Group: Analytics Avengers

PROEx Project:

Understanding the Impact of Student Background on Academic Performance

Project Goal

To analyze and understand how various background factors influence students' academic performance using two complementary datasets.

Student-Parent Dataset

1. Source: Portuguese secondary education
2. Sample size: 1044 students
3. Focus: Family background and academic performance
4. Key features:
   1. Parent education levels
   2. Urban/Rural location
   3. Gender
   4. Final grades (G3)

Academic-Social Dataset

- Source: Research Information study (Kaggle)
- Sample size: 493 students
- Focus: Social and academic factors
- Key features:
  - Academic scores (HSC, SSC)
  - Social factors (Gaming, Extra activities)
  - Study habits (Attendance, Preparation time)
  - Family income

- **Research Questions**

1. How do family background factors affect student performance?
2. What is the relative importance of social vs academic factors?
3. Are there consistent patterns across different educational contexts?

- **Methodology**

- Feature encoding for categorical variables
- Linear regression analysis
- Visualization plots
- Comparative visualization of effects across the dataset.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob	reason	guardian	traveltime	studytime	failures	schoolsup	famsup	paid	activities	nursery	higher	internet	romantic	famrel	freetime	goout
2	GP	F	18	U	GT3	A	4	4	at_home	teacher	course	mother	2	2	0	yes	no	no	no	yes	yes	no	no	4	3	
3	GP	F	17	U	GT3	T	1	1	at_home	other	course	father	1	2	0	no	yes	no	no	no	yes	yes	no	5	3	
4	GP	F	15	U	LE3	T	1	1	at_home	other	other	mother	1	2	3	yes	no	yes	no	yes	yes	yes	no	4	3	
5	GP	F	15	U	GT3	T	4	2	health	services	home	mother	1	3	0	no	yes	yes	yes	yes	yes	yes	yes	3	2	
6	GP	F	16	U	GT3	T	3	3	other	other	home	father	1	2	0	no	yes	yes	no	yes	yes	no	no	4	3	
7	GP	M	16	U	LE3	T	4	3	services	other	reputation	mother	1	2	0	no	yes	yes	yes	yes	yes	yes	no	5	4	
8	GP	M	16	U	LE3	T	2	2	other	other	home	mother	1	2	0	no	no	no	no	yes	yes	yes	no	4	4	
9	GP	F	17	U	GT3	A	4	4	other	teacher	home	mother	2	2	0	yes	yes	no	no	yes	yes	no	no	4	1	
10	GP	M	15	U	LE3	A	3	2	services	other	home	mother	1	2	0	no	yes	yes	no	yes	yes	yes	no	4	2	
11	GP	M	15	U	GT3	T	3	4	other	other	home	mother	1	2	0	no	yes	yes	yes	yes	yes	yes	no	5	5	
12	GP	F	15	U	GT3	T	4	4	teacher	health	reputation	mother	1	2	0	no	yes	yes	no	yes	yes	yes	no	3	3	
13	GP	F	15	U	GT3	T	2	1	services	other	reputation	father	3	3	0	no	yes	no	yes	yes	yes	yes	no	5	2	
14	GP	M	15	U	LE3	T	4	4	health	services	course	father	1	1	0	no	yes	yes	yes	yes	yes	yes	no	4	3	

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
Department	Gender	HSC	SSC	Income	Hometown	Computer	Preparation	Gaming	Attendance	Job	English	Extra	Semester	Last	Overall		
Business Administration	Male	4.17	4.84	Low (Below 15,000)	Village	3	More than 3 Hours	0-1 Hour	80%-100%	No	3	Yes	6th	3.22	3.35		
Business Administration	Female	4.92	5	Upper middle (30,000-50,000)	City	3	0-1 Hour	0-1 Hour	80%-100%	No	3	Yes	7th	3.467	3.467		
Business Administration	Male	5	4.83	Lower middle (15,000-30,000)	Village	3	0-1 Hour	More than 3 Hours	80%-100%	No	4	Yes	3rd	4	3.72		
Business Administration	Male	4	4.5	High (Above 50,000)	City	5	More than 3 Hours	More than 3 Hours	80%-100%	No	5	Yes	4th	3.8	3.75		
Business Administration	Female	2.19	3.17	Lower middle (15,000-30,000)	Village	3	0-1 Hour	2-3 Hours	80%-100%	No	3	Yes	4th	3.94	3.94		
Computer Science and Engineering	Male	4.75	4.05	Lower middle (15,000-30,000)	Village	3	0-1 Hour	More than 3 Hours	Below 40%	No	4	No	2nd	1	1		
Computer Science and Engineering	Male	4.42	5	High (Above 50,000)	Village	4	0-1 Hour	More than 3 Hours	60%-79%	No	2	No	2nd	1.06	1.06		
Computer Science and Engineering	Male	4.5	4.81	Upper middle (30,000-50,000)	City	3	2-3 Hours	More than 3 Hours	80%-100%	No	4	Yes	11th	2.95	1.25		
Computer Science and Engineering	Male	3.32	4.5	Low (Below 15,000)	City	4	0-1 Hour	More than 3 Hours	Below 40%	No	3	Yes	5th	1.42	1.44		
Computer Science and Engineering	Female	3.33	4.95	Lower middle (15,000-30,000)	City	3	0-1 Hour	More than 3 Hours	Below 40%	No	4	No	2nd	1.5	1.5		
Computer Science and Engineering	Male	4.17	5	Lower middle (15,000-30,000)	Village	3	0-1 Hour	More than 3 Hours	Below 40%	No	4	No	2nd	1.56	1.56		
Computer Science and Engineering	Male	3.58	4.56	Upper middle (30,000-50,000)	Village	4	0-1 Hour	More than 3 Hours	Below 40%	No	4	No	9th	1.57	1.57		
Computer Science and Engineering	Male	4.5	4.75	Lower middle (15,000-30,000)	Village	1	0-1 Hour	More than 3 Hours	60%-79%	No	2	Yes	8th	1.57	1.57		

