Bias Analysis towards Fair AI in Education

By utilizing Student Score Prediction Dataset

Team: Multi-Agents
Artur Ganzha, Detijon Lushaj, Lasse Müller

Problem Definition

- Identify potential inequalities in student performance based on race/ethnicity, parental education and gender.
- Addressing unfair opportunities is crucial for creating a more equitable learning environment.
- Target Group:
  - Educators, researchers, and students interested in educational equity.
- Use of Insights:
  - Adjusting teaching strategies and student grouping to promote fairness.
  - Provides actionable insights to help create equitable educational environments.

Goals and Objectives

- Showcase inequalities in student performance.
- Provide data-driven insights to support fairer educational practices.
- Validate through ML models

- **Key Questions:**
    - Are there disparities in performance among different groups?
    - Could these disparities suggest unfair opportunities?
    - How can educators use this data to reduce bias?

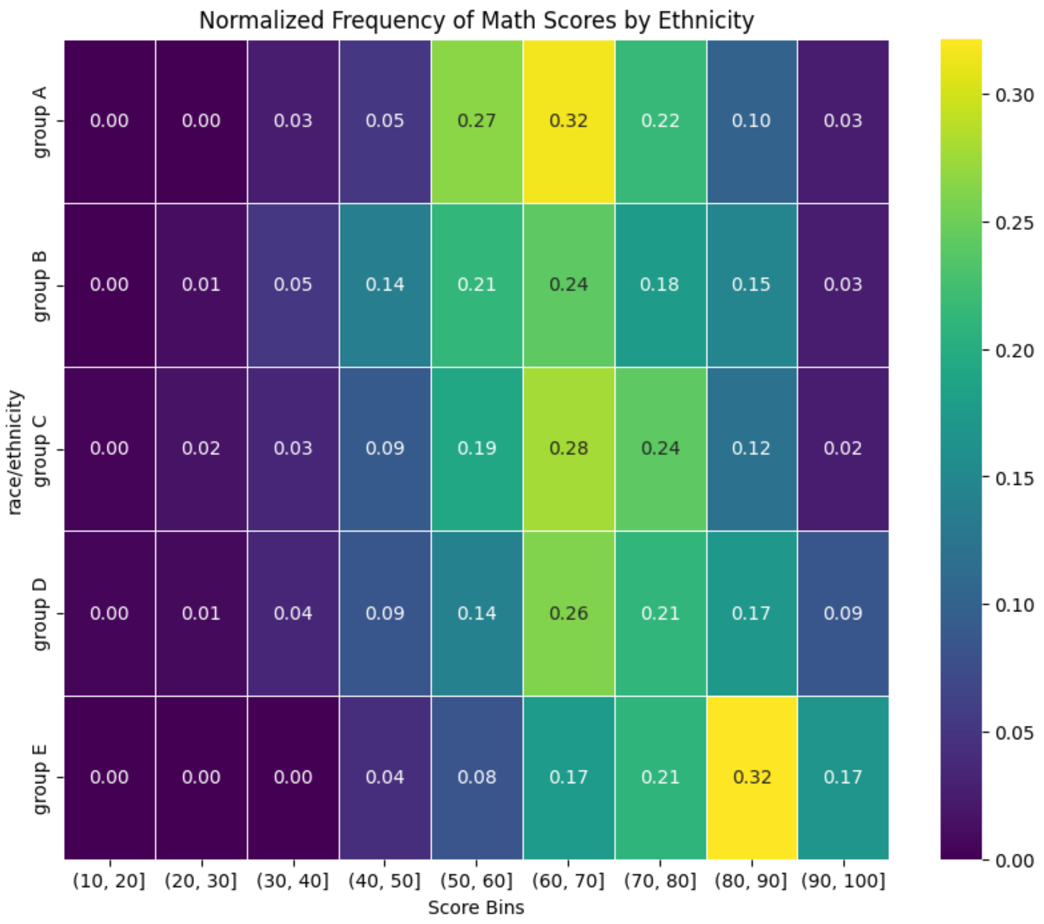
- **Goals:**
    - Highlight inequalities in student performance.
    - Provide insights for educators to enhance fairness in educational practices.

