Study Time vs Final Grades Analysis Report

Introduction

This analysis investigates the relationship between study time and academic performance, represented by final grades (G3). The hypothesis is as follows:

Null Hypothesis: Study hours have no effect on academic performance.

Alternative Hypothesis: Study hours positively affect academic performance.

Methodology

Dataset: The "Student Performance" dataset from the UCI Machine Learning Repository, containing 395 records and 33 variables.

Variables of Interest:

studytime: Weekly study time (1 to 4, representing categories).

G3: Final grades (numeric, range 0–20).

Tools: Python with pandas, matplotlib, seaborn, and statsmodels libraries.

The analysis involves:

1. Exploratory Data Analysis (EDA)
2. Outlier detection and data cleaning
3. Correlation analysis
4. Linear regression modeling

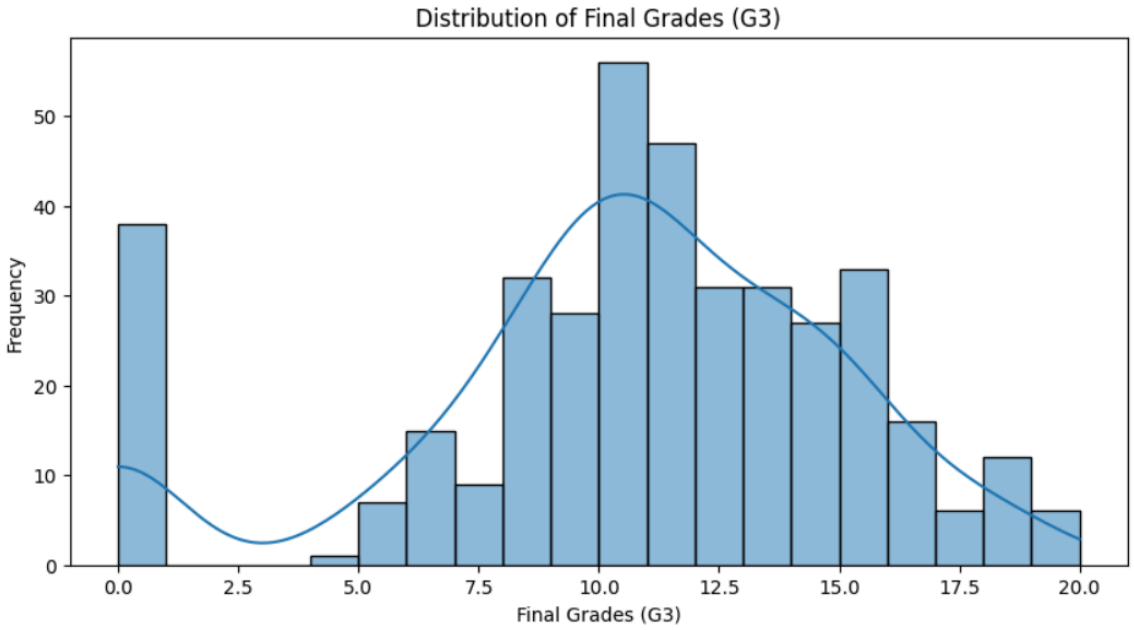
Results

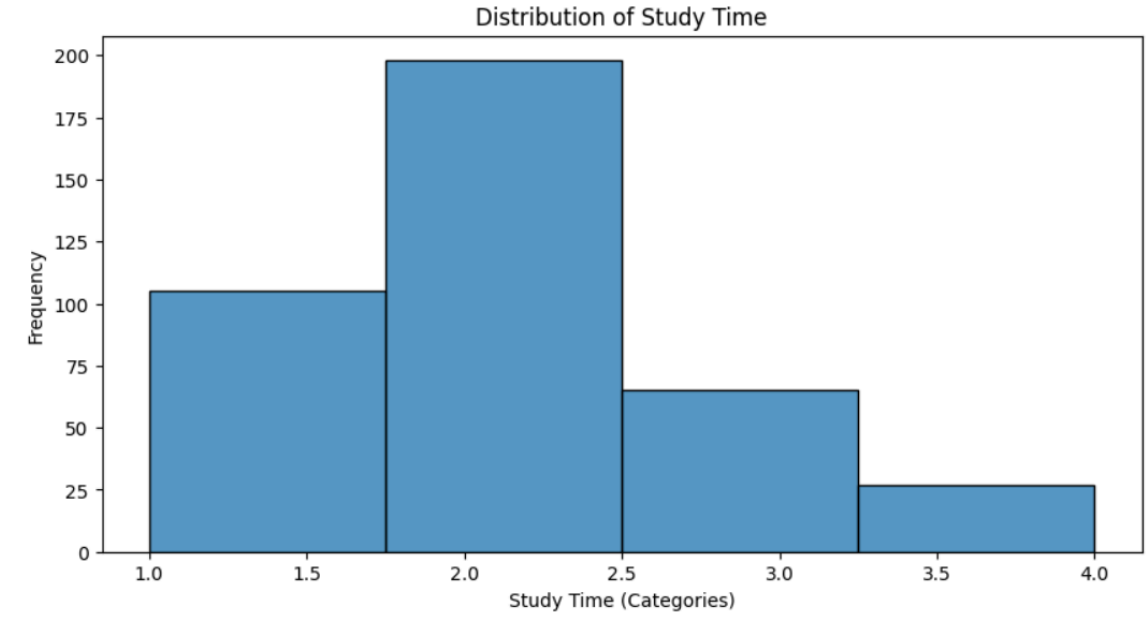
Data Exploration

1. Study Time:
    a. Mean: 2.03 (mostly in categories 1 and 2).
    b. Range: 1–4.
2. Final Grades (G3):
    a. Mean: 10.42.
    b. Range: 0–20.

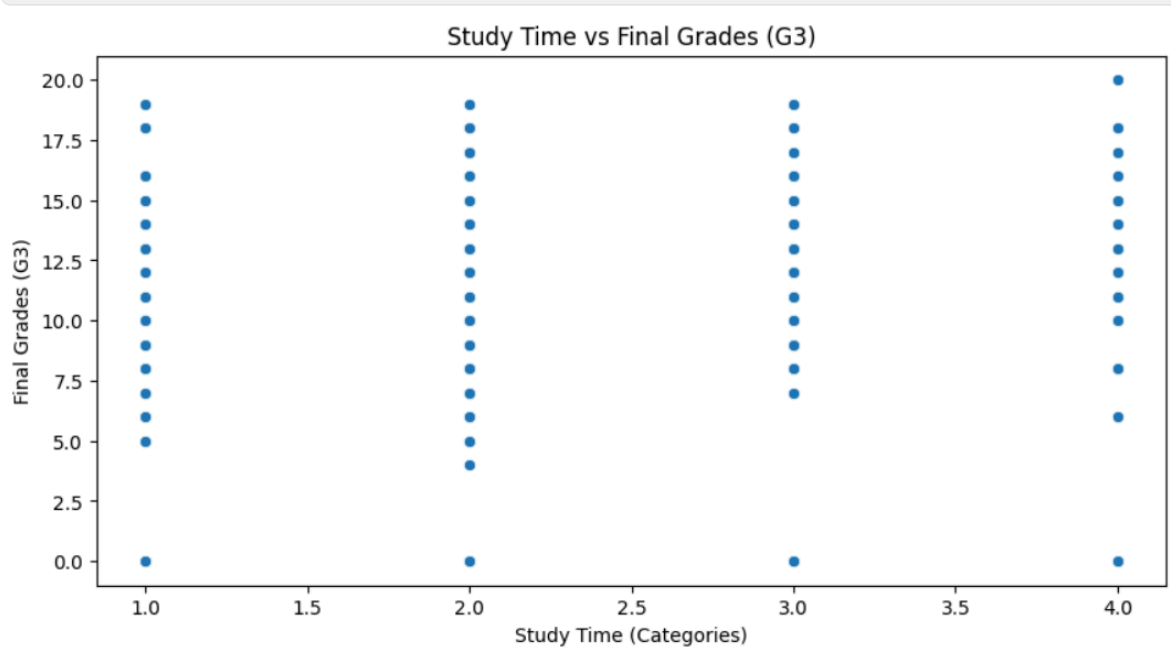
Visualizations

1. **Histograms**:
    - Study time distribution shows most students study less (categories 1 and 2).
    - Grades distribution is slightly right-skewed, peaking around 8–14.





