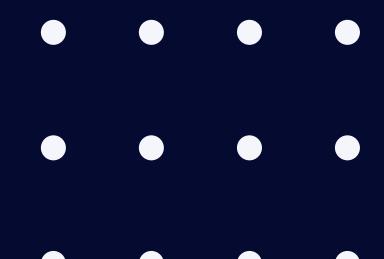
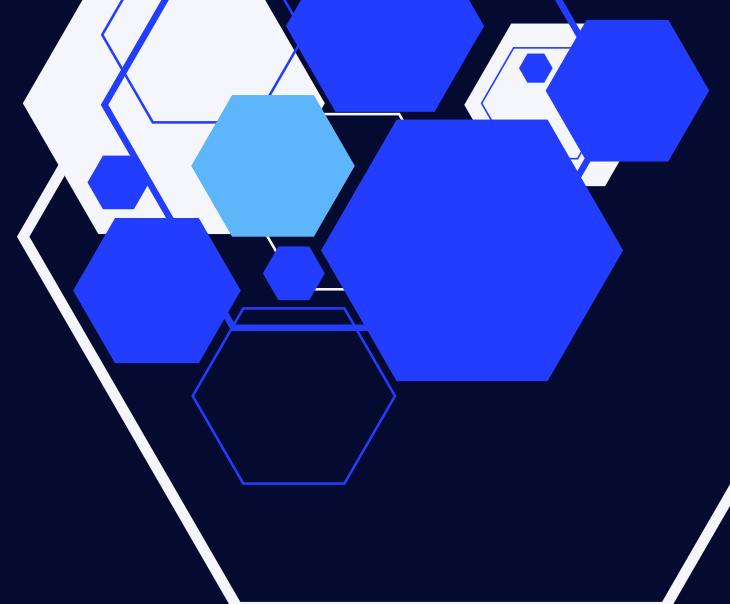


**INSIGHTS FROM EARLY CHILDHOOD EDUCATION DATA IN THE  
SCHOOL DISTRICT OF BELOIT**

# CASE STUDY 1

Presentation by: Aadi Joshi, RJ Pouzar, Prince Upadhyay



# Table of Contents

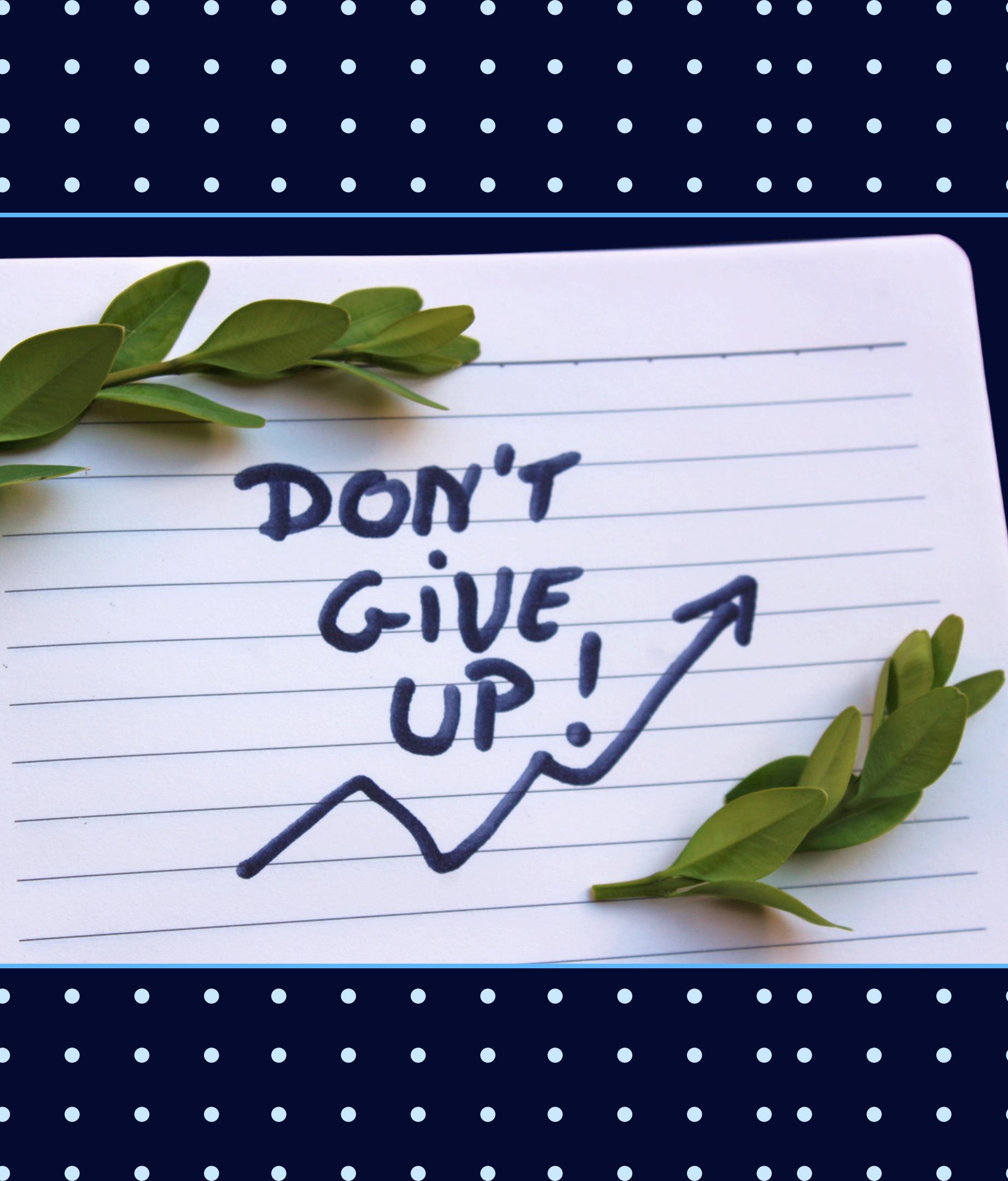
- 01 Introduction
- 02 Motivation
- 03 All About our Data
- 04 Descriptive Analysis & Visuals
- 05 All About our Research
- 06 Data Modeling
- 07 Outcome & Recommendations
- 08 Limitations and Future Work

