

Automated ML Analysis Report

Dynamic Machine Learning Report

Automated ML Analysis Report

About the Dataset

The dataset contains 1000 observations across 8 columns. Among them, 3 are numerical and 5 are categorical variables. A robust understanding of the data structure aids in selecting appropriate analysis and modeling techniques.

Exploratory Data Analysis

The feature 'math score' ranges from 0.00 to 100.00. The mean value is 66.09 with a standard deviation of 15.16, indicating moderate variability. Visual analysis confirms that values cluster within this range, with few potential outliers influencing distribution tails.

Boxplots and violin plots of math score reveal the central tendency around the median and help detect asymmetry and extreme values that may affect model assumptions.

The feature 'reading score' ranges from 17.00 to 100.00. The mean value is 69.17 with a standard deviation of 14.60, indicating moderate variability. Visual analysis confirms that values cluster within this range, with few potential outliers influencing distribution tails.

Boxplots and violin plots of reading score reveal the central tendency around the median and help detect asymmetry and extreme values that may affect model assumptions.

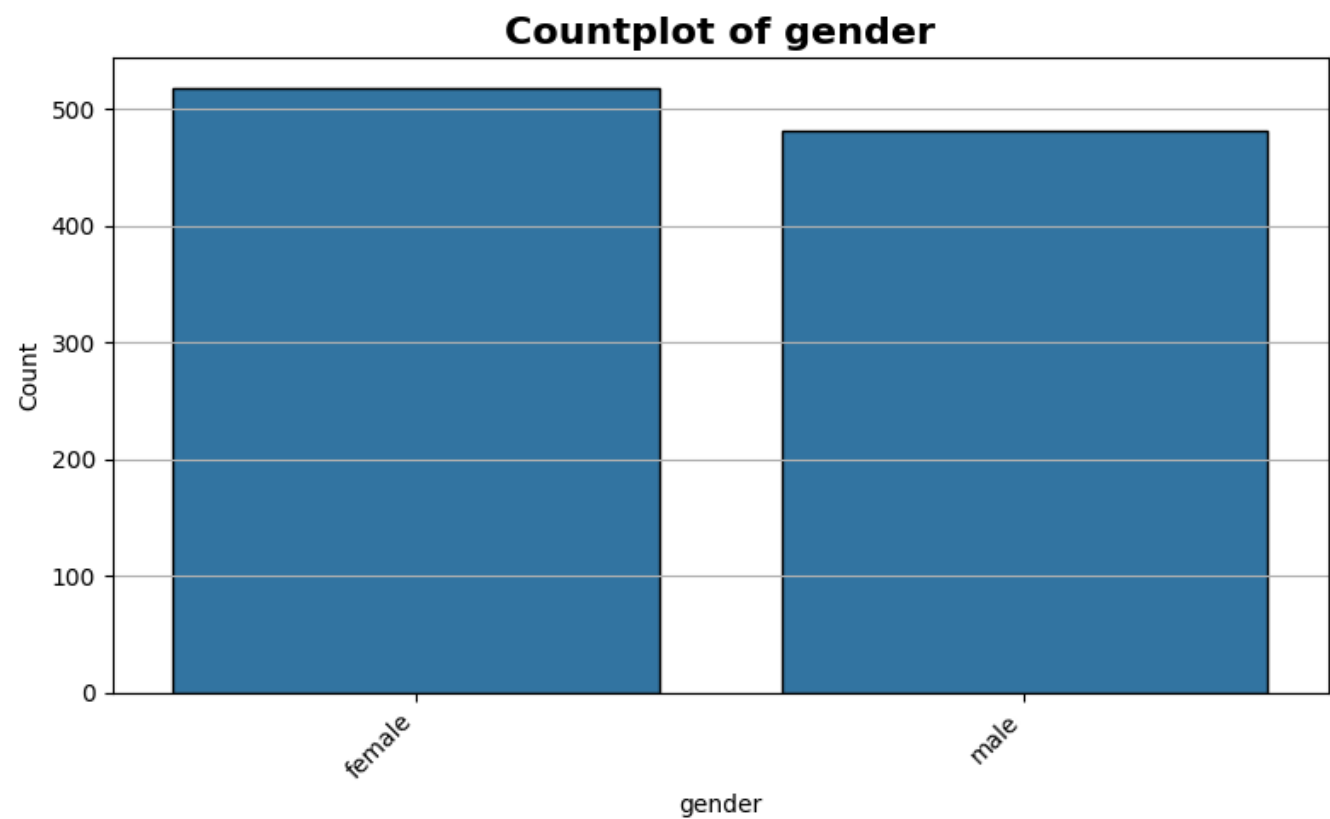
The feature 'writing score' ranges from 10.00 to 100.00. The mean value is 68.05 with a standard deviation of 15.20, indicating moderate variability. Visual analysis confirms that values cluster within this range, with few potential outliers influencing distribution tails.

