Insight Report: Analyzing School Learning Outcomes

Objective

This report presents findings from a multivariate analysis of government school performance data across three districts. The goal was to identify key factors influencing Math scores and segment schools to target interventions more effectively.

Step 1: Feature Description

%_Language_Score - Average language score

%_Science_Score - Average science score

Teacher_Student_Ratio - Number of students per teacher

Avg_Teacher_Experience_Years - Teacher experience in years

Teacher_Trainings_Attended - Training sessions attended

Electricity_Available - Binary: 1 = yes, 0 = no

Internet_Available   - Binary: 1 = yes, 0 = no

Library_Available - Binary: 1 = yes, 0 = no

%_Marginalized_Students - Share of marginalized students

%_Students_With_Home_Internet - Share of students with internet at home

Parent_Literacy_Rate - Parental literacy level.

Step 2: Data Preprocessing

- Dropped rows with missing %_Math_Score
- Filled missing:
    - Teacher_Student_Ratio with median
    - Internet_Available with 0
    - Parent_Literacy_Rate with Block-wise mean, then overall median
- Converted categorical features to appropriate types
- Normalized numerical features using StandardScaler.

Step 3: Multiple Linear Regression

- Target: %_Math_Score

- Strong predictors:

    - **Positive**: %_Science_Score, %_Language_Score, Parent_Literacy_Rate

    - **Negative**: Teacher_Student_Ratio, %_Marginalized_Students

- Interpretation: Parental education and strong teaching environment drive better math performance.

Step 4: Clustering with KMeans

- KMeans applied with **k = 3** (based on Silhouette Score)

- Resulting clusters:

    - **Cluster 0**: High-performing schools, better infrastructure

    - **Cluster 1**: Moderate performance, mixed infrastructure

    - **Cluster 2**: Low-performing, rural, higher marginalization

Step 5: Principal Component Analysis (PCA)

- Used for dimensionality reduction and visualization

- PCA showed clear cluster separation on 2D plot

- Helped confirm the strength of clustering structure

Highlights:

1.Correlation Highlights

Correlation analysis showed:

- Strong positive correlation between Math scores and Science/Language scores

- Moderate positive link between Math scores and Parent Literacy Rate

- Weak or negative correlation with Teacher-Student Ratio and % Marginalized Students

2. Key Insights by Cluster

- Cluster 0 schools are typically urban with better infrastructure and outcomes

- Cluster 2 schools are mostly rural, with lower access to internet, toilets, and libraries

- Parent literacy and home internet access are major differentiators across clusters

❖ Recommendations

- Focus infrastructure development in Cluster 2 regions (rural blocks lacking basics)

-Promote parent literacy initiatives and community learning programs

-Boost digital access through local internet hubs or tablet provisions

-Invest in teacher training in high student-load areas (high T-S ratio)

Appendix: Key Visualizations

Correlation Heatmap

School Clusters (PCA Projection)