# INTRODUCTION



# Why this case study is important?

- Early Intervention
- Data-Driven Policy Making
- Community Impact
- Informed Use of Resources

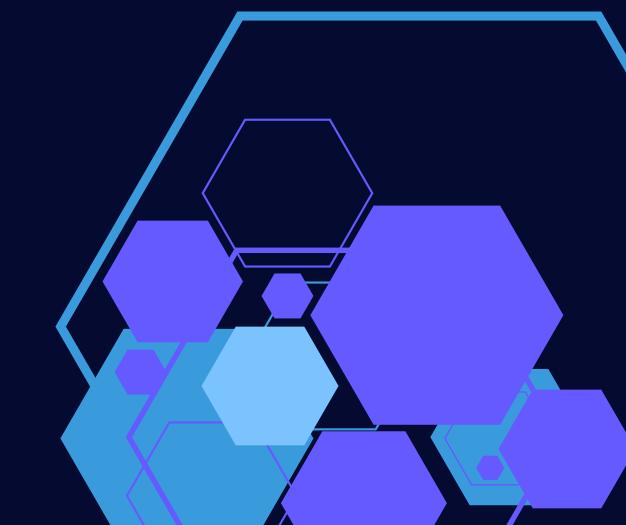
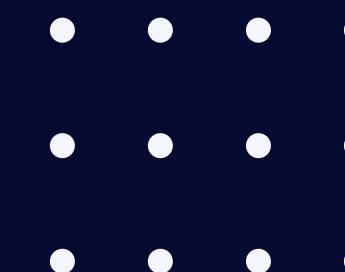


# ALL ABOUT OUR DATA



# Data Cleaning

- Variable Isolation
  - We began by isolating each variable independently, allowing for targeted and precise cleaning measures tailored to the unique characteristics of each variable.
- Visualization and Feature Interaction
  - To identify trends and outliers, we visualized each variable against performance scores using both Tableau and Python.
- Sorting Variables
  - Based on the insights gained from our visualizations, we sorted the variables into predictors and control variables.
- Ambiguity Resolution
  - We thoroughly cleaned data that was ambiguous and contributing to noise in the dataset.
- Handling Missing Values: Missing values were addressed systematically. Each field was considered individually:
  - Manual Correction
  - Median Imputation



# Tools and Purpose

- **Google Sheets**

- We used this application to manually clean the data.
- Filtering and sorting of data.

- **Python**

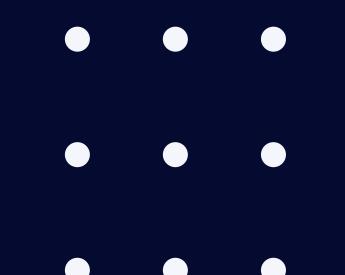
- Exploratory Data Analysis (EDA), handling missing values, converting categorical to continuous labels, & regression models training.

- **Tableau**

- We used this feature for visualizing the data in order to provide us more information on our target, predictors, & control variables.