This phase involves understanding our data and transforming it into a format suitable for analysis. We need to ensure all variables are properly encoded for our statistical analysis.

Converting categorical variables into numerical values allows us to measure their impact on grades quantitatively.

Data Quality

- No missing values in both datasets
- Balanced class distributions
- Consistent grade scales

→ Clean, balanced data ensures reliable analysis results and valid conclusions.

- **Academic-Social Dataset**
- **Original Features:**
  - Academic: HSC, SSC scores, English level
  - Social: Gamin time, Extra activities
  - Study Habits: gPreparation time, Attendance
  - Background: Income level, Hometown
  - Target Variable: Overall grade
- **Preprocessing**
1. Numeric Conversion:
   1. Income levels (0-3 scale)
   2. Preparation/Gaming time (0-2 scale)
   3. Attendance (percentage)
2. **Binary Encoding:**
   1. Gender, Extra activities, Job status

   Standardizing different types of measurements (time, percentages, categorical) enables fair comparison of their effects.

- **Student-Parent Dataset**
- **Original Features:**
- Demographic: age, sex, address
- Family Background:
  - Parents' education (Medu, Fedu: 1-4 scale)
  - Parents' jobs (Mjob, Fjob)
  - Guardian type
- Target Variable: **Final grade (G3)**

- **Preprocessing:**
1. Binary Encoding:
   1. Gender (Male=1, Female=0)
   2. Urban/Rural (Urban=1, Rural=0)
   3. Parent jobs (Teacher=1, Other=0)
   4. Guardian types (Mother=1, Father=1, Other=0)

Student-Parent Dataset

6]:

	age	sex	Medu	Fedu	urban	mother_teacher	father_teacher	guardian_type	G3
0	18	F	4	4	1	0	1	0	6
1	17	F	1	1	1	0	0	1	6
2	15	F	1	1	1	0	0	0	10
3	15	F	4	2	1	0	0	0	15
4	16	F	3	3	1	0	0	1	10
...	...	...	...	...	...	...	...	...	...
644	19	F	2	3	0	0	0	0	10
645	18	F	3	1	1	1	0	0	16
646	18	F	1	1	1	0	0	0	9
647	17	M	3	1	1	0	0	0	10
648	18	M	3	2	0	0	0	0	11

1044 rows × 9 columns

Academic-Social Dataset

:

	Department	Gender	HSC	SSC	Income	Hometown	Computer	Preparation	Gaming	Attendance	Job	English	Extra	Semester	Last	Overall
0	Business Administration	1	4.17	4.84	0.0	0	3	2	0	1.00	0	3	1	6th	3.220	3.350
1	Business Administration	0	4.92	5.00	2.0	1	3	0	0	1.00	0	3	1	7th	3.467	3.467
2	Business Administration	1	5.00	4.83	1.0	0	3	0	2	1.00	0	4	1	3rd	4.000	3.720
3	Business Administration	1	4.00	4.50	3.0	1	5	2	2	1.00	0	5	1	4th	3.800	3.750
4	Business Administration	0	2.19	3.17	1.0	0	3	0	1	1.00	0	3	1	4th	3.940	3.940
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
488	Public Health	1	3.50	4.00	1.0	0	2	2	1	0.59	1	2	1	7th	3.250	3.100
489	Public Health	1	4.50	4.00	2.0	1	4	1	1	0.79	0	3	1	9th	3.500	3.670
490	Sociology	0	4.56	3.67	1.0	1	3	0	1	0.59	0	2	0	2nd	3.560	3.560
491	Sociology	1	4.23	4.50	0.0	0	3	2	1	1.00	0	3	1	10th	3.610	3.750
492	Sociology	0	3.83	5.00	1.0	0	3	2	1	1.00	0	4	1	12th	4.000	3.860