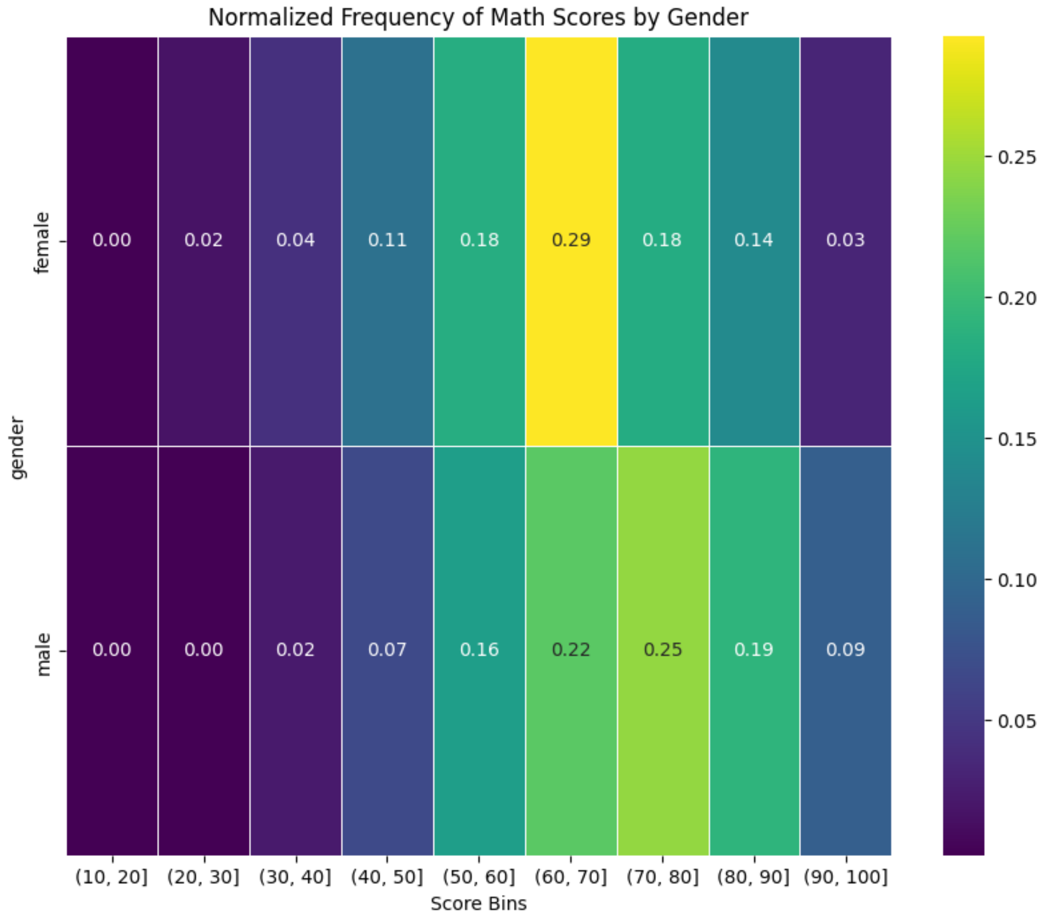
Student Performance Prediction dataset

- Data Source: Kaggle - Student Performance Prediction dataset.
  - Gender
  - race/ethnicity
  - parental level of education
  - test scores across subjects (math, reading, writing)

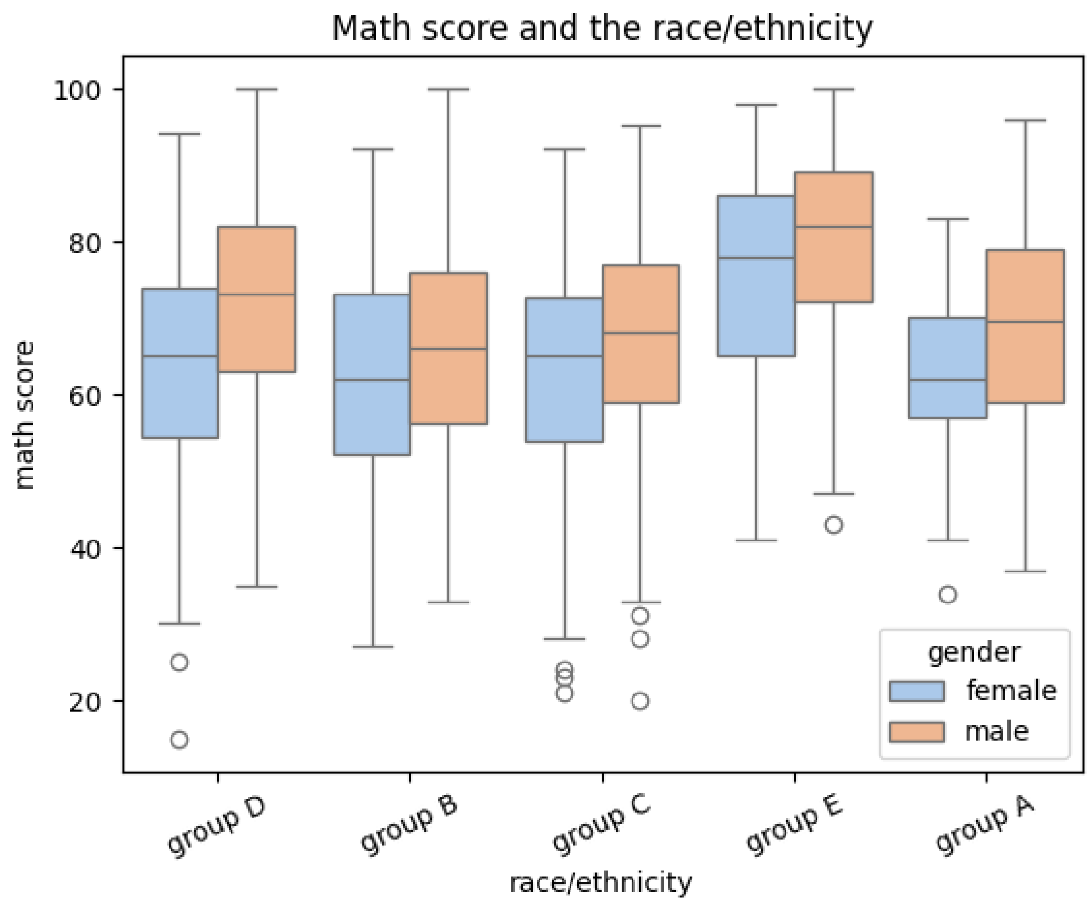
	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
0	female	group D	some college	standard	completed	59	70	78
1	male	group D	associate's degree	standard	none	96	93	87
2	female	group D	some college	free/reduced	none	57	76	77
3	male	group B	some college	free/reduced	none	70	70	63
4	female	group D	associate's degree	standard	none	83	85	86