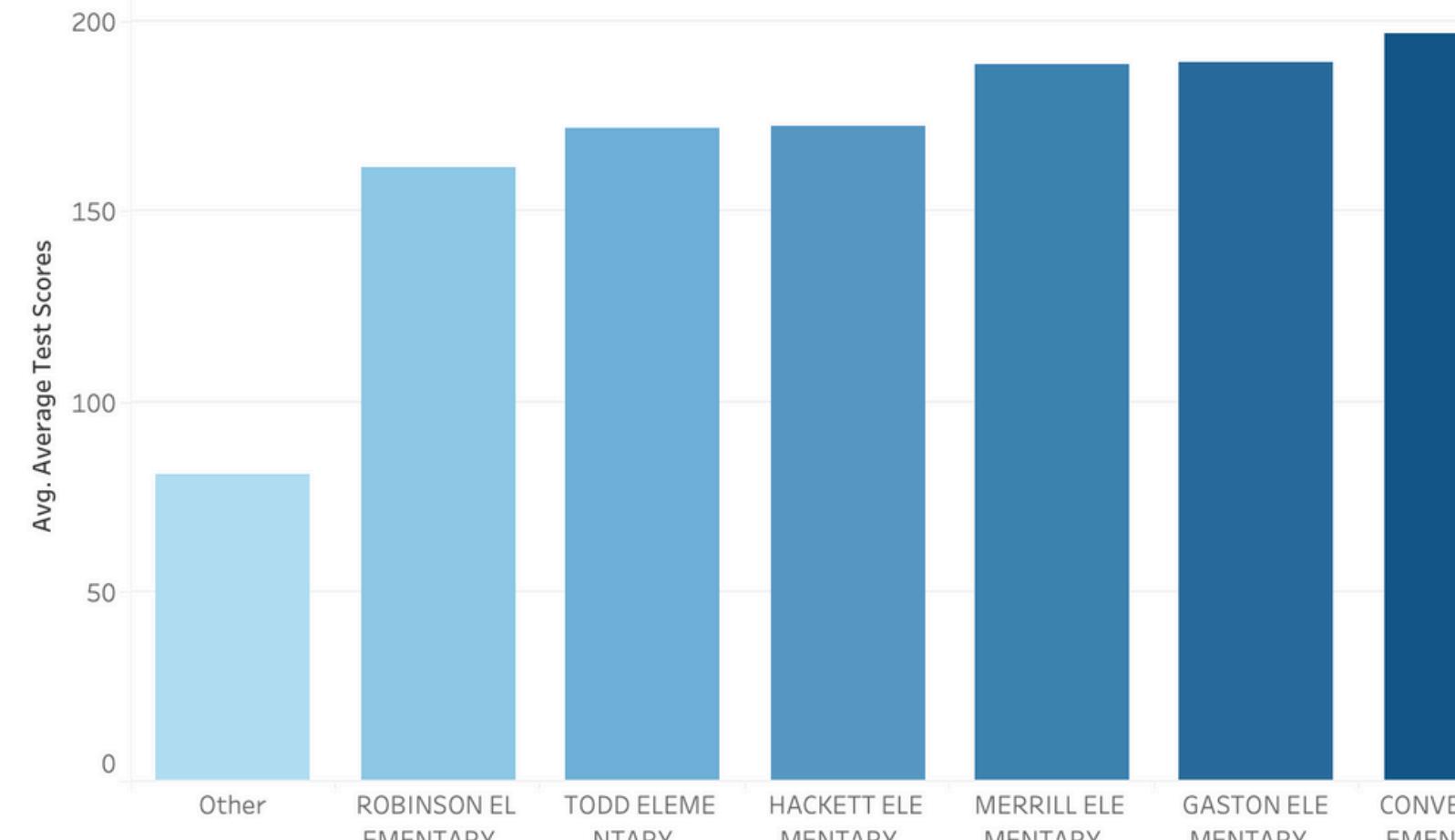
- **R-programming**

- Deeper linear regression model trainings, model evaluation ( $R^2$ , MAE).