Impact of Background Factors on Student Grades

This horizontal bar chart shows the relative importance of different background factors on student grades, where:

1. Positive Impacts (from strongest):

   1. Urban Location (0.896): Students in urban areas significantly outperform rural students
   2. Mother's Education (0.472): Higher maternal education strongly correlates with better grades
   3. Father's Teaching Job (0.461): Having a father who teaches positively impacts performance

2. Negative Impacts:

   1. Being Male (-0.524): Female students tend to perform better academically
   2. Student Age (-0.230): Younger students generally achieve higher grades

→ **The graph clearly shows location and parental education/occupation are the most influential positive factors**.

Impact of Background Factors on Student Grades
Positive values indicate positive effect on grades
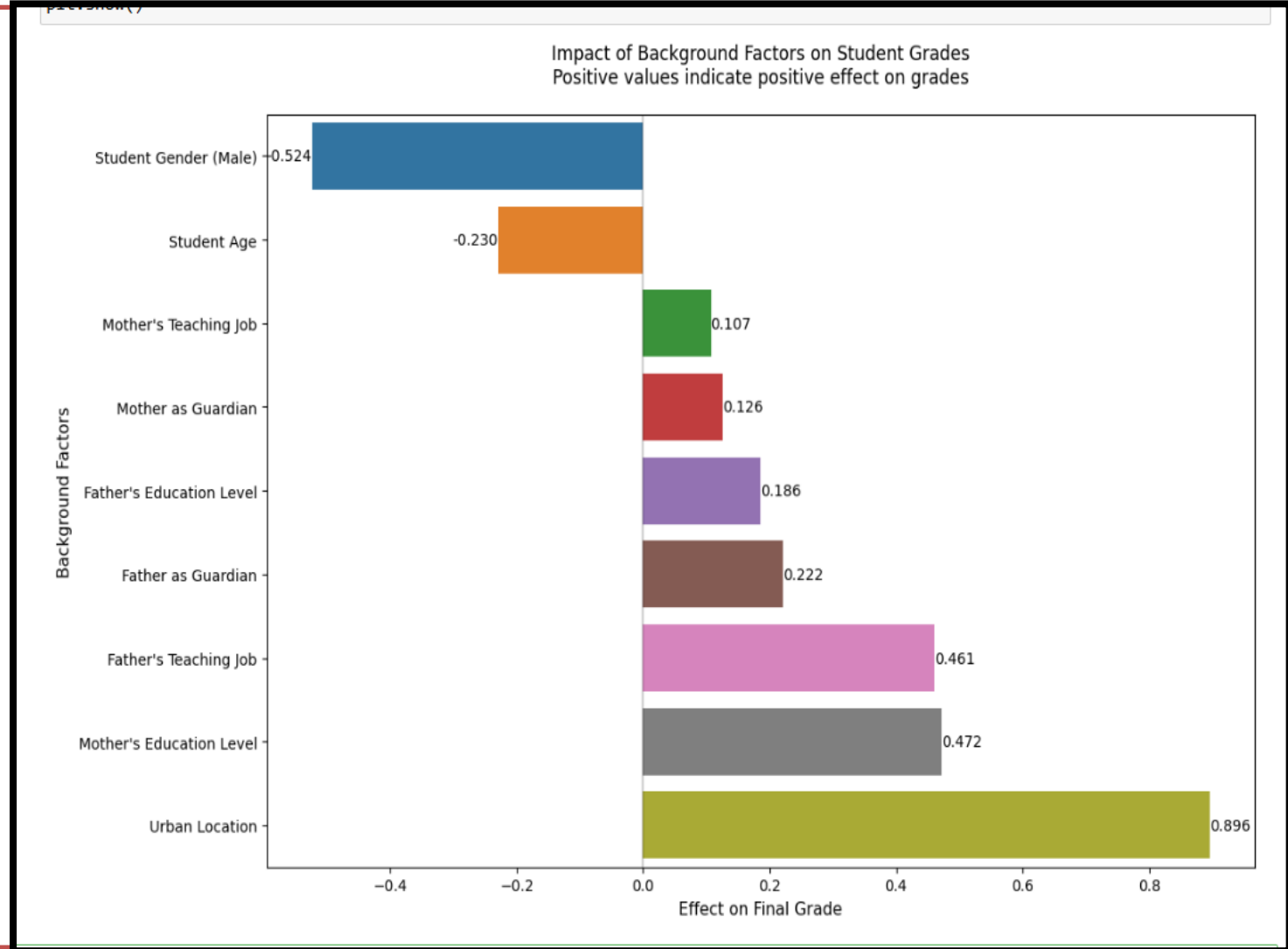
These box plots compare the distribution of grades across different parental education levels (1-4 scale):

