

Navigation

Go to

- ☒ Data Overview
- ☐ Model Evaluation
- ☐ Clustering Analysis
- ☐ Anomaly Detection
- ☐ Predictive Insights

Student Performance Model Training and Analysis

Data loaded successfully.

Data Preprocessing

Initial Data Shape

(17375, 42)

Missing Values in Each Column

	0
District	17,375
Gender	17,375
Ethnicity	17,375
AssessmentSubjectcode	17,375
SE	17,375
ELL	17,242
EconomicallyDisadvantaged	17,375
Achievement	17,375
AchievementCode	2,960
AssessmentMeasureDesc	17,375

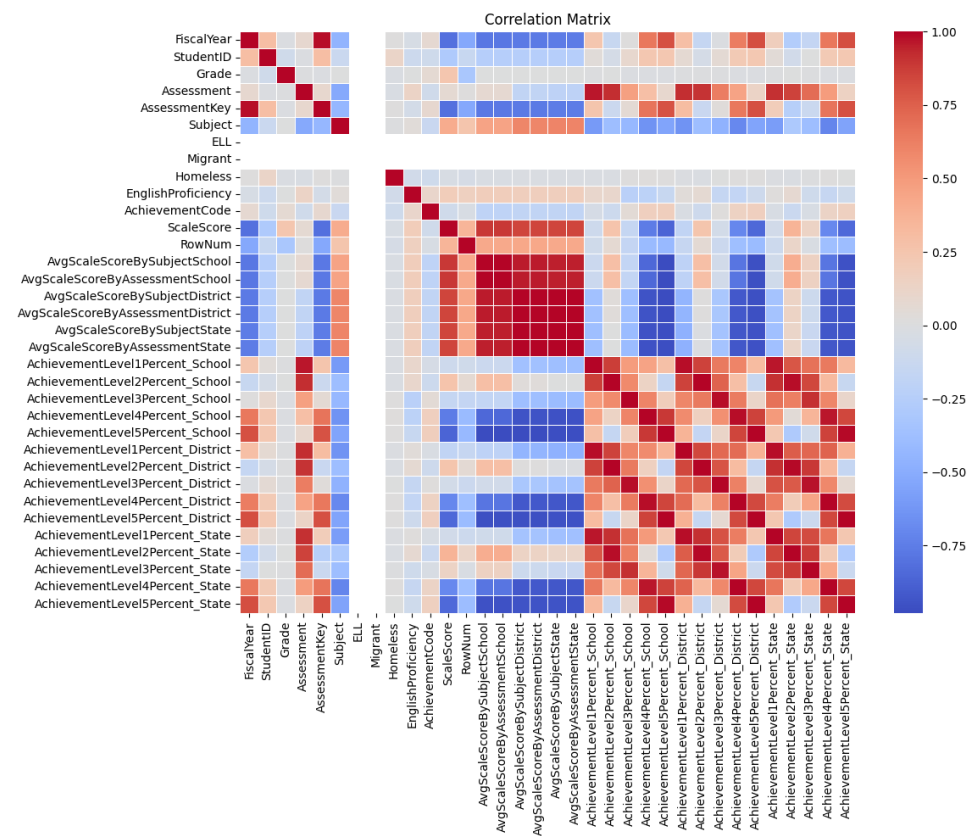
Dropping Columns with All Missing Values: ['District', 'Gender', 'Ethnicity', 'AssessmentSubjectcode', 'SE', 'EconomicallyDisadvantaged', 'Achievement', 'AssessmentMeasureDesc']

Data Shape after Imputing Missing Values: (17375, 34)

Data Shape after Dropping Remaining NaNs: (17375, 34)