2. **Scatter Plot**:
    - Displays a weak positive trend between study time and grades.



Data Cleaning

1. Outliers in grades (G3) were assessed using the IQR method.
2. No significant outliers were detected, so the dataset remained unchanged.

Correlation Analysis

The Pearson correlation coefficient between study time and G3 is 0.098, indicating a weak positive relationship.

Linear Regression Analysis

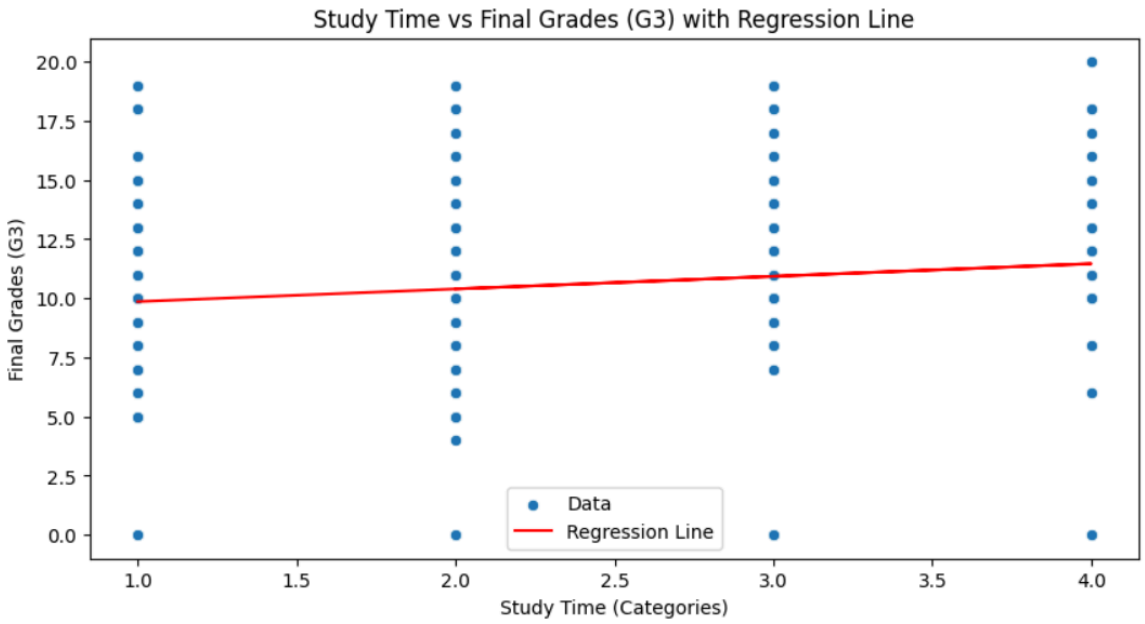
The model equation:

Key metrics:

1. Intercept : 9.33 — Predicted grade for zero study time.
2. Slope : 0.534 — Each additional study category increases grades by ~0.534 points.
3. R-squared: 0.010 — Only 1% of the variance in grades is explained by study time.
4. P-value : 0.052 — Borderline statistical significance.

Regression Visualization

The regression line overlaid on the scatter plot confirmed a weak upward trend.



Interpretation

1. The weak correlation and low value suggest that study time alone is not a strong predictor of final grades.
2. The borderline p-value (0.052) indicates that the effect of study time is statistically inconclusive.

Conclusions

1. The hypothesis that study time positively affects grades is weakly supported but not conclusive.
2. Academic performance likely depends on multiple factors beyond study time.
3. Therefore we reject alternative hypothesis

Recommendations

1. Include additional variables (e.g., sleep patterns, class participation, extracurricular activities) for a more comprehensive analysis.
2. Use a larger dataset for improved robustness.
3. Explore nonlinear models or interactions between variables.