1.Mother's Education Plot (Left):

1. Clear upward trend in median grades as mother's education increases
2. Wider grade spread at higher education levels
3. Some outliers at lower grades across all levels
4. Most consistent improvement between levels 3 and 4

2.Father's Education Plot (Right):

1. Similar but more gradual increase in median grades
2. More consistent grade spread across all levels
3. Slightly less pronounced effect compared to mother's education
4. More outliers in the lower grade range

Both plots demonstrate that higher parental education generally correlates with better student performance, with mother's education showing a slightly stronger pattern.
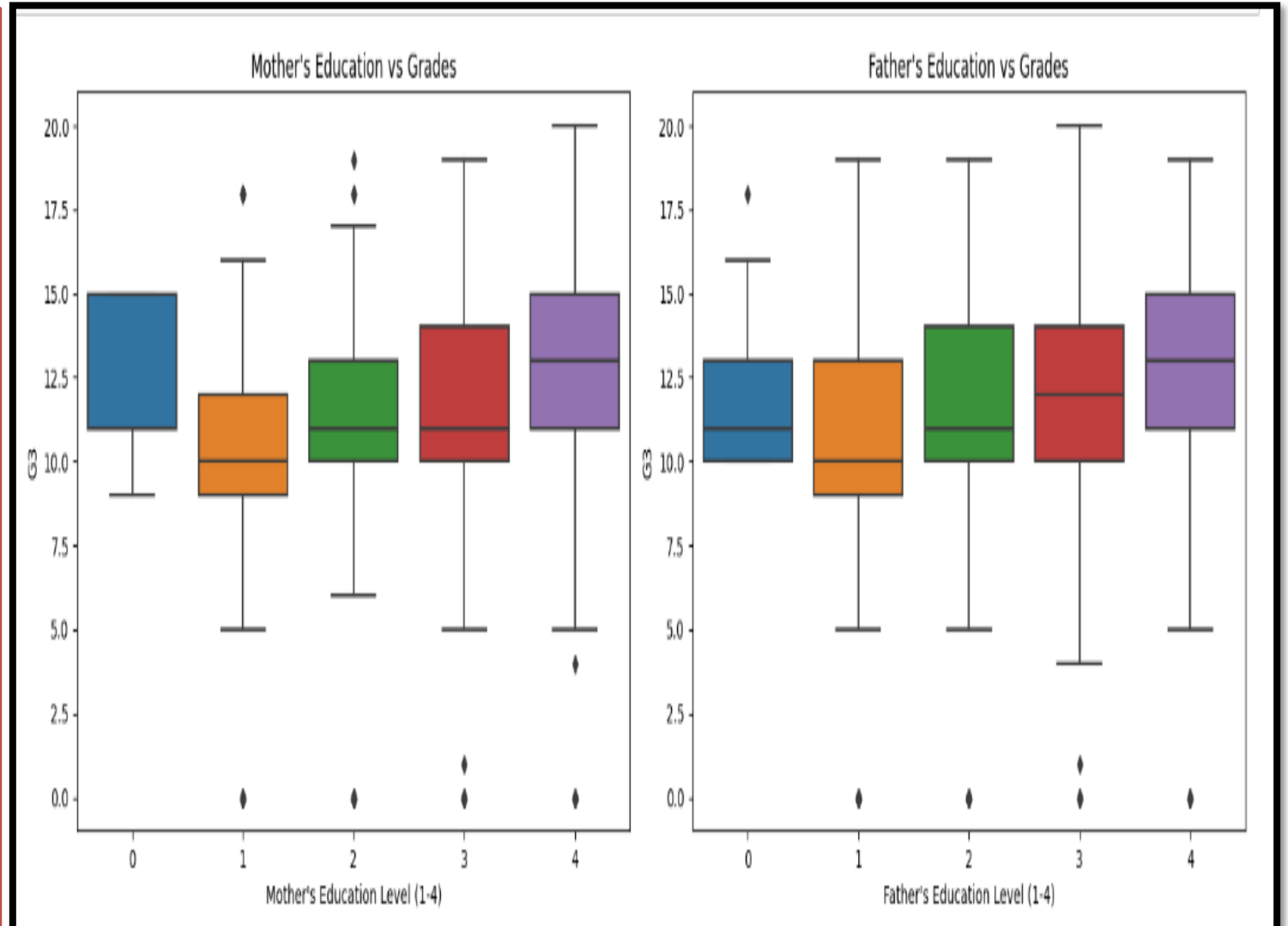
Relationship Between Background Factors and Grades

Age Impact:
- •Negative trend: Younger students generally achieve better grades
- •Widest grade variation in 15-18 age group
- •Performance tends to decline and become more scattered in older students (19-22)