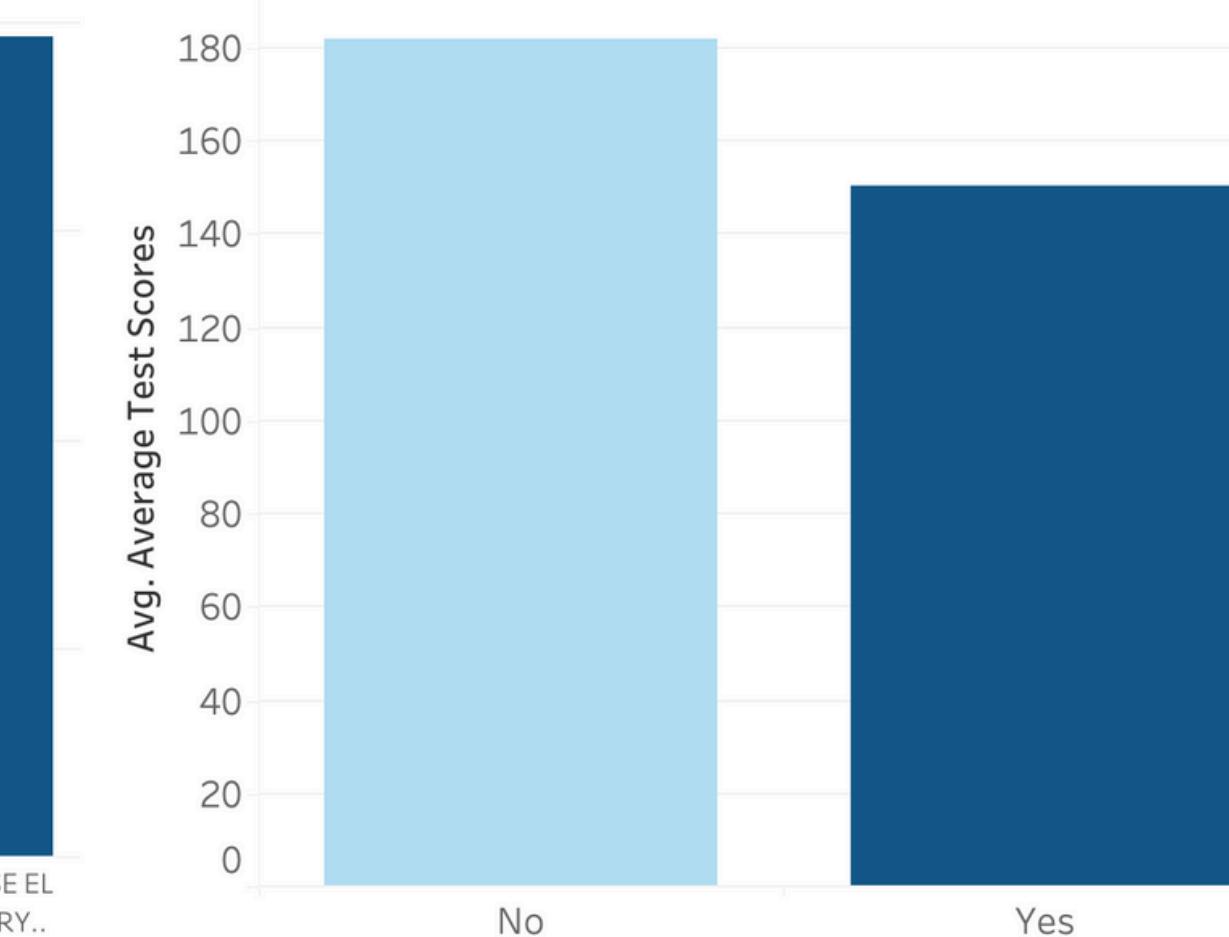


# DESCRIPTIVE VISUALS

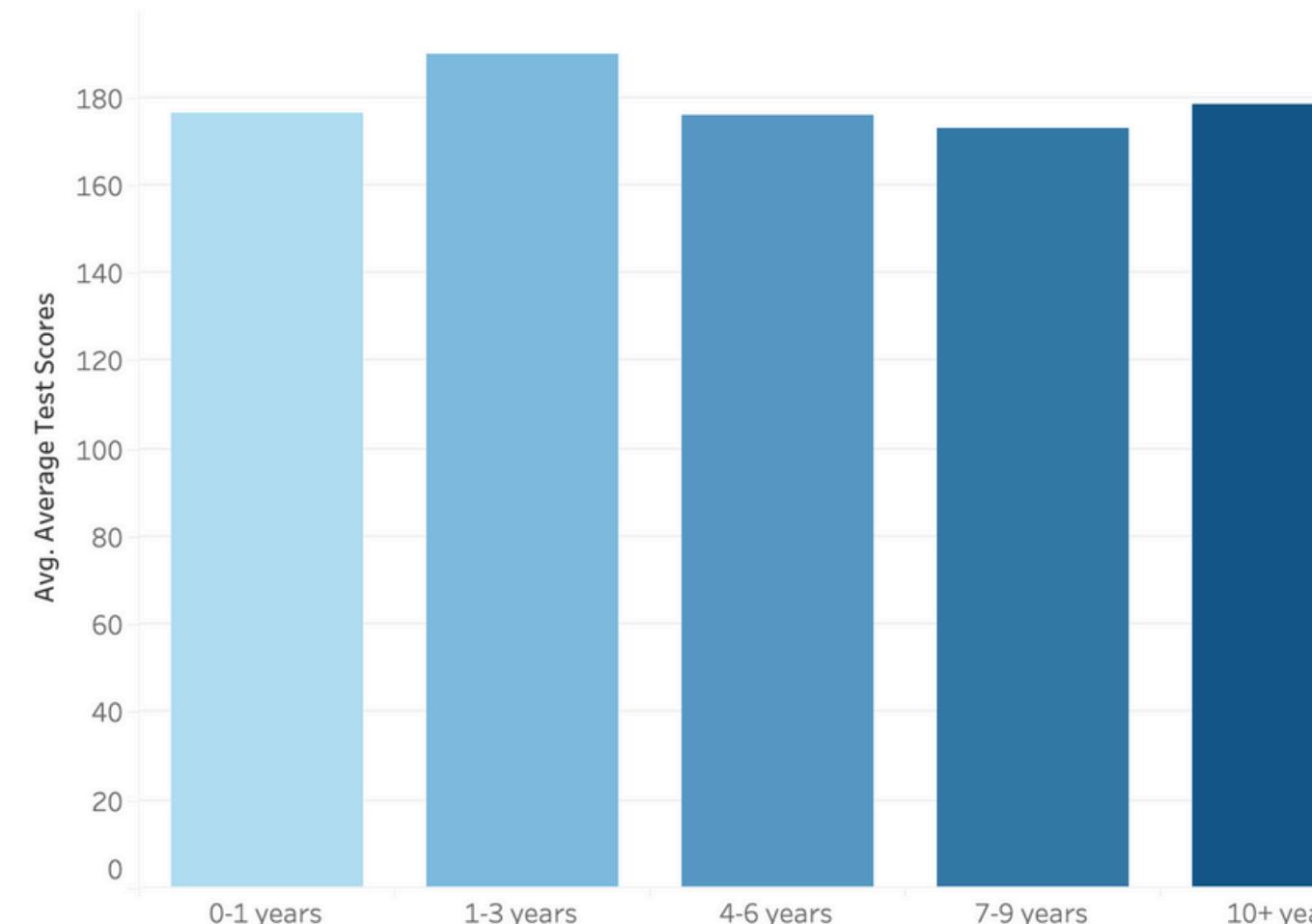
Average Test Scores for Each School



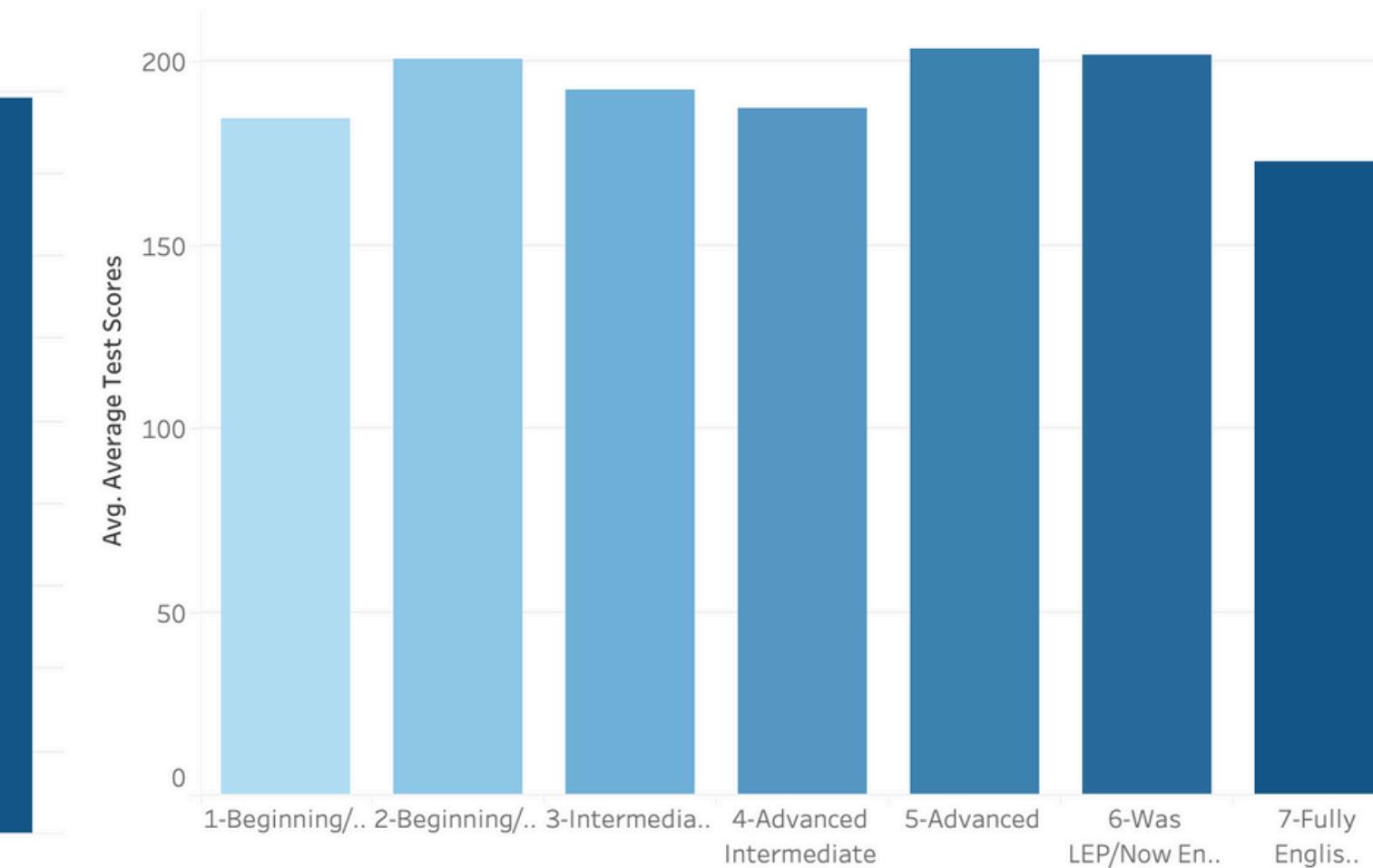
Avg Test Scores vs Special Education Needs



Average Test Scores vs. Number of Years Lived in Beloit



English Proficiency Level vs. Average Test Score



# OUR RESEARCH QUESTION

"How do factors such as old schooling, special education status, English proficiency, and duration of residency in Beloit influence academic achievement among students?"