Regression Coefficients Impact
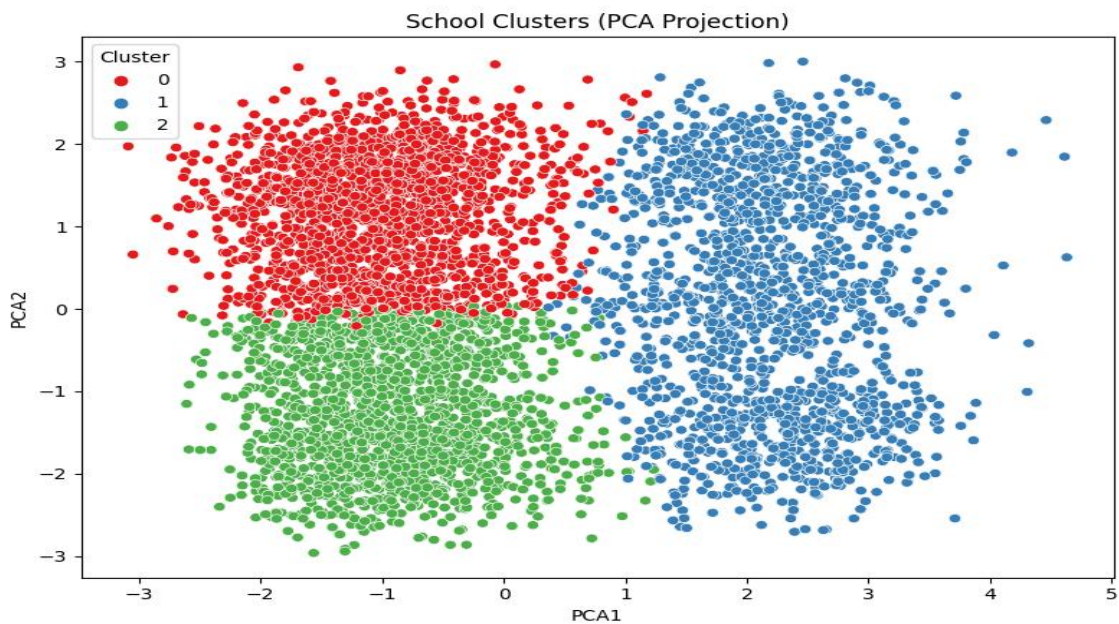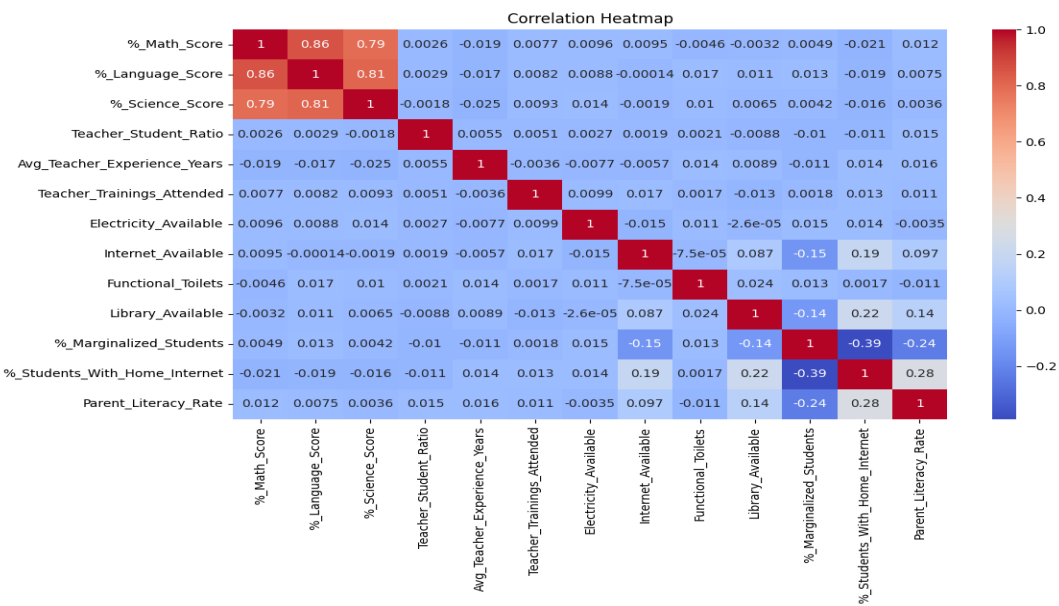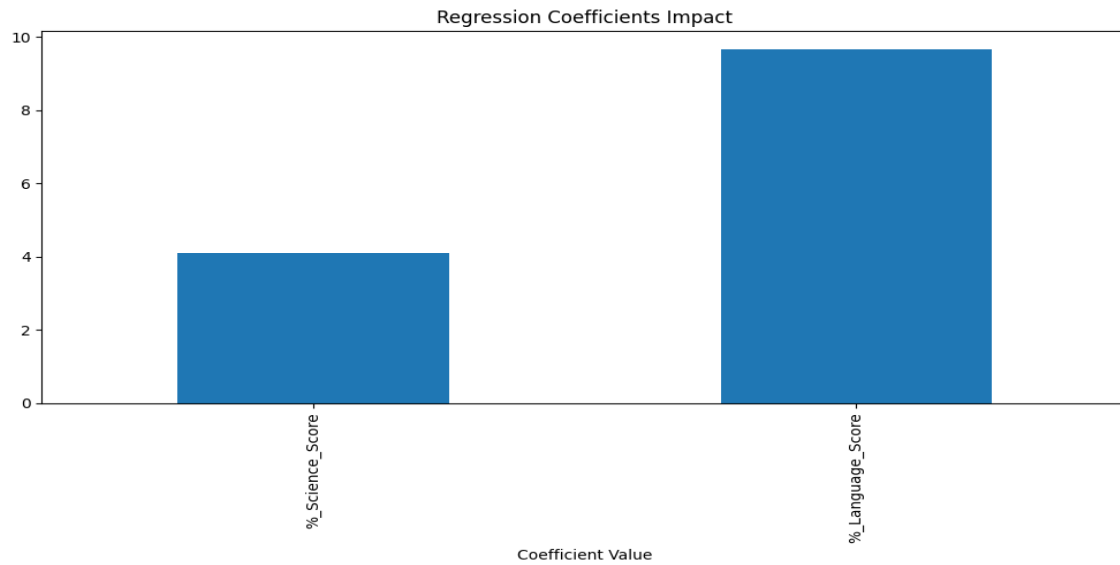
Conclusion :

This multivariate analysis revealed key drivers of math performance in schools. PCA and clustering helped segment schools meaningfully. Targeted interventions should focus on:

- Improving basic infrastructure (Cluster 2)

- Promoting parental education and digital access

- Supporting marginalized groups with focused educational programs