Parents' Education Effect:
- •Both show positive influence on grades
- •Mother's education: Stronger positive impact (steeper trend)
- •Father's education: Similar but more moderate effect
- •Most grades cluster between 7.5-15 for all education levels

Summary

These trends confirm our earlier findings: younger age and higher parental education correlate with better academic performance, with mother's education showing the strongest positive influence.
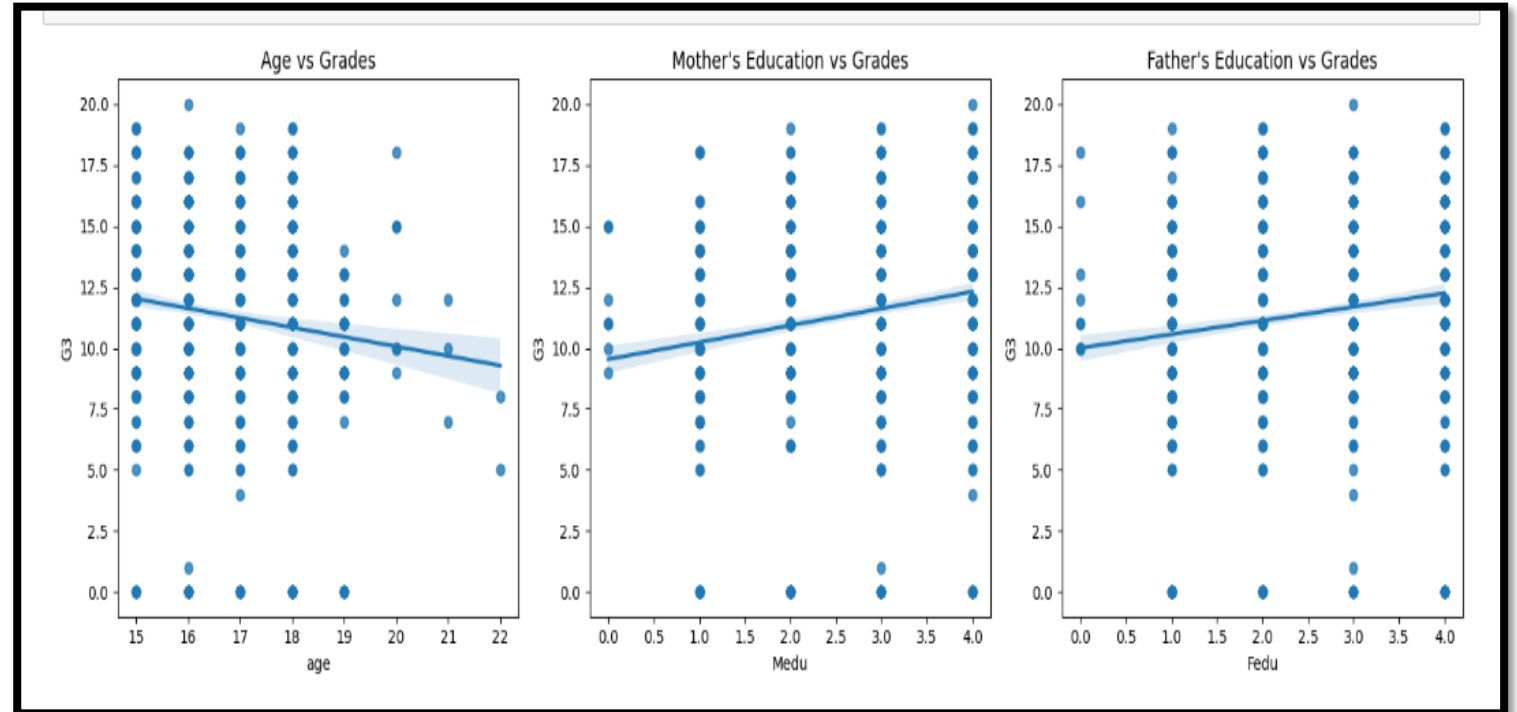
Impact of Academic and Social Factors on Grades

1. Attendance Rate (1.594): By far the strongest positive factor
  1. Demonstrates crucial importance of consistent class participation

