared (uncentered):	0.909
Method:	Least Squares	F-statistic:	548.1
Date:	Thu, 07 Nov 2024	Prob (F-statistic):	3.07e-84
Time:	00:32:37	Log-Likelihood:	-884.24
No. Observations:	164	AIC:	1774.
Df Residuals:	161	BIC:	1784.
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Sex	14.1361	7.246	1.951	0.053	-0.174	28.446
Race	12.4991	4.541	2.753	0.007	3.532	21.466
Eng_Prof	20.5294	1.660	12.366	0.000	17.251	23.808

Omnibus:	5.818	Durbin-Watson:	2.128
Prob(Omnibus):	0.055	Jarque-Bera (JB):	5.885
Skew:	-0.321	Prob(JB):	0.0527
Kurtosis:	3.670	Cond. No.	12.1

Model 2 Summary: Effect of Parental Involvement Score and Per_Person_Budget on Total_Score_18_19

OLS Regression Results

Dep. Variable:	Total_Score_18_19	R-squared (uncentered):	0.918
Model:	OLS	Adj. R-squared (uncentered):	0.916
Method:	Least Squares	F-statistic:	449.8
Date:	Thu, 07 Nov 2024	Prob (F-statistic):	6.85e-86
Time:	00:32:37	Log-Likelihood:	-995.25
No. Observations:	164	AIC:	1998.
Df Residuals:	160	BIC:	2011.
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Parental_Involvement_Score	13.5480	1.735	7.809	0.000	10.122	16.974
Per_Person_Budget	0.0047	0.001	3.705	0.000	0.002	0.007
Sex	47.7303	15.120	3.157	0.002	17.869	77.591
Race	35.1673	8.930	3.938	0.000	17.532	52.803

Omnibus:	11.991	Durbin-Watson:	2.200
Prob(Omnibus):	0.002	Jarque-Bera (JB):	12.686
Skew:	-0.674	Prob(JB):	0.00176
Kurtosis:	3.198	Cond. No.	2.37e+04

Model 1: Effect of Demographics on Reading Test

- Key Predictors: English proficiency (coef = 20.53) has the strongest positive impact on reading scores, while race (coef = 12.50) and sex (coef = 14.14) also influence reading performance.
- Model Fit: The model explains 91.1% of the variance (R-squared = 0.911) in reading scores, showing that demographic factors are significant predictors of academic performance.
- Statistical Significance: English proficiency and race are both statistically significant predictors (p-values < 0.01), while the impact of sex is marginally significant (p = 0.053).

Model 2: Effect of Parental Involvement and Economical Condition on overall Academic Performance

- Key Predictors: Parental involvement score (coef = 13.55) and economic condition (coef = 0.0047) significantly impact total academic performance, with parental involvement having a strong positive effect.
- Model Fit: The model explains 91.8% of the variance (R-squared = 0.918) in total academic scores, indicating that family involvement and economic condition play important roles.
- Statistical Significance: All predictors, including sex and race, are statistically significant (p-values < 0.01), showing clear relationships between the independent variables and total academic scores.

Policy Focus Questions

Our questions surrounding key variables:

1) How does the amount of parental involvement have an impact on a students testing capabilities?

2) Does English proficiency have an impact on test scores, in particular reading?

Stakeholder Viewpoint:

Through the eyes of a stakeholder or even as a faculty member at the school they may view the policies as more work on their plate or potentially in a more harm than good mindset.

Due to the lack of guidance especially for families who may not be proficient in speaking, writing, or comprehending english they may not be as involved in the school day of their student due to this disconnect and they may need help adjusting.

By assisting not only the students in adjusting but also the parents these methods may be difficult for all parties involved but the effort is well worth the wait in the sense that the students will be able to learn and hopefully further their education while also allowing the school district as a whole to benefit from higher testing scores and potentially more volunteers in the form of parental involvement.

Policy Recommendations

Parental interaction policy:

The policy idea that we have hypothesized through the use of our cleaned and implemented data revolves around lack of parental involvement having a negative impact on academic achievement.

By visualizing data using different demographics standards we are able to see how different households involvement impact

Our policy recommendation is to create a system in which individuals of all demographics not only the ones whose students are struggling are mandated to

English Proficiency Improvement policy:

A secondary policy idea that could be used to improve the testing scores within this school district was to potentially implement an English proficiency course for not only students who are not as english proficient but all students to increase grammar and reading comprehension understanding at every level but also potentially for their parents in order to allow the parents to be included in their child's learning even if they aren't english proficient.

Ethical Concerns

Some ethical concerns we have addressed is the anonymity of students income status as well as the potential feelings of inadequacy or discrimination these parents and children may feel when given special treatment or mandated to attend events and such to expand their parental involvement.

In order to address these concerns we want to ensure that we implement individuals at every income level as well as in every demographic in order to not single anyone out or make anyone feel excluded.

Conclusion

In conclusion, after careful consideration we have identified that english proficiency as well as parental involvement are key variables in which there has been shown to be an effect on the School District of Beloit's test scores during this time.

Through our analysis we believe that these key factors could have a favorable impact on testing within this school district.

In order to assist in the increase of test scores within this student population we are proposing a parental involvement policy along with assistance to individuals who are struggling with english proficiency without the discrimination or targeting of a group of individuals.

Limitations of Analysis

While analyzing our data set we discovered that due to a lack of information we are unable to do certain policies that we had discussed such as missing information about the parents english ability as well as their own testing abilities, and any special education or learning disabilities they may have.

Without these variables we can’t bring about entirely accurate conclusions revolving around the students true background and any genetic influences.

Future Work

In the future, within the report we are going to create we are going to strive towards making a detailed and structured report that will be understandable from not only a data science and data analytics perspective but also leave room for understanding of the policies from the students, parents, and faculties of the school.

By ensuring that the visualizations are implemented well and provide the most accurate and concise data within the report can help with the spread of information to groups of all ages, education levels, and backgrounds.

Thank you very much for listening and to Professor Diep Phan for providing the Data Set!

Please feel free to ask any questions you may have. We appreciate your time.