# EXPERIMENTAL MODEL

$$\text{AA}_i = \beta_0 + \beta_1 \times (\text{OSL})_1 + \beta_2 \times (\text{SE})_2 + \beta_3 \times (\text{EP})_3 + \beta_4 \times (\text{NY})_4 + \epsilon_i$$

**AA** → Academic Achievement (Outcome)

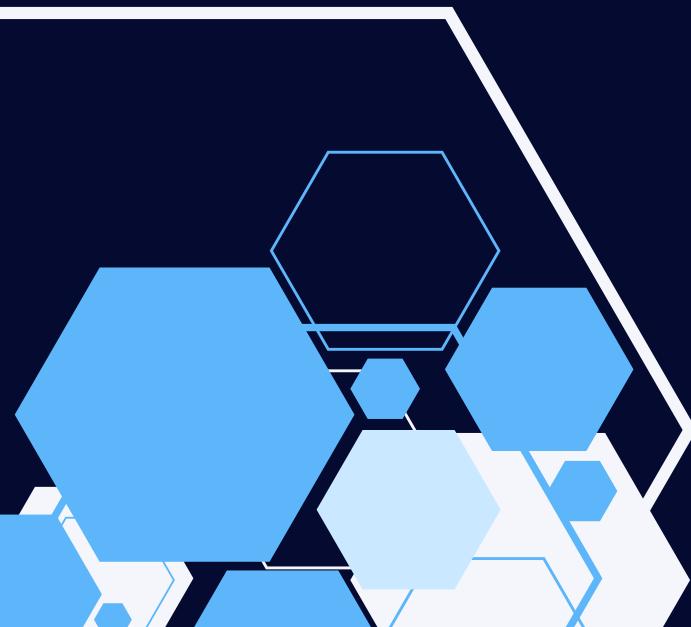
**OSL** → Old School Label

**SE** → Special Education

**EP** → English Proficiency

**NY** → # of years lived in Beloit

**$\epsilon_i$**  → Error Term



# OUTCOME

Residuals:

Min	1Q	Median	3Q	Max
-103.330	-16.130	3.034	21.150	73.391

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	119.11892	28.40135	4.194	4.25e-05	***
old_School_Label	6.01077	1.33246	4.511	1.15e-05	***
Sex	-4.61760	4.97522	-0.928	0.354562	
Special_Ed	-27.52978	12.04914	-2.285	0.023466	*
Disability	1.22525	9.65305	0.127	0.899135	
Fed_RaceAsian	-23.71699	37.00024	-0.641	0.522323	
Fed_RaceBlack or African American	-9.17626	9.25620	-0.991	0.322810	
Fed_Racewhite	0.05034	7.61560	0.007	0.994734	
Eng_Prof	6.12752	1.69541	3.614	0.000389	***
Q1_3	2.66060	3.54744	0.750	0.454208	
Q1_7	2.84274	2.02354	1.405	0.161756	
Q2_1	1.31662	4.73769	0.278	0.781399	
Q2_2	-4.20287	4.21271	-0.998	0.319753	
children_Per_Adult	8.69045	7.19923	1.207	0.228930	
Q2_3	-4.92158	2.30116	-2.139	0.033775	*
Q2_5	0.76982	0.93070	0.827	0.409231	
Q3_1	8.28008	8.76677	0.944	0.346161	
Q3_2	-2.11593	1.96231	-1.078	0.282320	
Q3_4	-0.36004	1.61714	-0.223	0.824062	
Q4_1	-0.75359	5.68624	-0.133	0.894711	
Q4_2	0.19277	2.27620	0.085	0.932599	
Q4_4	1.23445	1.68410	0.733	0.464489	
Q5_1	2.69052	7.17492	0.375	0.708100	
Q6_1	0.50049	2.14803	0.233	0.816022	
Q6_3	-0.04626	2.72117	-0.017	0.986456	
Q6_4	1.63471	2.81237	0.581	0.561777	
Q6_5	-0.81766	2.06757	-0.395	0.692955	
Q6_7	-2.30984	2.92448	-0.790	0.430645	
Q6_8	2.44345	2.75605	0.887	0.376465	

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 33.66 on 184 degrees of freedom

Multiple R-squared: 0.3431, Adjusted R-squared: 0.2431

F-statistic: 3.432 on 28 and 184 DF, p-value: 2.803e-07

# OUTCOME

Minimum	1st Quartile	Median	3rd Quartile	Maximum
-111.34	-15.57	5.303	22.354	73.967

Coefficients	Estimate Std.	Error	t value	Pr(> t )
Intercepts	143.081	9.848	14.529	< 2e^-16 ***
Old_School_Label	5.974	1.238	4.825	2.70e^-06 ***
Special_Ed	-29.854	7.491	-3.985	9.32e^-05 ***
Eng_Prof	5.876	1.162	5.055	9.41e^-07 ***
Q2_3 (Years in Beloit)	-4.561	2.091	-2.181	0.0303 *