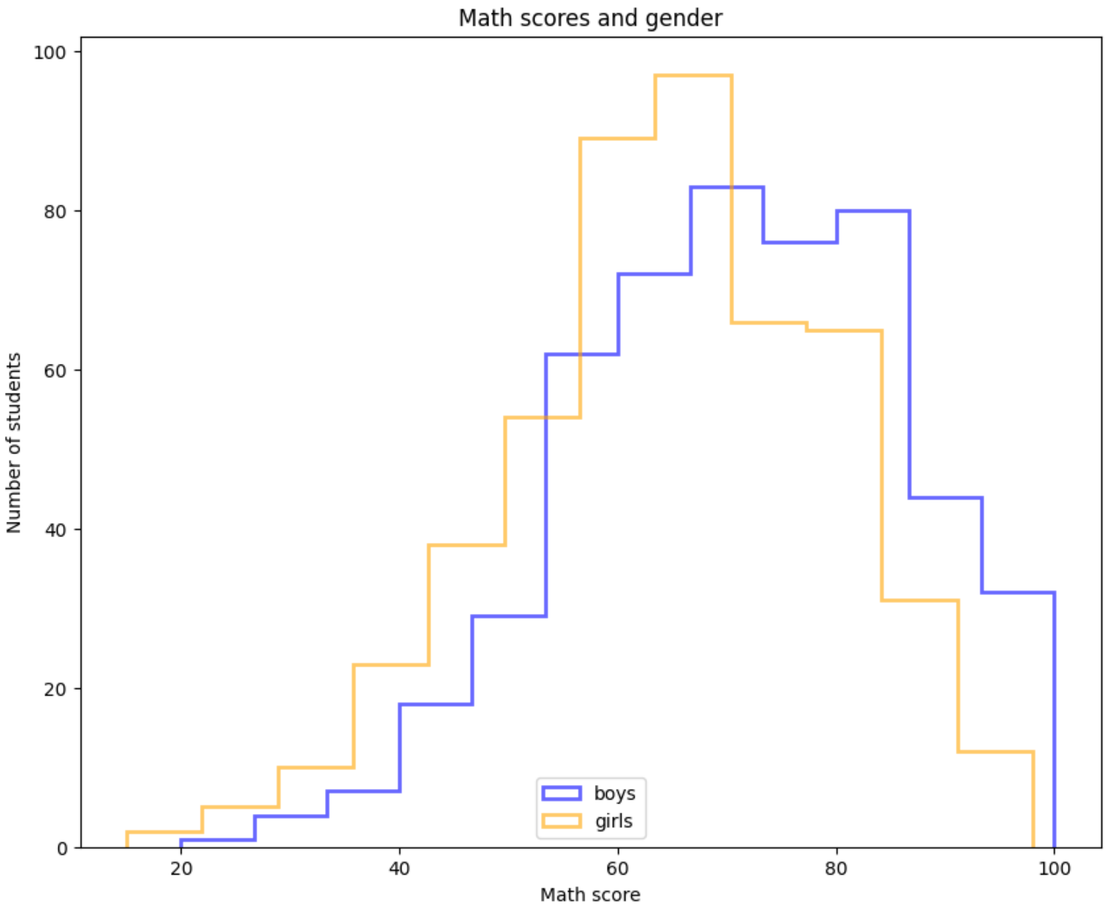
Student Performance Prediction dataset





Student Performance Prediction dataset





ANOVA Table

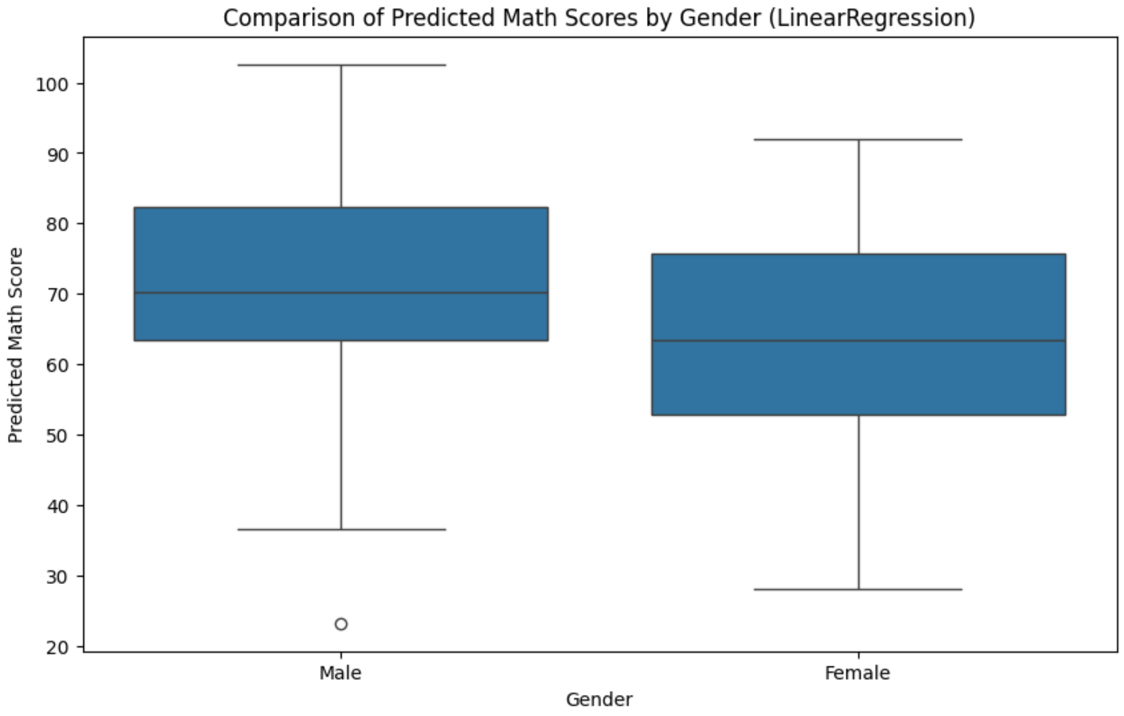
	sum_sq	df	F	PR(>F)
C(gender)	8092.095570	1.0	40.159076	3.551026e-10
C(parents_education)	6351.401902	5.0	6.304088	9.051222e-06
C(race)	18155.689375	4.0	22.525553	7.968034e-18

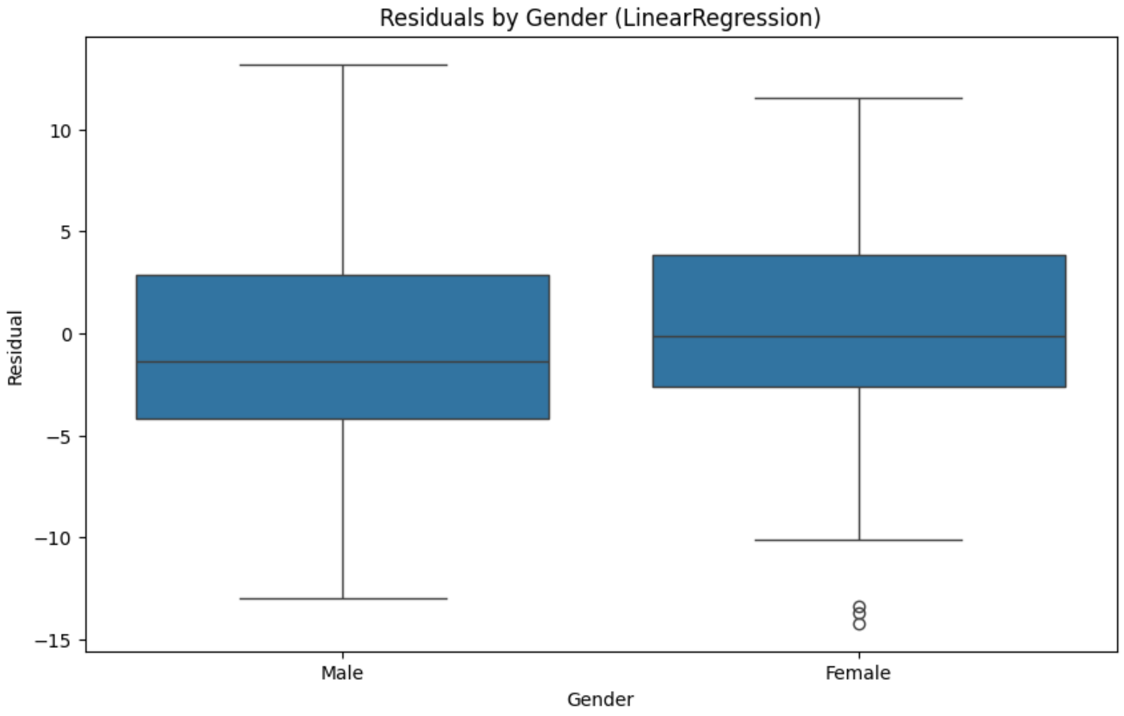
- **Gender (C(gender)):**
  - F-statistic: **40.16**, p-value: **3.55e-10**
  - Strong evidence that gender impacts student performance.
- **Parental Education (C(parents_education)):**
  - F-statistic: **6.30,** p-value: **9.05e-06**
  - Suggests a significant, but lesser impact compared to gender.
- **Race (C(race)):**
  - F-statistic: **22.53,** p-value: **7.97e-18**
  - Indicates a strong influence of race on student performance.