Correlation Analysis

Correlation Matrix



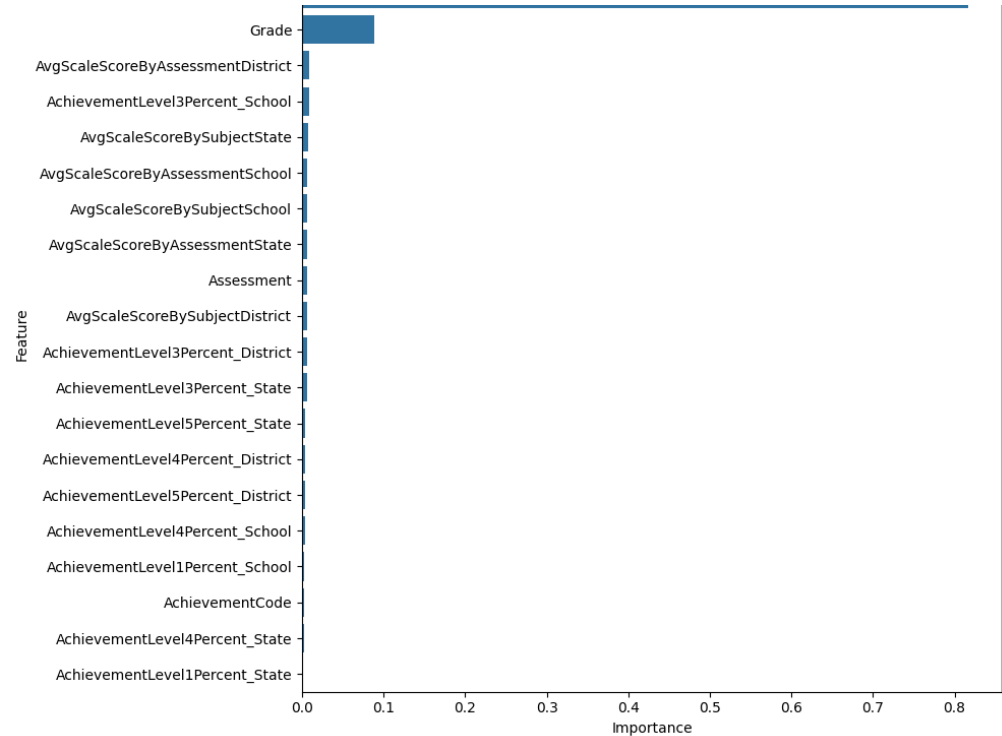
Top 10 Features Correlated with ScaleScore

	ScaleScore
ScaleScore	1
AvgScaleScoreByAssessmentSchool	0.8876
AvgScaleScoreBySubjectSchool	0.8876
AchievementLevel5Percent_School	0.8685
AvgScaleScoreByAssessmentDistrict	0.8532
AvgScaleScoreBySubjectDistrict	0.8532
AchievementLevel5Percent_District	0.8521
AvgScaleScoreByAssessmentState	0.8462
AvgScaleScoreBySubjectState	0.8462
AchievementLevel5Percent_State	0.8428

Feature Importance

Feature Importance Plot





Feature Importance Data

	Feature	Importance
4	AssessmentKey	0.8171
2	Grade	0.0881
15	AvgScaleScoreByAssessmentDistrict	0.0085
20	AchievementLevel3Percent_School	0.0083
16	AvgScaleScoreBySubjectState	0.0071
13	AvgScaleScoreByAssessmentSchool	0.0066
12	AvgScaleScoreBySubjectSchool	0.0064
17	AvgScaleScoreByAssessmentState	0.0063
3	Assessment	0.0062
14	AvgScaleScoreBySubjectDistrict	0.006