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

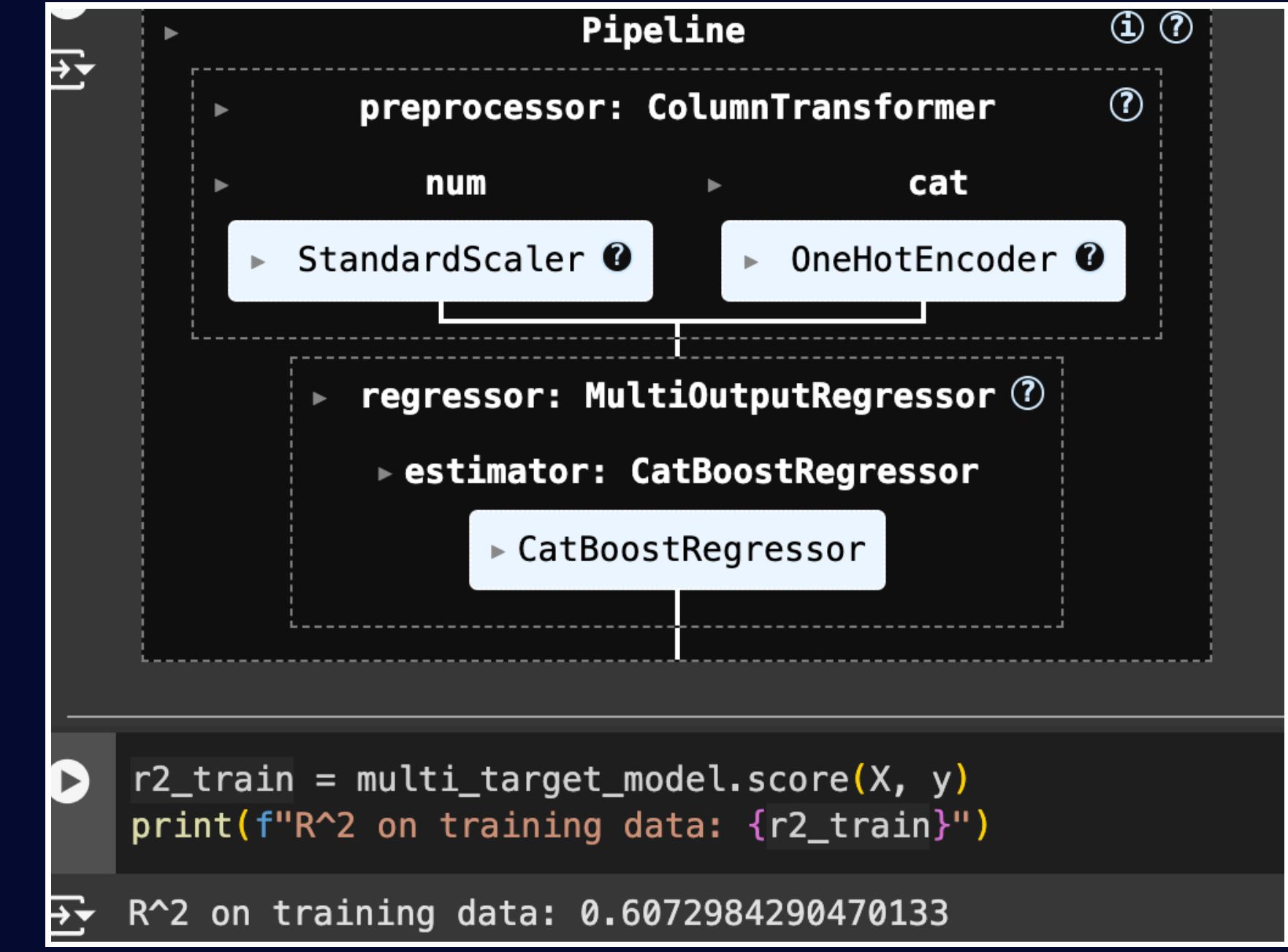
Residual standard error: 33.13 on 208 degrees of freedom  
Multiple R-squared: 0.2808, Adjusted R-squared: 0.267  
F-statistic: 20.3 on 4 and 208 DF, p-value: 3.919e-14

# OTHER MODELS

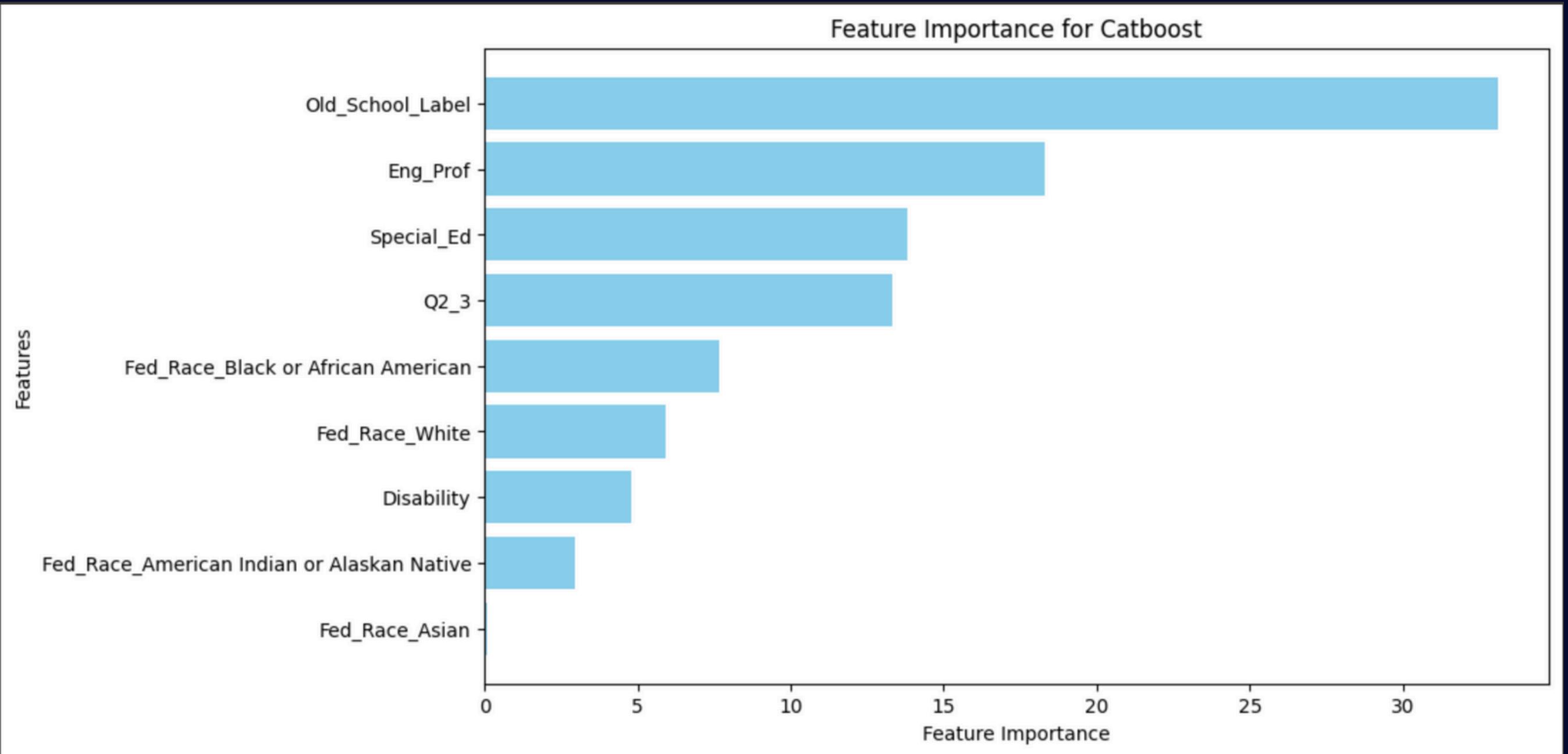
## 1. XGBoost: - 0.6 ( $R^2$ )