Boxplots and violin plots of writing score reveal the central tendency around the median and help detect asymmetry and extreme values that may affect model assumptions.

The categorical feature 'gender' shows the frequency of each category. Understanding dominant or

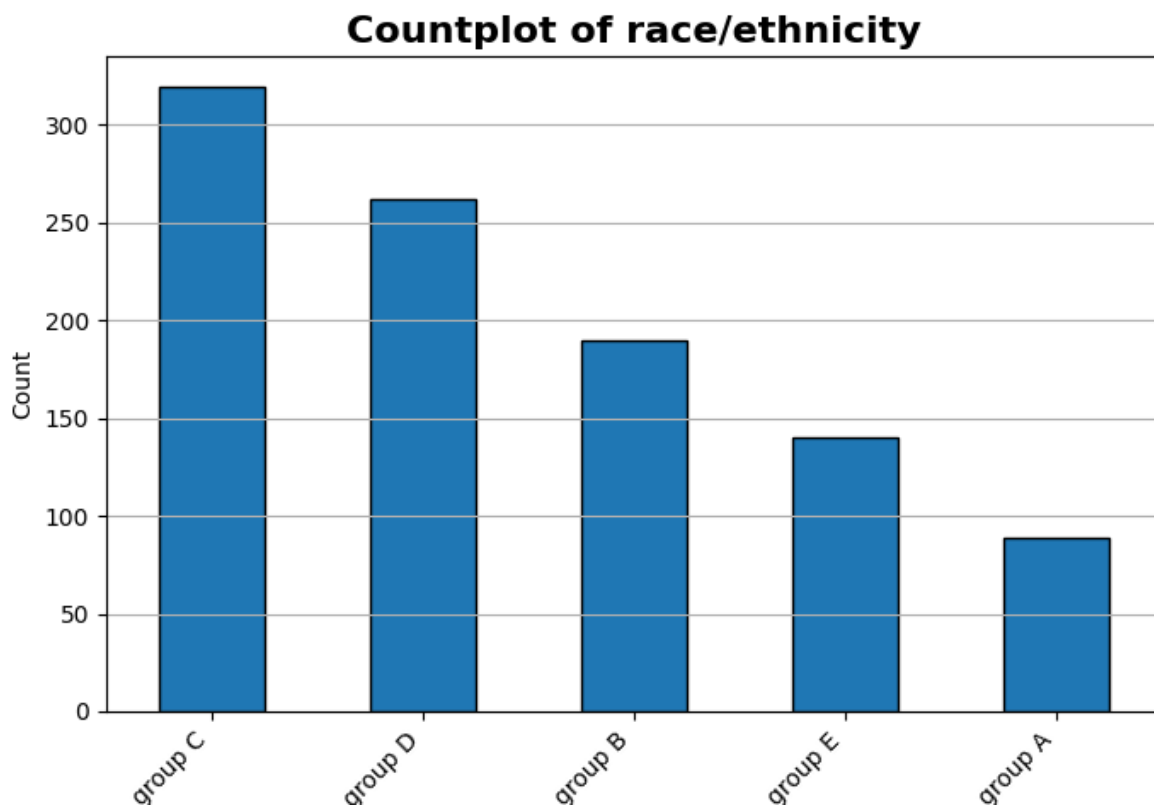
Automated ML Analysis Report

underrepresented classes is important for ensuring model fairness and effectiveness.