Model Evaluation

Random Forest Regressor

Random Forest RMSE: 8.15

Random Forest R² Score: 1.00

Linear Regression

Linear Regression RMSE: 108.95

Linear Regression R² Score: 0.90

Gradient Boosting Regressor

Gradient Boosting RMSE: 8.98

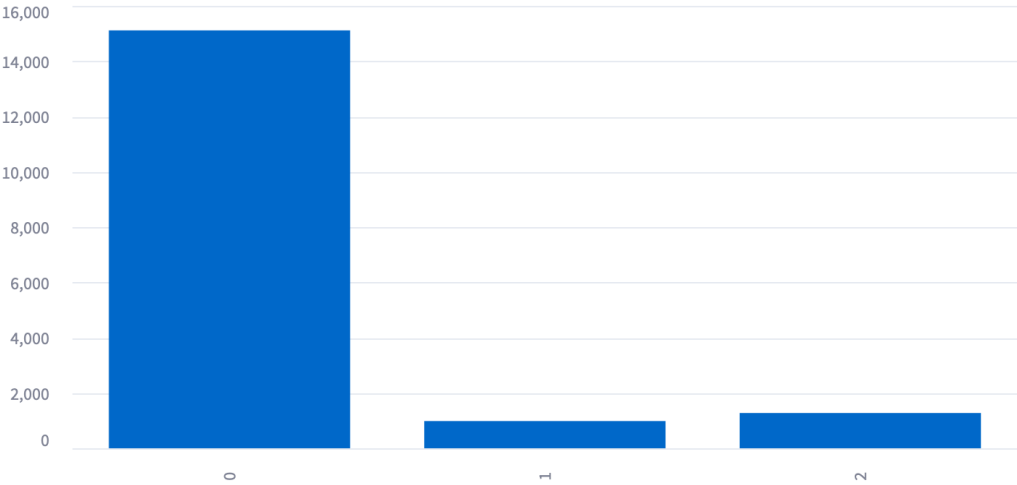
Gradient Boosting R² Score: 1.00

Model Evaluation Summary

	Model	RMSE	R ² Score
0	Random Forest	8.1506	0.9994
1	Linear Regression	108.9457	0.8995
2	Gradient Boosting	8.9776	0.9993

Clustering Analysis

Cluster Counts



Cluster Summaries

Cluster 0 Summary

	FiscalYear	StudentID	Grade	Assessment	AssessmentKey	Subject	ELL	Migrant	Homeless
count	15,114	15,114	15,114	15,114	15,114	15,114	15,114	15,114	15,114
mean	3.2344	130,083,955.4225	6.3705	3.8409	1,817.1153	3.2355	1	0	0.0001
std	2.4289	86,594.5277	1.2186	2.1508	665.9072	1.6209	0	0	0.0001
min	0	128,511,253	5	0	723	0	1	0	
max	5	130,083,955	7	5	5,015	5	1	1	

25%	1	130,042,687	5	1	1,247	2	1	0	
50%	3	130,086,870	6	4	1,737	4	1	0	
75%	6	130,130,099	8	6	2,567	5	1	0	
max	7	130,181,280	8	6	2,900	5	1	0	

Cluster 2 Summary

	FiscalYear	StudentID	Grade	Assessment	AssessmentKey	Subject	ELL	Migrant	Homeless
count	1,276	1,276	1,276	1,276	1,276	1,276	1,276	1,276	1
mean	0.2476	126,834,417.8738	7.8793	2.9483	885.3119	4.203	1	0	0.0008
std	0.5039	1,382,872.6309	0.426	2.448	212.9276	1.163	0	0	0.111
min	0	123,063,508	5	0	723	2	1	0	
25%	0	127,068,563	8	1	752	4	1	0	
50%	0	127,335,845	8	1	825	5	1	0	
75%	0	127,426,411	8	6	827	5	1	0	
max	3	128,336,799	8	6	1,796	5	1	0	

Cluster 1 Summary

	FiscalYear	StudentID	Grade	Assessment	AssessmentKey	Subject	ELL	Migrant	Homeless
count	985	985	985	985	985	985	985	985	
mean	5.6193	150,041,331.2345	6.3188	3.4985	2,459.4914	2.6173	1	0	0.1118
std	1.5269	22,366.7227	1.1876	1.9096	349.1132	1.7393	0	0	0.333
min	2	150,000,903	5	0	1,472	0	1	0	
25%	5	150,026,605	5	1	2,552	1	1	0	
50%	6	150,040,647	6	4	2,570	3	1	0	
75%	7	150,058,609	7	5	2,639	4	1	0	
max	7	150,093,211	8	6	2,900	5	1	0	

Statistical Analysis

Gender columns not found for statistical analysis.