2. Academic Effort
  1. Preparation Time (0.125): Second strongest positive impact
  2. HSC Score (0.093): Previous academic performance matters
  3. English Score (0.085): Language proficiency helps
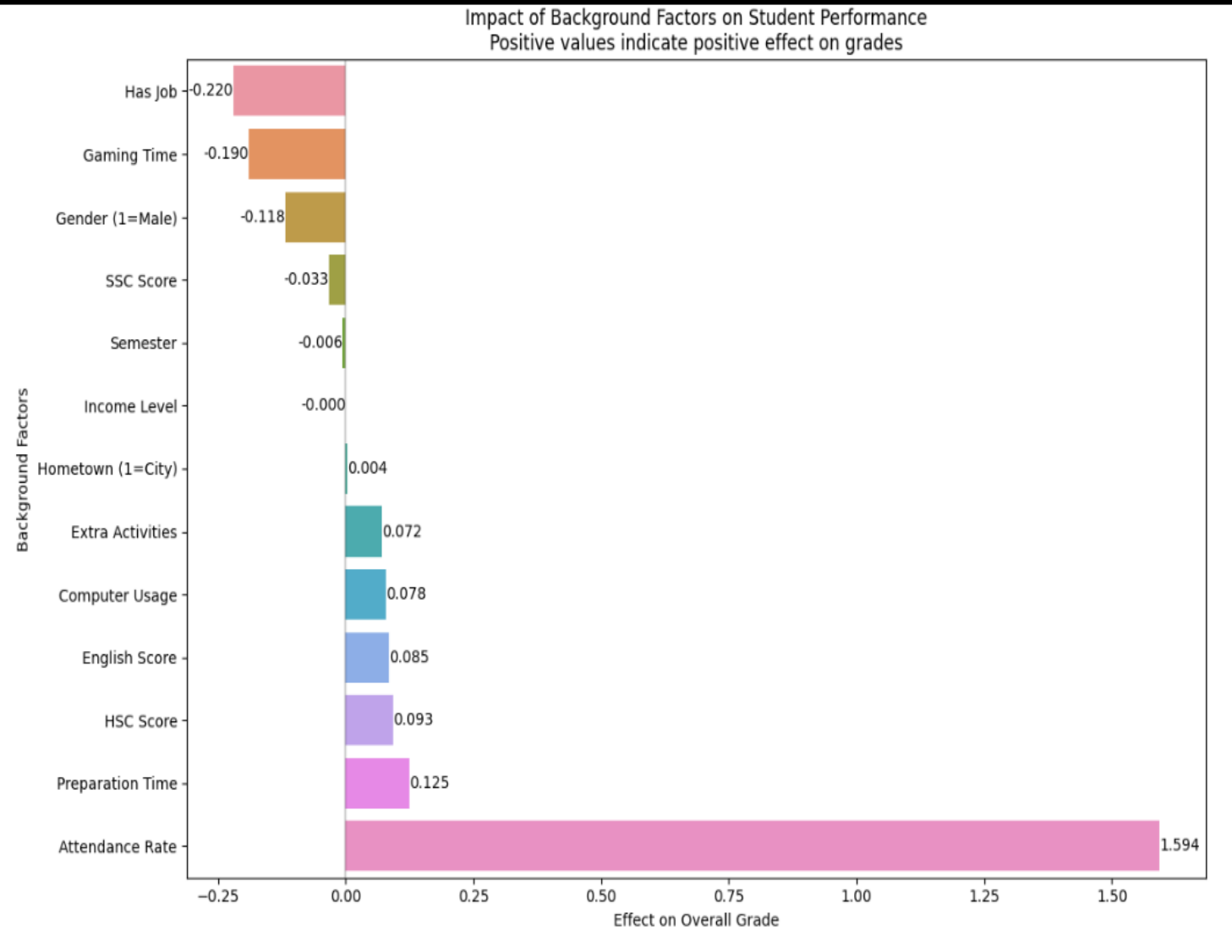
3. External Commitments
  1. Having a Job (-0.220): Strongest negative impact
  2. Gaming Time (-0.190): Second most negative factor

4. Gender Effect
  1. Being Male (-0.118): Shows moderate negative correlation

Summary:

Results suggest academic success is most strongly influenced by attendance and time management, while external commitments (work, gaming) can significantly hinder performance.

Impact of Background Factors on Student Performance
Positive values indicate positive effect on grades

Background Factor	Effect on Overall Grade
Has Job	-0.220
Gaming Time	-0.190
Gender (1=Male)	-0.118
SSC Score	-0.033
Semester	-0.006
Income Level	-0.000
Hometown (1=City)	0.004
Extra Activities	0.072
Computer Usage	0.078
English Score	0.085
HSC Score	0.093
Preparation Time	0.125
Attendance Rate	1.594

Gender and Income Effects on Academic Performance

Gender Impact

- Female students show slightly higher median grades
- Male students show more consistent but lower performance
- Confirms regression coefficient: -0.118 (negative effect for males)

Income Level Impact

- Minimal differences across income levels (0-3)
- All income levels show similar grade distributions
- Supports regression finding: negligible effect (-0.000)

Comparison with First Dataset

- Gender effect consistent but stronger in first dataset (-0.524 vs -0.118)
- Economic impact varies: parental education matters more than income level
- Both datasets show female students generally performing better

Summary:

Gender consistently influences performance across both contexts, while economic factors show varying impacts depending on measurement method.