The categorical feature 'race/ethnicity' shows the frequency of each category. Understanding dominant or underrepresented classes is important for ensuring model fairness and effectiveness.

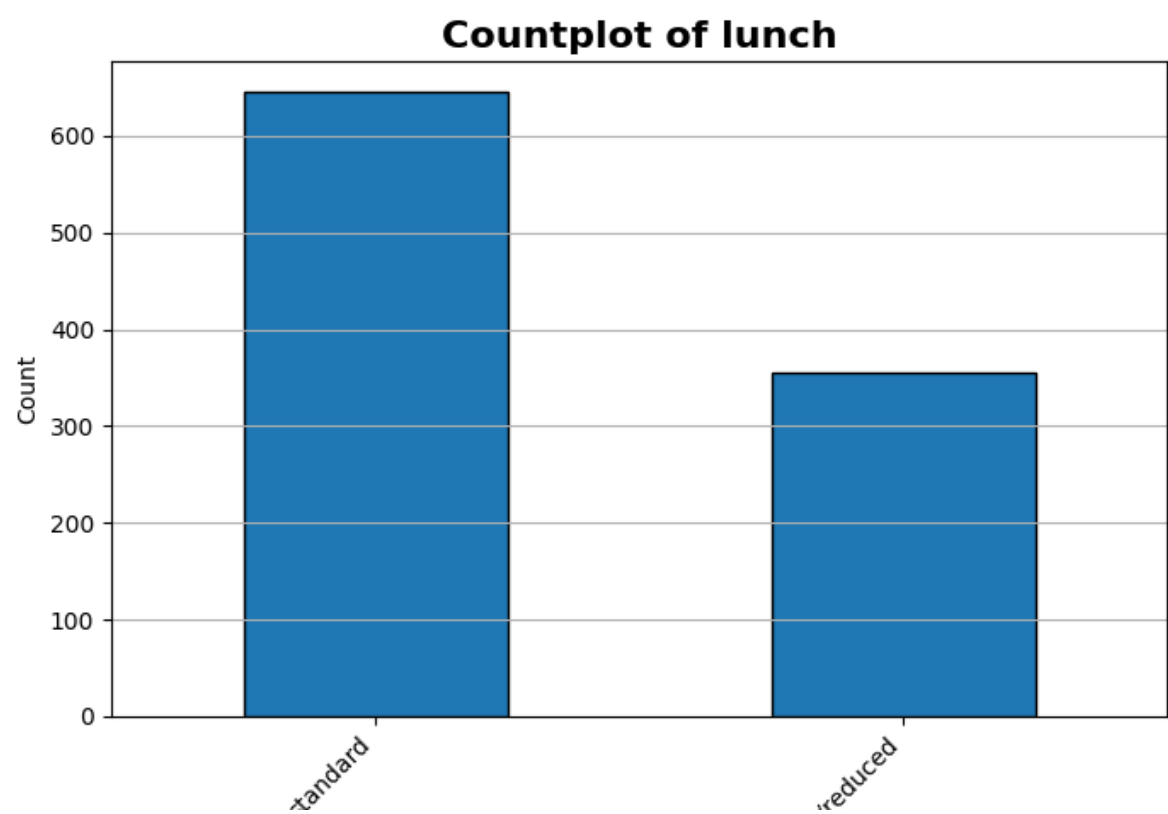
Automated ML Analysis Report



The categorical feature 'parental level of education' shows the frequency of each category. Understanding dominant or underrepresented classes is important for ensuring model fairness and effectiveness.