Methodology

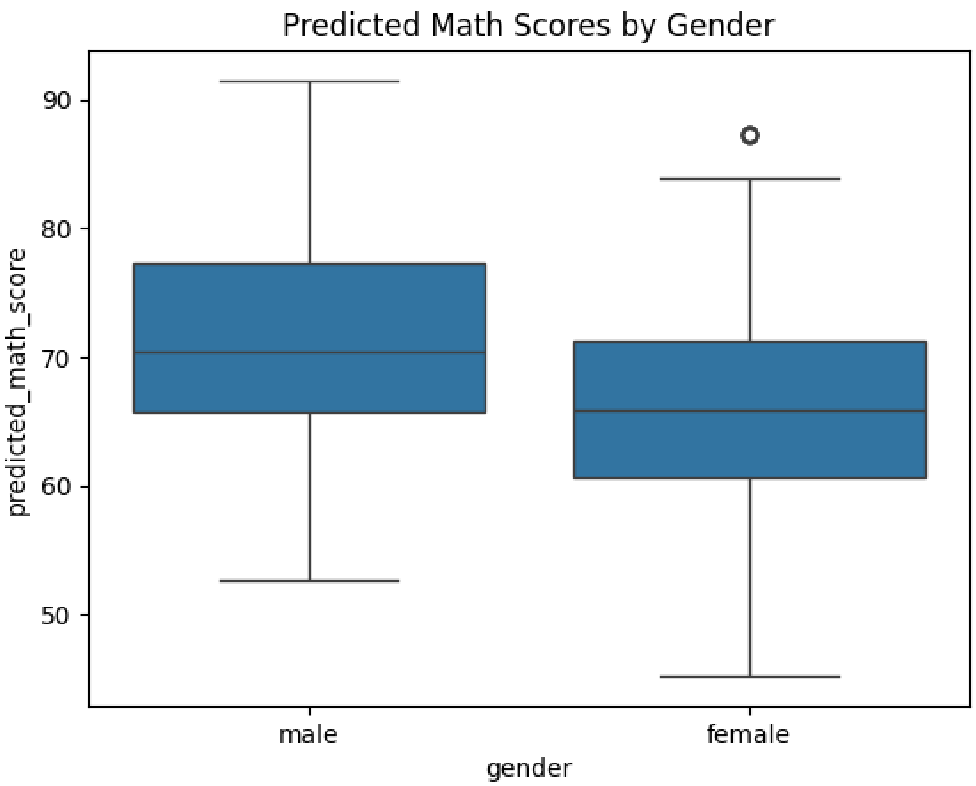
- Data Preparation:
  - Trainings/Test split
  - Normalization of Data
- Model Selection:
  - Linear Regression, Random Forests
  - Neural Networks
    - 3 hidden layer, ReLU, Dropout-rate 0.2
    - Learning-rate 0.0005
- Data Splitting:
  - K-Fold Cross-Validation
    - 70-20-10 split
    - K = 5
- Metrics for Evaluation:
  - R2 Score
  - MSE/MAE