Grades by Gender (0=Female, 1=Male)

Grades by Income Level
(0=Low, 1=Lower middle, 2=Upper middle, 3=High)
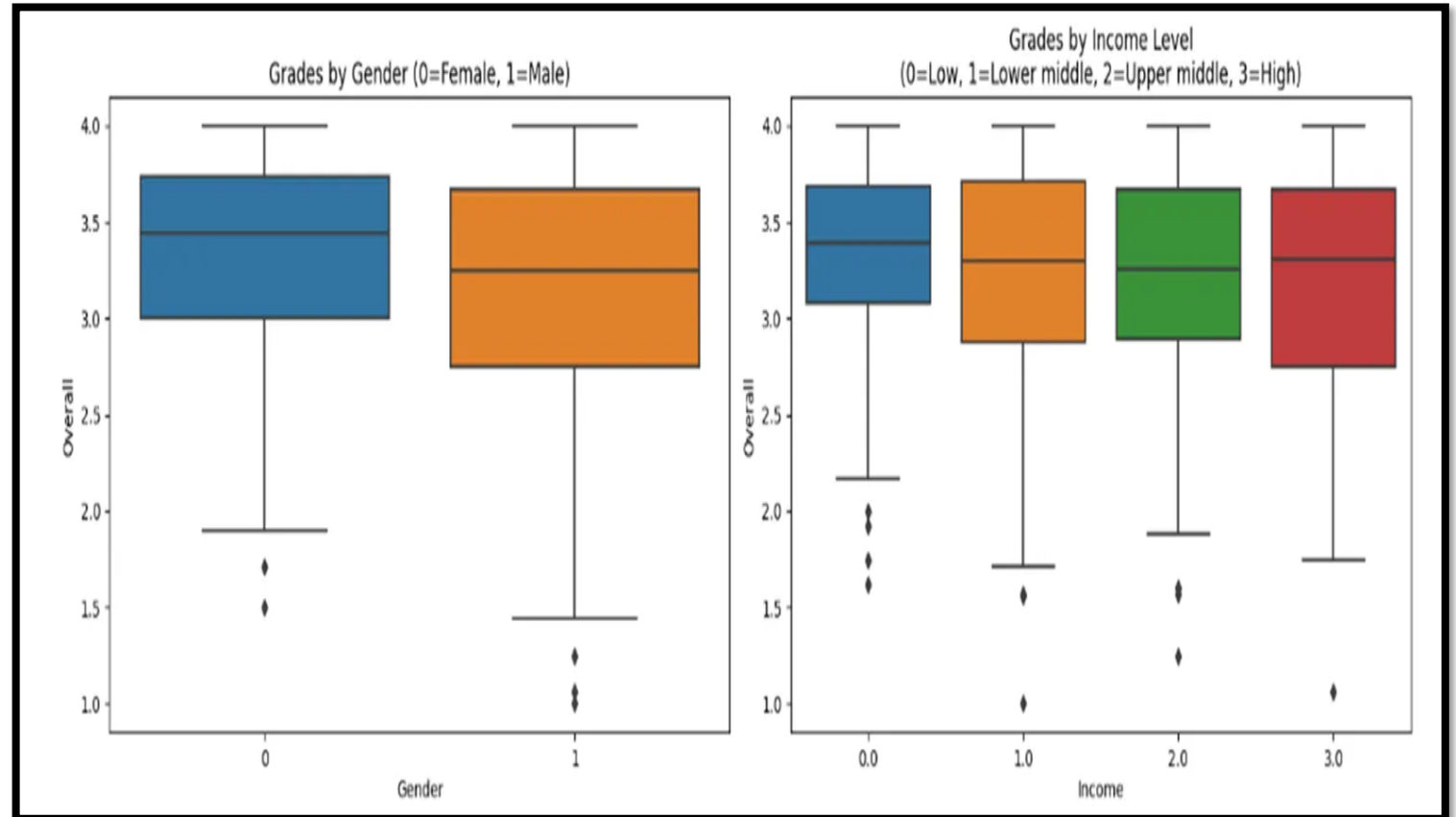
Cross-Dataset Comparison of Background Factors

This visualization compares the impact of three key background factors (Gender, Location, and Family Background) across both datasets, showing how their effects on grades vary in different educational contexts.

1. Location Impact

1. First dataset shows strong urban advantage **(0.896)**
2. Second dataset shows minimal effect (0.004)
3. Suggests location importance varies by educational system/level

2. Gender Effect

1. Consistently negative for male students
2. Stronger effect in first dataset (-0.523 vs -0.118)
3. Shows persistent gender gap across different contexts

3. Family Background

1. First dataset (parent education): significant positive effect (0.472)
2. Second dataset (income): negligible impact **(-0.000)**
3. Indicates educational background matters more than economic status

Summary:

The comparison reveals that background factors can have drastically different impacts depending on the educational context and how they are measured.