The categorical feature 'lunch' shows the frequency of each category. Understanding dominant or underrepresented classes is important for ensuring model fairness and effectiveness.

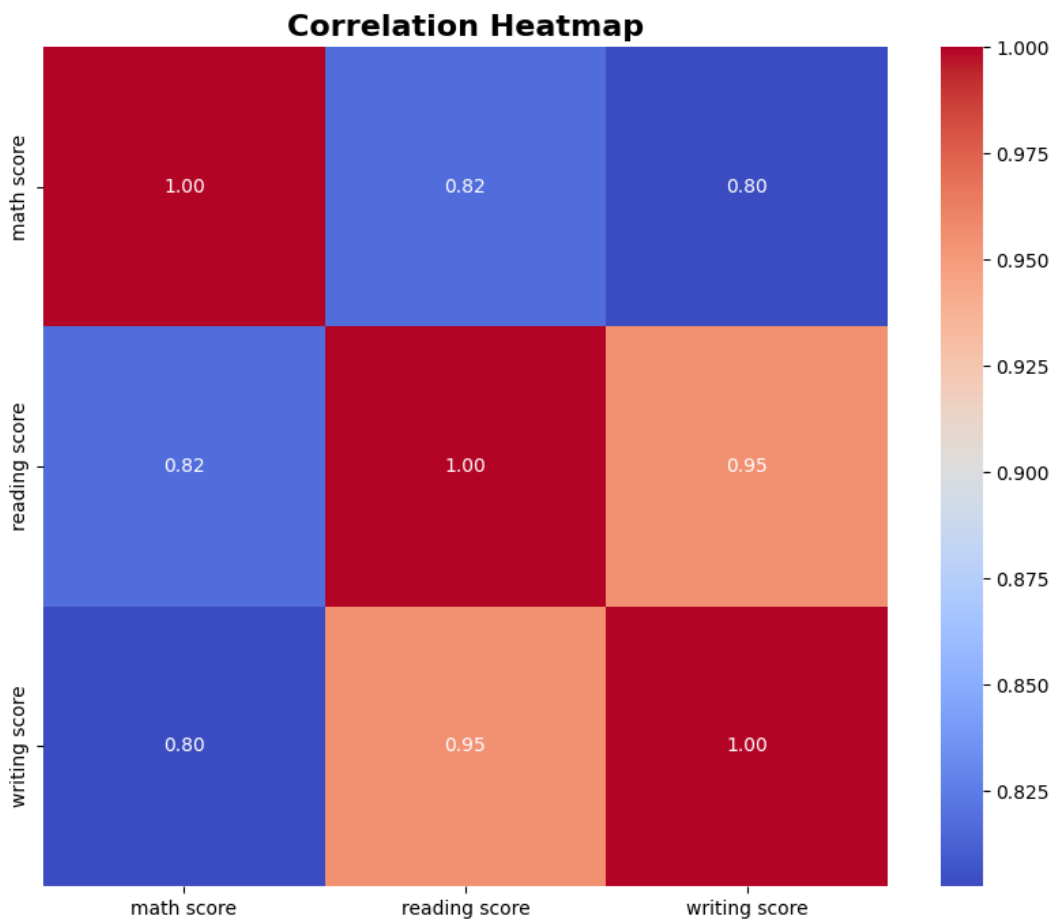
Automated ML Analysis Report



The categorical feature 'test preparation course' shows the frequency of each category. Understanding dominant or underrepresented classes is important for ensuring model fairness and effectiveness.

The correlation heatmap visualizes relationships among numerical features. High correlations may indicate multicollinearity, which can bias model coefficients and reduce performance. Addressing these during preprocessing improves model robustness.