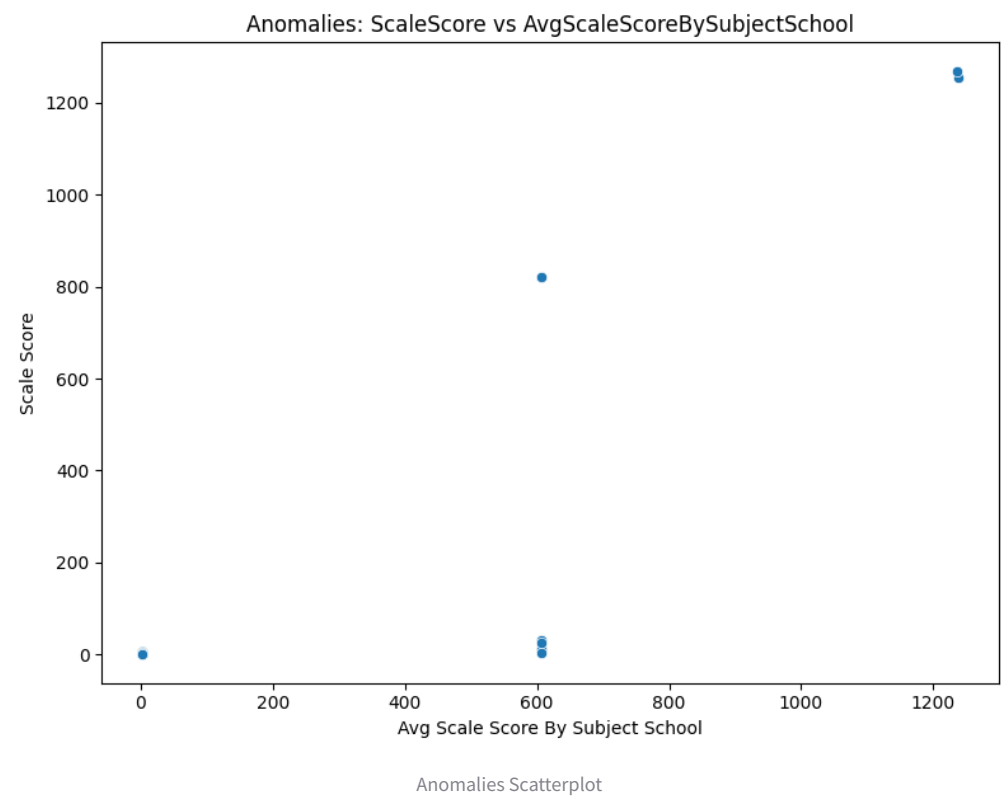
Anomaly Detection

Number of Anomalies Detected: 174

Sample Anomalies

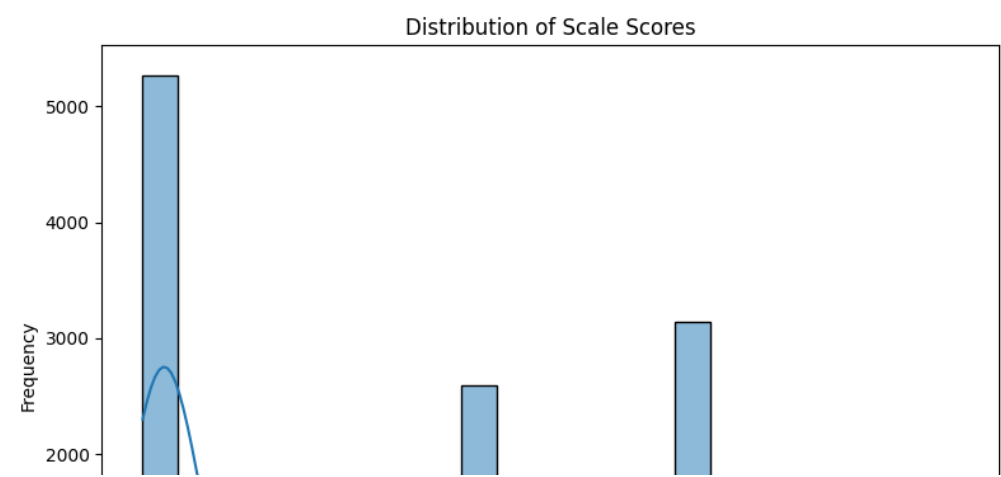
--	--	--	--	--	--	--	--	--	--

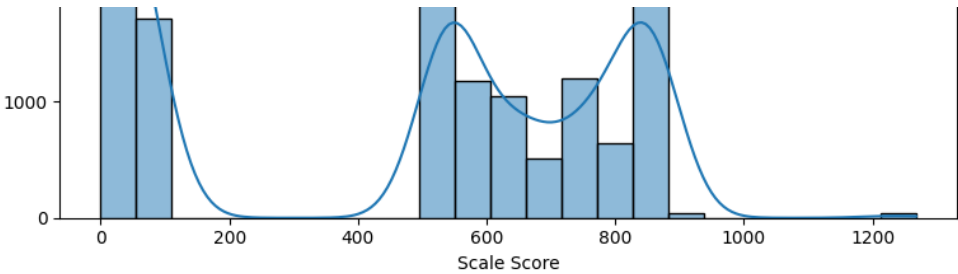
	FiscalYear	StudentID	Grade	Assessment	AssessmentKey	Subject	ELL	Migrant	Homeless
93	0	130,097,404	8	0	1,116	3	1	0	0
264	7	150,038,460	5	0	2,900	3	1	0	1
270	7	150,038,460	5	1	2,663	5	1	0	1
332	0	130,024,529	5	0	1,116	3	1	0	0
636	1	130,173,809	5	0	1,308	3	1	0	1



