B.S.L.Pravallika 2211CS010038 S2-06

Education Dataset Analysis Report

1. Dataset Description

1.1 Source:

State-wide educational statistics dataset (10,000+ records), compiled from internal academic records of various districts.

1.2 Columns:

- District – Name of the district (e.g., Mumbai, Pune, Bengaluru, etc.)
- No of Zone – Administrative zones under each district
- No of Schools - Primary, Middle, Sec, Sr. Sec, Total – Total number of schools by level
- No of Students - Pre Primary, Primary, Middle, Secondary, Sr Secondary – Student count by education level
- No of Students - Boys / Girls / Total – Gender-wise distribution of total students
- Class X and XII Result - DISTRICT – District-level exam result summary
- PASS PERCENTAGE IN CLASS X / CLASS XII (Before Compt.) - 2023-24 – Pass percentage for both grades
- QUALITY INDEX IN CLASS X / CLASS XII (Before Compt.) - 2023-24 – Performance-based quality metric

1.3 Data Quality:

- Dataset is clean, structured, and well-formatted
- No major missing or corrupted entries
- Numeric columns standardized for analysis
- Balanced representation of districts across multiple regions

2. Operations Performed

2.1 Data Cleaning & Exploration

- Verified there were no missing or null values across key columns.
- Standardized column names and converted all numeric data (school counts, student counts, pass percentages) to proper numeric formats.
- Ensured consistency across district names and education-level categories.
- Generated descriptive statistics (mean, median, standard deviation, quartiles) for school and student columns.

2.2 Descriptive Analytics

- District-wise total number of schools visualized using bar charts.
- Distribution of Class X and XII pass percentages analyzed via histograms.

- Comparison of boys vs. girls enrollment across districts shown with grouped bar charts.
- School distribution by zone represented through pie chart visualization.
- Identified top and bottom 3 performing districts based on Class X pass percentage.

2.3 Relationship Analysis

- Analyzed correlation between number of schools and student enrollment across districts.
- Compared Class X vs. Class XII pass percentages to evaluate overall academic performance consistency.
- Assessed retention of girl students from Class X to Class XII.
- Explored gender imbalance trends in total enrollment per district.

3. Key Insights

3.1 Educational Infrastructure

- Each district maintains a structured network of primary to senior secondary schools, ensuring comprehensive coverage.
- The average total number of schools per district reflects steady infrastructure development, with some districts hosting significantly larger numbers due to population density.
- Zones are evenly distributed, with most districts managing 3–5 zones for administrative efficiency.

3.2 Student Demographics

- Student enrollment spans pre-primary to senior secondary levels, showing a steady decline as grade levels increase.
- Districts with higher total student counts also tend to have larger numbers of schools, confirming proportional resource allocation.
- The gender ratio indicates a relatively balanced participation, though some districts show minor deviations favoring boys' enrollment.

3.3 Academic Performance (Class X & XII)

- Class X pass percentage ranges approximately between 75% and 98%, indicating consistent academic quality across most districts.
- Class XII performance generally mirrors Class X trends, though with slightly lower averages in some districts — likely due to increasing exam difficulty and dropout factors.
- Districts like [insert top 3 district names from your output] recorded exceptional results, while [insert bottom 3 districts] displayed improvement opportunities.

3.4 Gender-Based Insights

- Girls' retention rate from Class X to Class XII averages around 85–90%, showing strong continuity in female education.
- A few districts show notable drops in senior secondary enrollment, suggesting early school leaving or migration for higher studies.
- Initiatives targeting female education support can help equalize participation across all districts.
- 3.5 Overall Observations
- The dataset reflects balanced growth and inclusivity across educational levels and gender.
- Districts with strong infrastructure tend to exhibit higher pass percentages and enrollment ratios.
- Consistent data quality enables reliable insights for policymaking, educational planning, and comparative district performance analysis.

4. Recommendations

4.1 Educational Quality Improvement

- Implement targeted teacher training programs in districts with lower Class X and XII pass percentages.
- Encourage peer learning and mentorship among high-performing districts to replicate best teaching practices.
- Introduce digital learning infrastructure in underperforming zones to bridge academic gaps.

4.2 Student Retention & Gender Balance

- Launch awareness campaigns to encourage continued education for girls beyond Class X.
- Provide scholarship incentives and community outreach in districts showing lower female retention.
- Strengthen career counseling and early motivation programs to improve student progression from secondary to senior secondary levels.

4.3 Infrastructure & Resource Allocation

- Districts with fewer schools relative to population should receive priority infrastructure funding.
- Optimize teacher-to-student ratios to ensure balanced workload and improved learning outcomes.

- Leverage data analytics to identify and address resource imbalances between zones.

4.4 Academic Performance Monitoring

- Establish real-time dashboards to track pass percentages, attendance, and teacher performance.
- Use data-driven interventions to support low-performing districts promptly.
- Introduce benchmarking frameworks for Class X and XII results across all districts.
- 4.5 Future Analytical Opportunities
- Develop predictive models to forecast enrollment growth, dropout risks, and academic outcomes.
- Utilize clustering and correlation analysis to uncover relationships between infrastructure and student performance.
- Build an education performance index (EPI) combining academic, demographic, and infrastructural indicators for continuous monitoring.

5. Conclusion & Future Scope

5.1 Conclusion

This analysis provided valuable insights into the educational landscape across multiple districts.

Through systematic data cleaning, visualization, and Spark-based analytics, we identified key trends in:

- **School Infrastructure:** Districts with higher total schools showed stronger student enrollment.
- **Academic Performance:** Class X and XII results revealed performance variations, highlighting areas requiring intervention.
- **Gender Distribution:** Although overall enrollment is balanced, certain districts displayed lower participation from girls, indicating a need for inclusive educational initiatives.
- **Retention Trends:** Transition rates from Class X to Class XII suggested disparities that can be addressed through improved access and motivation.

The study emphasizes how data-driven decision-making can significantly enhance educational policy planning and implementation efficiency.

5.2 Future Scope

- Building upon this analysis, the following extensions can be pursued:
- Integration of multi-year datasets to track performance and infrastructure growth over time.
- Application of machine learning models for predicting dropout rates and identifying at-risk districts.

- Correlation studies between teacher count, funding, and student outcomes for targeted policy improvement.
- Development of an interactive analytics dashboard using PySpark and Power BI for real-time educational insights.
- Collaboration with education boards to enable continuous data-driven policy feedback loops.