Automated ML Analysis Report



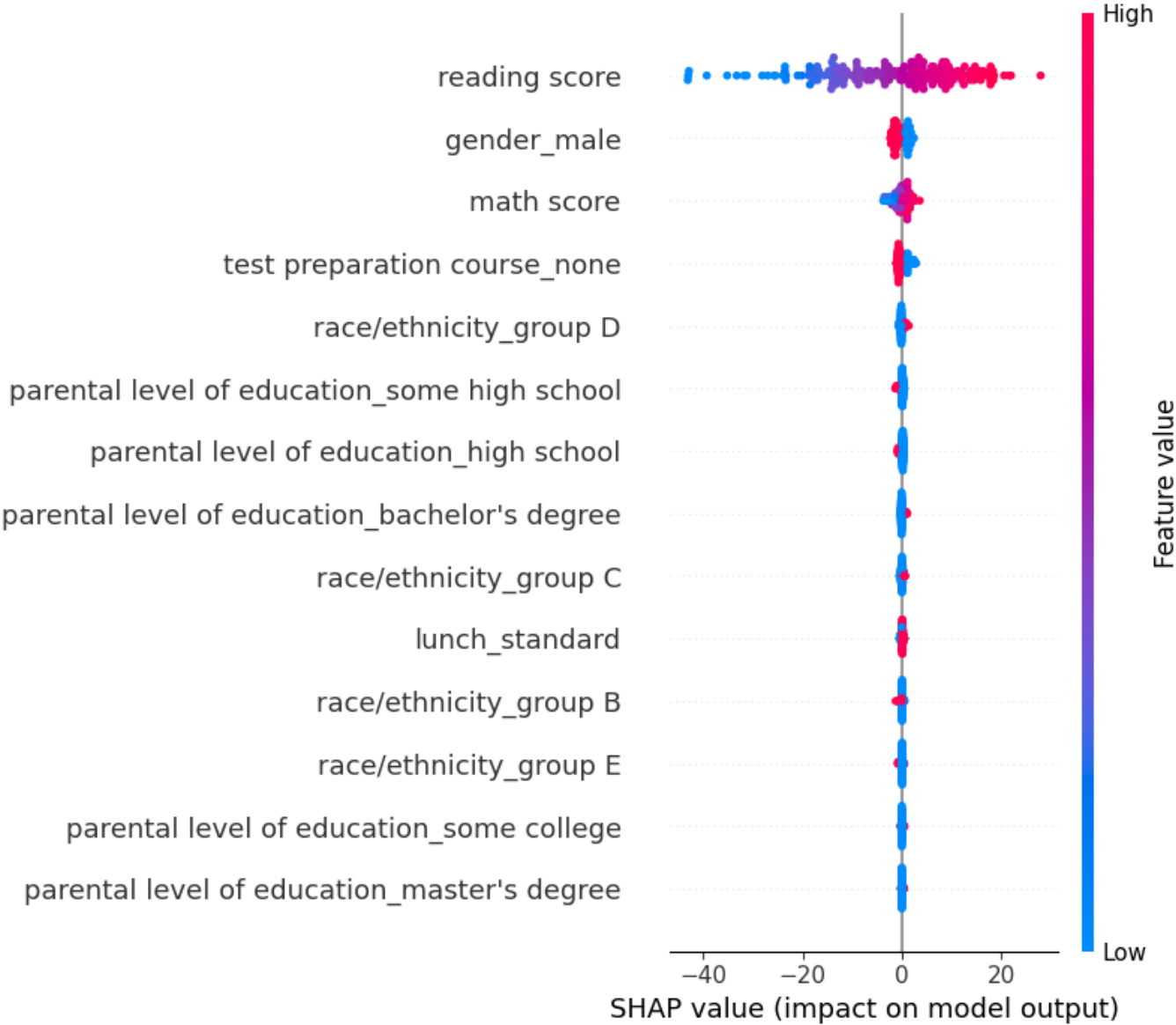
Model Building and Evaluation

The task was identified as Regression. A Random Forest model was used for its ability to capture non-linear relationships effectively without intensive preprocessing. The model achieved a performance score of 0.9168. An RMSE of 4.4793 indicates acceptable prediction accuracy for this domain.