Data Visualization

Distribution of Scale Scores





Scale Score by Grade

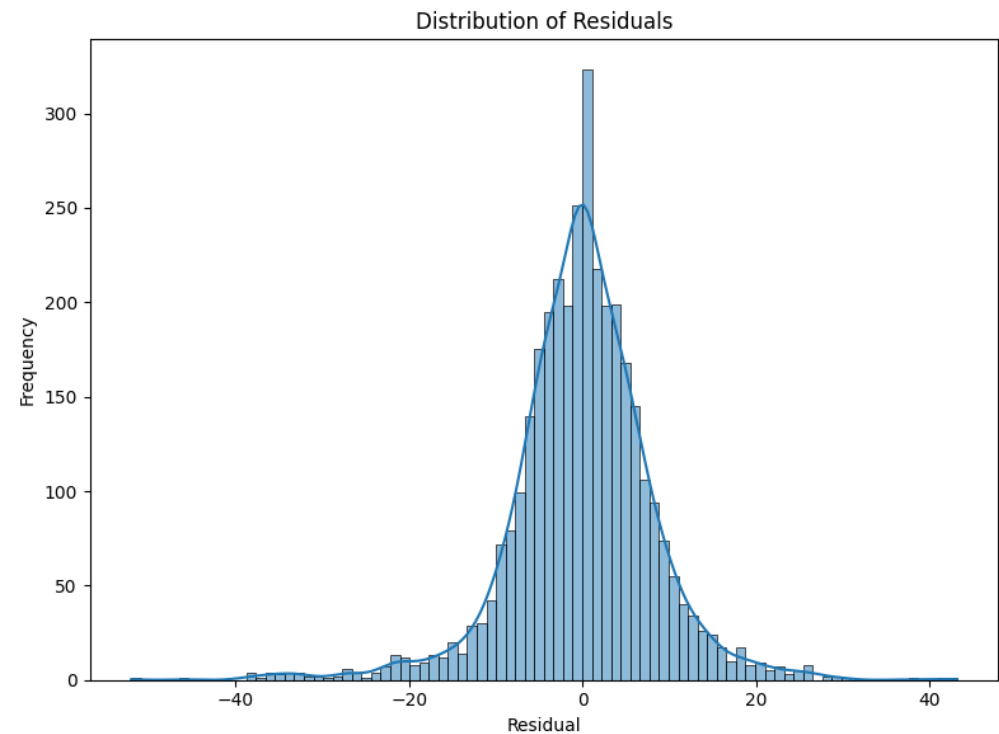
Grade columns not found for boxplot.

Predictive Insights

Students with Lowest Predicted Scale Scores

	PredictedScaleScore	ActualScaleScore	Residual
284	0	0	0
15,635	0	0	0
953	0	0	0
11,195	1	1	0
14,902	1	1	0

Residuals Analysis



Save Evaluation Metrics

Model evaluation summary saved as 'model_evaluation_summary.csv'.

	Model	RMSE	R ² Score
0	Random Forest	8.1506	0.9994
1	Linear Regression	108.9457	0.8995
2	Gradient Boosting	8.9776	0.9993

Data Overview

	FiscalYear	StudentID	Grade	Assessment	AssessmentKey	Subject	ELL	Migrant	Homeless
0	1	130,072,706	5	0	1,308	3	1	0	0
1	2	130,072,706	6	0	1,536	3	1	0	0
2	3	130,072,706	7	0	1,809	3	1	0	0
3	1	130,072,706	5	1	1,247	5	1	0	0
4	1	130,072,706	5	1	1,243	5	1	0	0

Data Description

	FiscalYear	StudentID	Grade	Assessment	AssessmentKey	Subject	ELL	Migrant	Homeless
count	17,375	17,375	17,375	17,375	17,375	17,375	17,375	17,375	17,375
mean	3.1503	130,976,709.6822	6.4784	3.7559	1,785.1014	3.2715	1	0	0.0006
std	2.5006	4,765,119.1508	1.2412	2.1744	694.3289	1.6264	0	0	0.1111
min	0	123,063,508	5	0	723	0	1	0	
25%	1	130,031,667	5	1	1,188	2	1	0	
50%	3	130,084,911	6	4	1,737	4	1	0	
75%	6	130,139,223	8	6	2,567	5	1	0	
max	7	150,093,211	8	6	2,900	5	1	0	