Comparison of Background Factors Impact Across Both Datasets
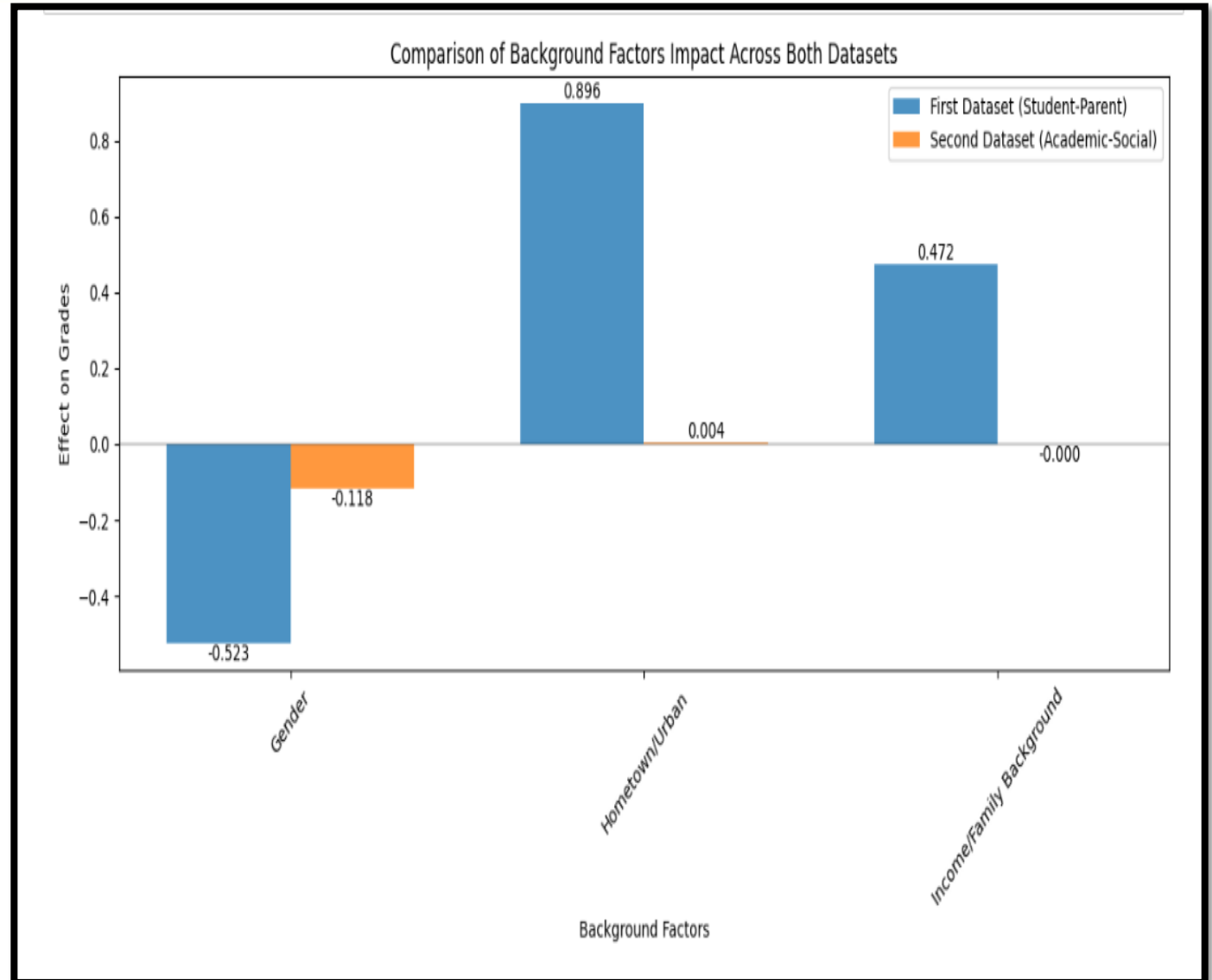
1. **How do family background factors affect student performance?**
   1. Parents' education shows significant positive impact (0.472)
   2. Urban location advantage is substantial in traditional education (0.896)
   3. Income level alone has minimal direct impact (-0.000)

2. **What is the relative importance of social vs academic factors?**
   1. Academic behaviors dominate:
      1. Attendance is the strongest positive factor (1.594)
      2. Preparation time shows positive impact (0.125)
   2. Social factors can negatively impact:
      1. Having a job (-0.220)
      2. Gaming time (-0.190)

3. **Are there consistent patterns across educational contexts?**
   1. Gender effect is consistent: female students perform better in both datasets
   2. Family influence varies: educational background matters more than income
   3. Location impact differs significantly between contexts

- **Key Takeaway:** While some background factors (like gender) show consistent effects, the impact of others varies by context. Academic behaviors and parental education emerge as stronger predictors of success than economic factors alone.

Thank you!