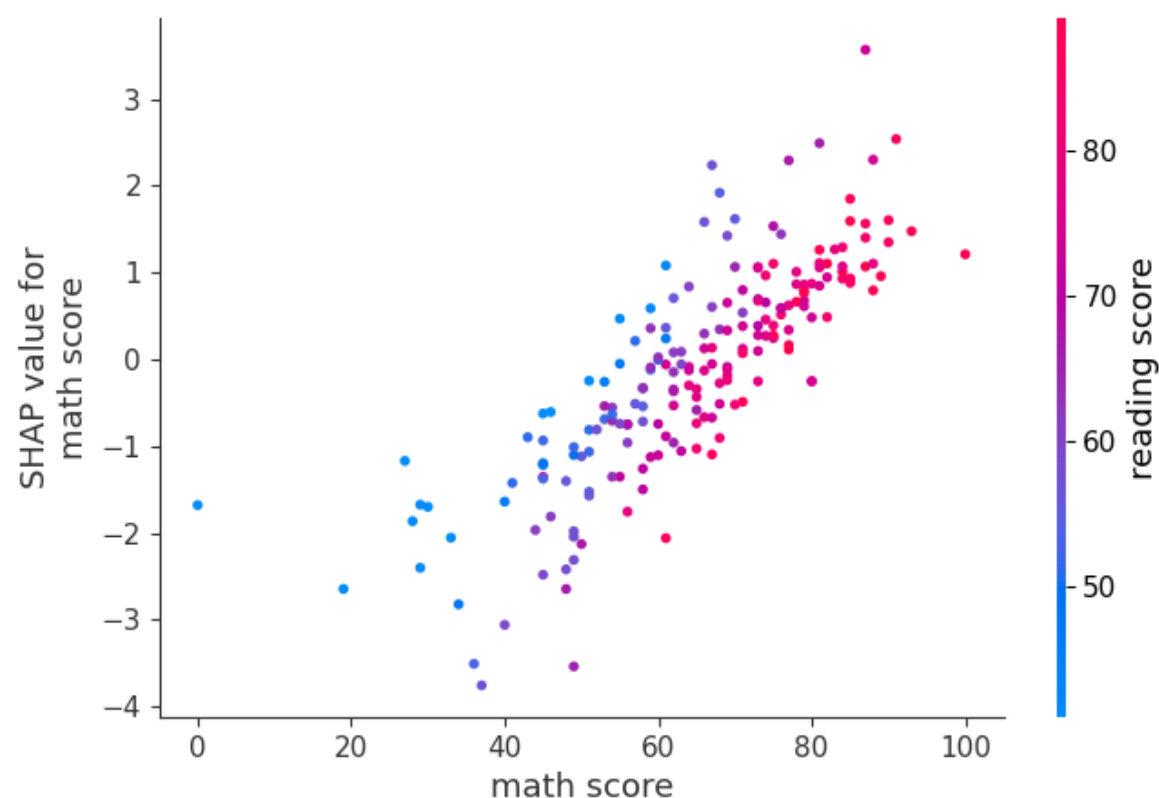
Explainable AI (XAI) Results

SHAP analysis identifies major feature influences globally, while dependence plots highlight feature interactions. LIME provides local, instance-based explanations enhancing model transparency.

Automated ML Analysis Report



Automated ML Analysis Report



Conclusion and Future Recommendations

This pipeline successfully automated dataset analysis, model building, evaluation, and interpretability. Future work can focus on model tuning, stacking methods like Gradient Boosting, and adding fairness and accountability checks for responsible AI deployment.