## 2. Catboost - 0.6 ( $R^2$ )



# FEATURE IMPORTANCE



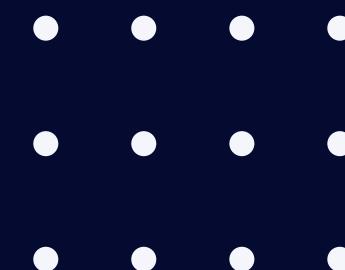
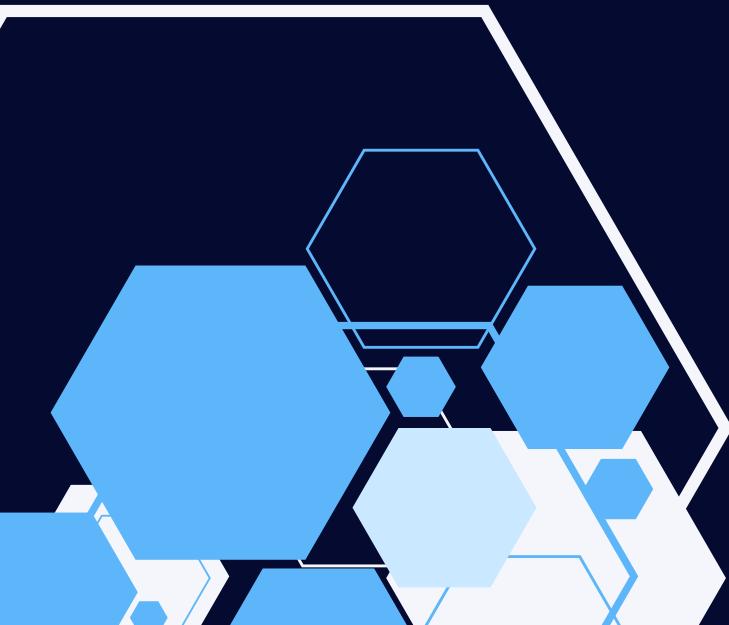
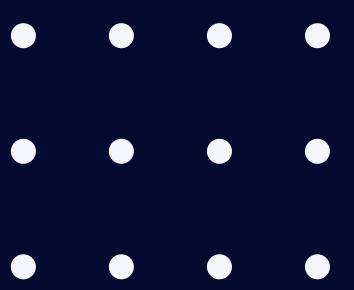
MAE FOR MATH 18-19: 20.72939415306791

MAE FOR READ 18-19: 32.20628298305619

OVERALL MAE: 26.46783856806205

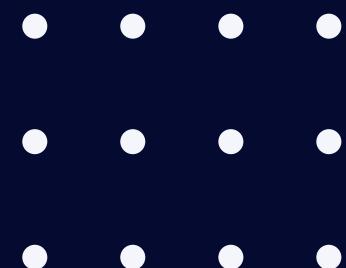
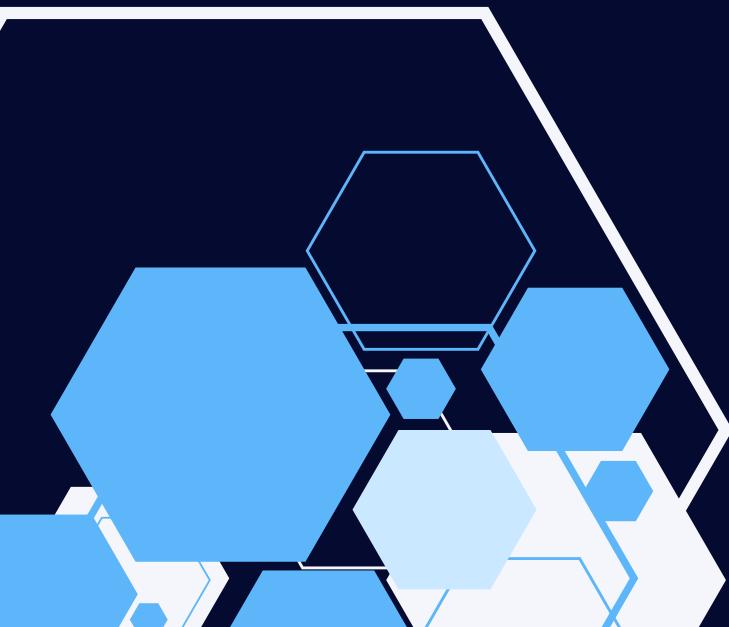
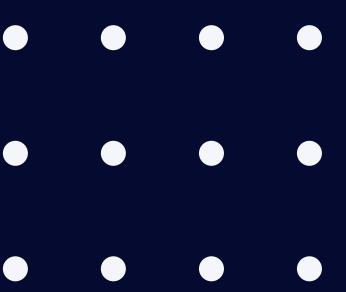
# POLICY RECOMMENDATIONS

1. Increase support for special education programs. This could include additional resources, specialized staff, and adaptive learning materials to address the needs of these students. Personalized learning plans and dedicated mentorship could help mitigate the negative impact.