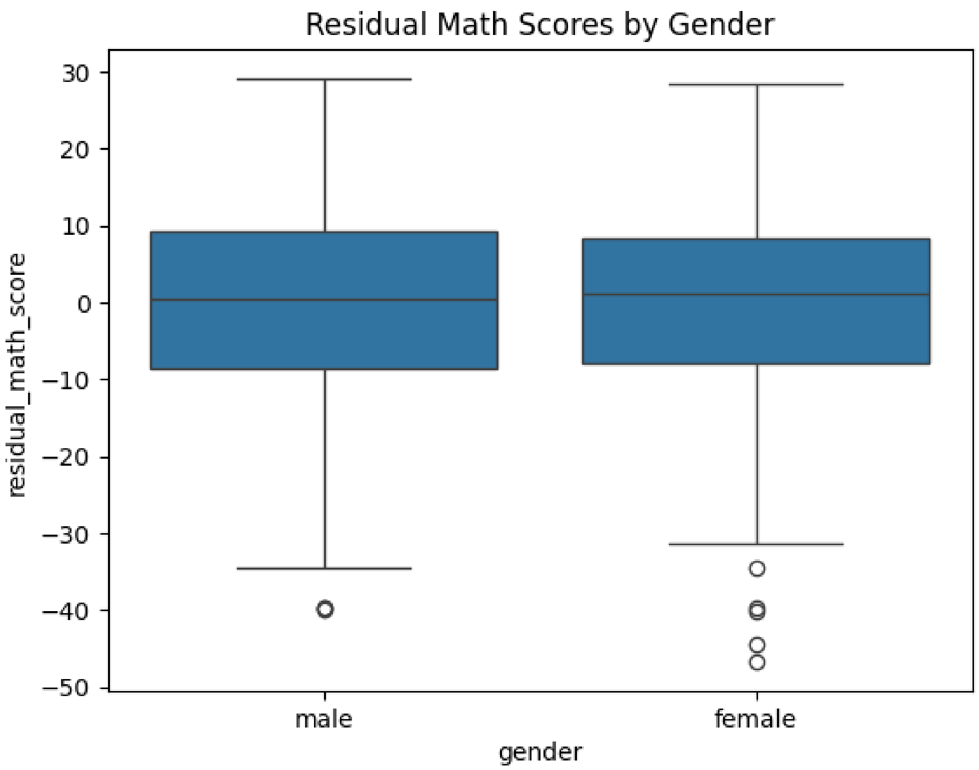
Linear Regression



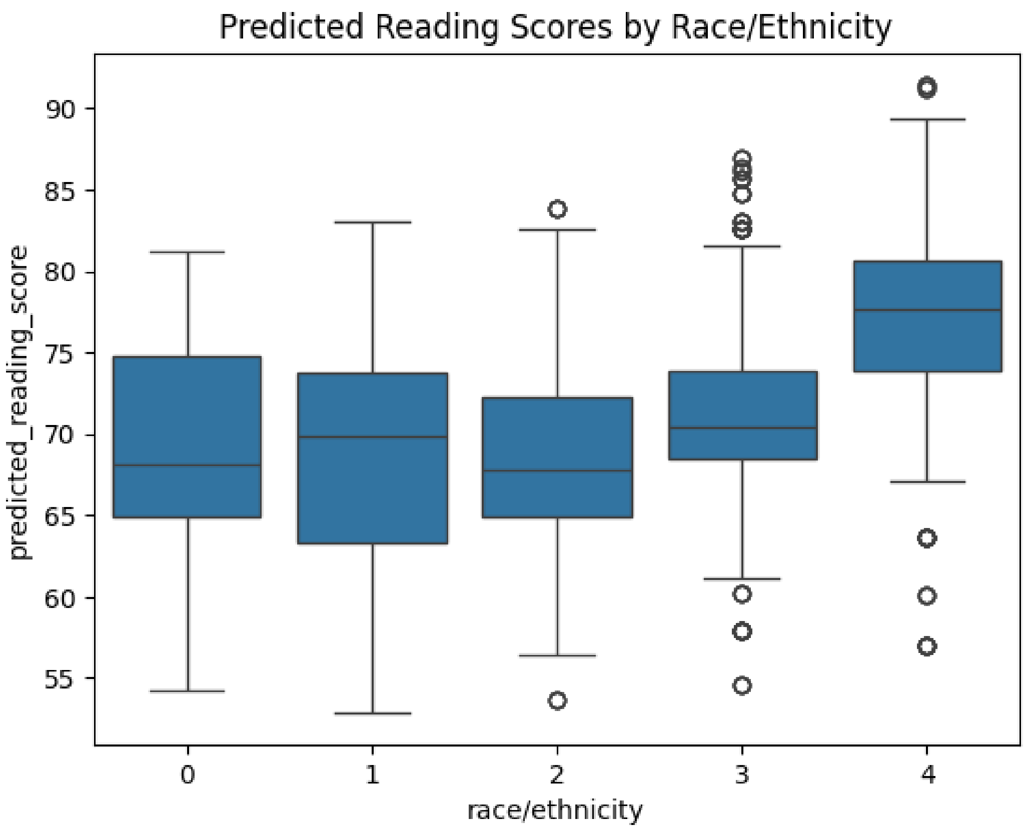


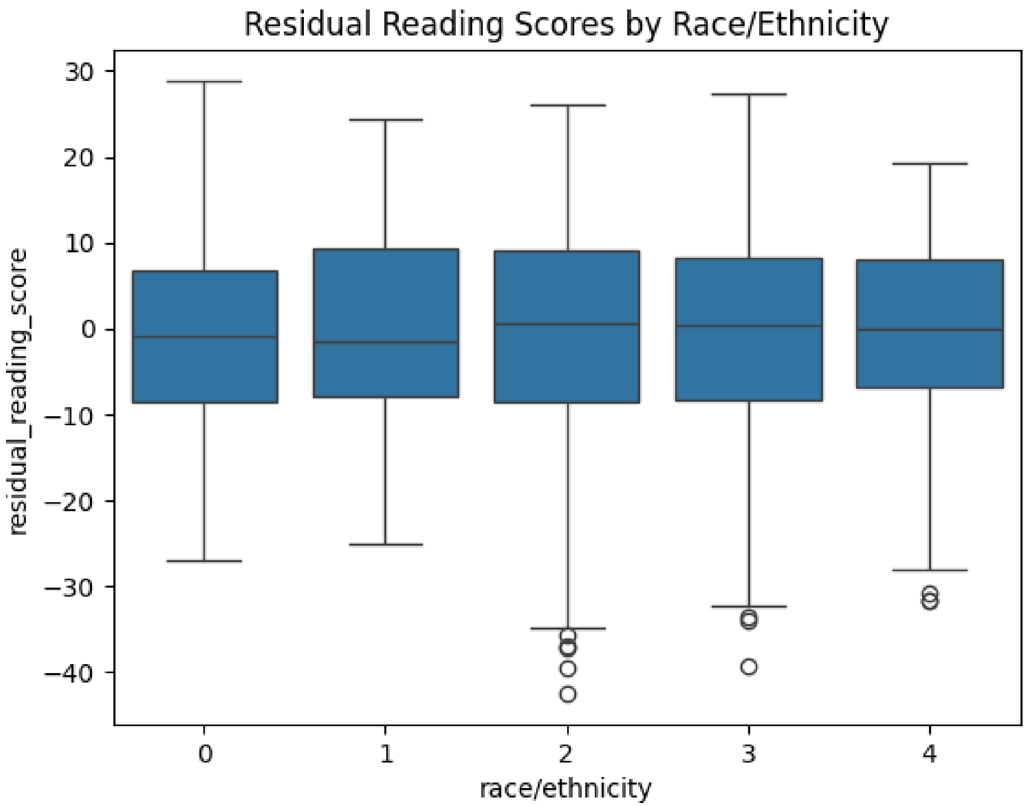
Neural Networks



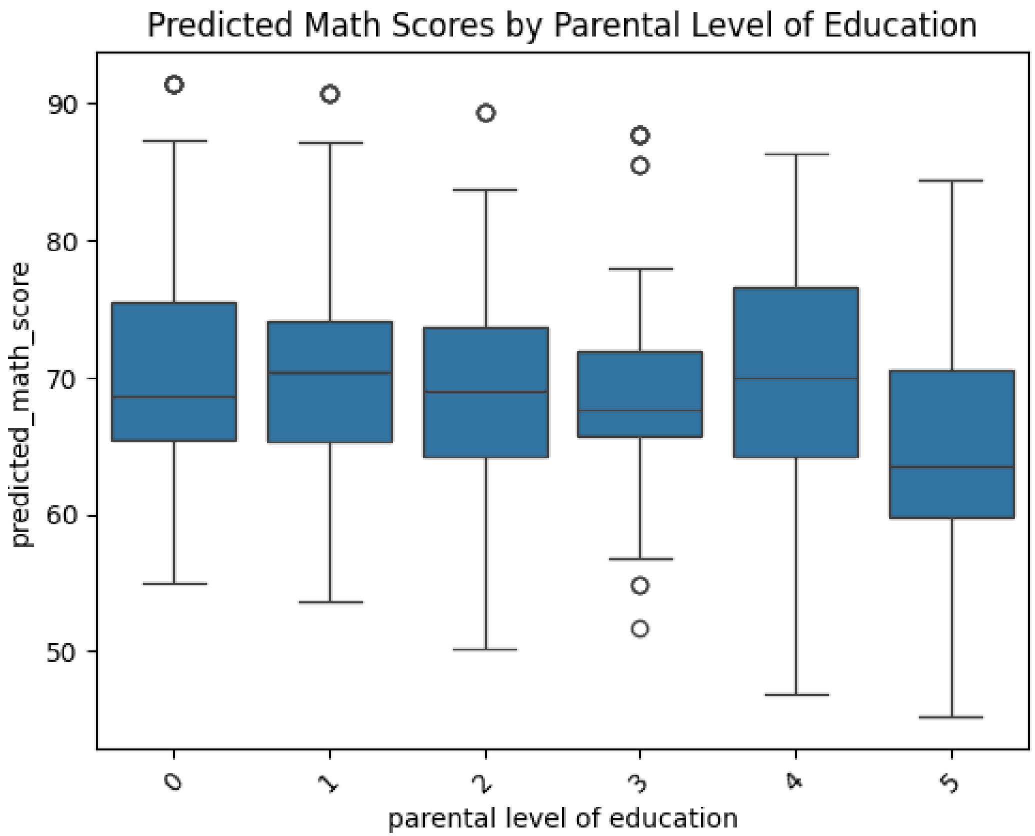


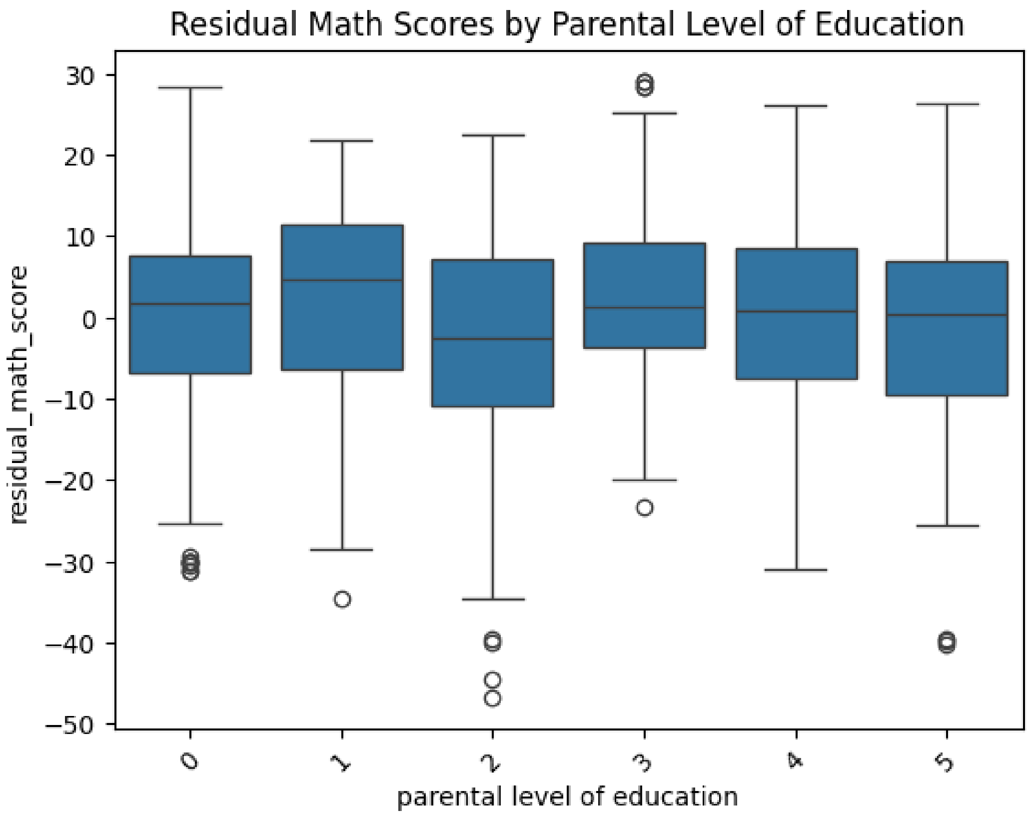
Neural Networks





Neural Networks





Key Questions: Answers

- Are there disparities in performance among different groups?
  - Yes there are disparities, but they are not necessarily caused by biases
- Could these disparities suggest unfair opportunities?
  - These disparities do not favor / disfavor new students (backed by prediction results)
- How can educators use this data to reduce bias?
  - Educators can shuffle student groups during task assignments to create more diversity

Results

- The prediction models gave similar results
  - We decided going with neural networks, due to flexibility
- There are disparities within gender and ethnicity
  - But the prediction models showed, that they have little to no residual
  - Hence we can conclude that the model reduces any bias as no class is favored / disfavored by the model
- Key take aways: teachers can shuffle students who are performing well with students with worse performances to increase their respective influence

Review of the Hackathon

- What went well?
  - Finding the right topic
  - Analyzing and understanding the dataset
  - Teamwork
- What can we improve next time?
  - Dig deeper into the data
  - Environment setup
  - Incorporate fairness metrics

Thank you for your attention!

https://github.com/arzx/bias-analysis-students