# POLICY RECOMMENDATIONS

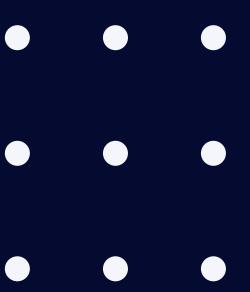
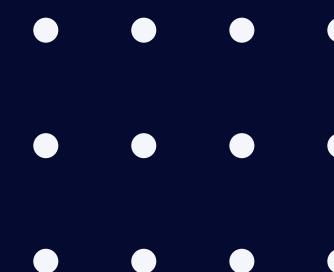
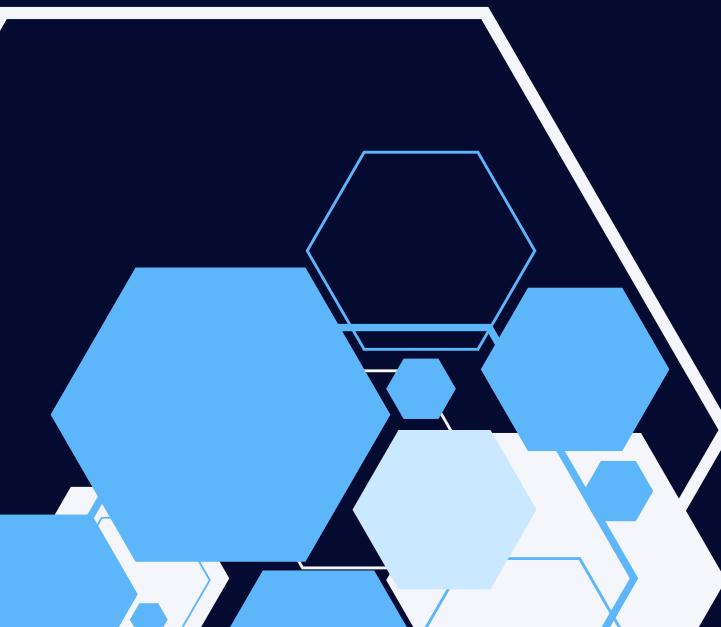
2. We believe that in order to support the student well-being schools must implement school-based health screenings and mental health resources, ensuring students have access to necessary healthcare services.



# POLICY RECOMMENDATIONS

3. Establish a comprehensive Mandatory English Language Development Program within educational institutions, aimed at students with lower English proficiency levels. This program should include the following key components:

- Curriculum Design, Assessment & Placement, & Monitoring Support



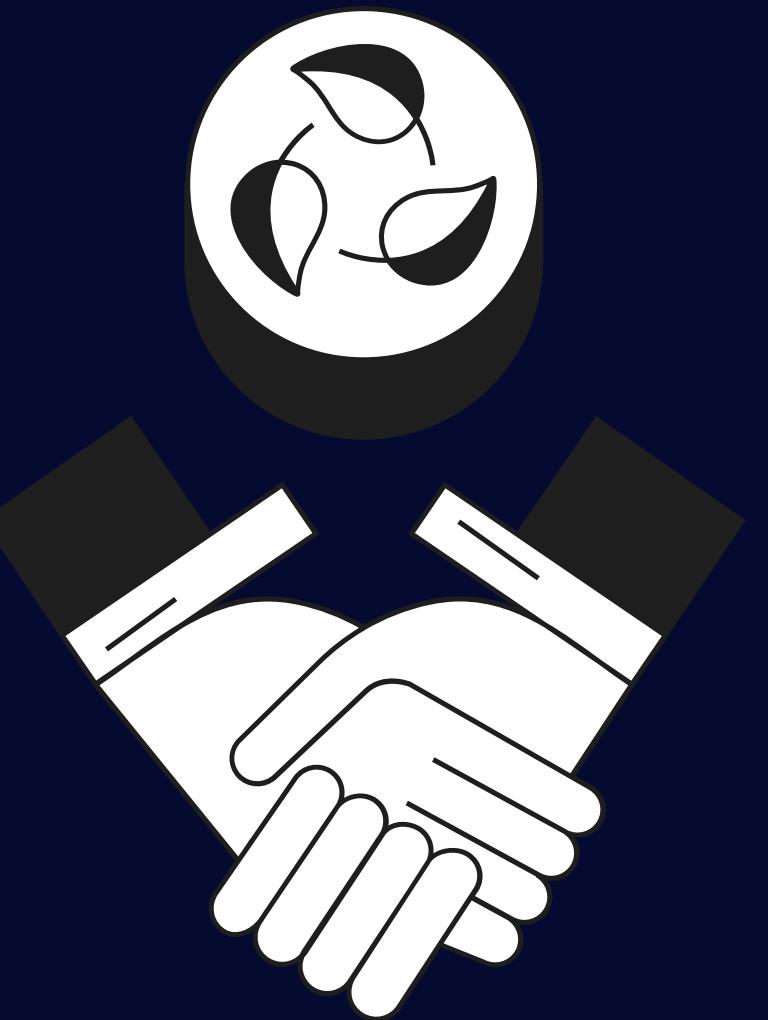
# Limitations of our Analysis !

- Data quality and completeness
- Self reported data
- Number of observations
- Lower R-squared value
- Time period of the survey taken
- Biased data



# Ethical Considerations

- Confidentiality and Privacy
- Informed Consent
- Accurate Representation of Data
- Impact on Educational Outcomes:
- Bias and Fairness
- Transparency



# Acknowledgements

We acknowledge the use of artificial intelligence tools in outlining our case study and assisting with code cleaning for our model. These AI technologies were instrumental in enhancing our understanding and streamlining our analysis, contributing significantly to the precision and efficiency of our research.

We also extend our gratitude to Dr.Phan for sharing the data with us.

# THANK YOU

Any Questions?