**School Performance and Standardized Test Data in the City’s School District**

**Executive Summary**

The role of the Chief Data Scientist for the city's school district is to aid the school board and the mayor in making well-informed decisions regarding school budgets and priorities. The initial task was to analyse district-wide standardized test results, incorporating math and reading scores alongside relevant school information. This analysis involved two key datasets: one focused on individual student data and the other providing insights into the schools themselves. These datasets encompass standardized test results, student particulars, and school characteristics.

The analysis aims to reveal valuable insights that can guide the district's future decisions and help improve educational outcomes. In the following sections, we will delve into the details of the analysis and present the findings and recommendations.

**Methodology**

**Data Sources**

The data was generated by Mockaroo, LLC (2022), and edX Boot Camps LLC for educational purposes. The datasets consisted of two files (school\_complete.csv and students\_complete.csv) in a zipped folder titled starter\_code. The student-level data included standardized test results for math and reading, along with details about each student. The School-level data contained information about the schools, including their names, types (Charter or District), sizes, total budgets, and per-student budgets.

To ensure the data's quality and relevance, it underwent cleaning, handling of missing values, and conversion of currency values into numerical formats. Grouping and aggregation were employed for group-level analysis based on school type, size, and spending ranges, following a structured methodology to ensure the data was well-processed and suitable for analysis.

**Data Processing and Pandas Integration**

The Python programming language (version 3.10.13 packaged by Anaconda, Inc.) alongside version 2.0.3 of the Pandas library is utilised within Jupyter Notebook (version 5.3.0) for data processing and analysis. Leveraging Pandas efficiently facilitated the organization and manipulation the information into a structured tabular format, primarily using the Pandas DataFrame as the primary data structure. The DataFrame facilitated data cleaning, handling of missing values, and overall data preparation for a meaningful and insightful analysis. Furthermore, the codes were generated using Pandas documentation (Pandas Documentation, 2023).

**Key Findings and Analysis**

Based on the data provided, the district has a total of 15 schools, serving 39,170 students, with a combined budget of $24,649,428.00. The average math score is approximately 78.99, and the average reading score is approximately 81.88. About 74.98% of students pass the math standardized test, while approximately 85.81% pass the reading test. The overall passing rate, considering both math and reading, is approximately 65.17%.

**School Performance by Type:**

It was observed that there were two school types: Charter and District.

Charter schools showed higher performance in both math and reading, with significantly higher passing rates. This indicates that charter schools may have more effective teaching methods or resources. District schools on the other hand, showed lower average scores and passing rates. This suggests that additional support may be required in district schools to improve student outcomes.

**School Performance by Size:**

The analysis categorized schools based on their sizes as Small (<1000 students), Medium (1000-2000 students), and Large (2000-5000 students).

Smaller schools displayed higher average math and reading scores, along with better passing rates in math and reading. Larger schools showed lower average scores and passing rates. This could be attributed to challenges related to managing larger student populations.

**School Performance by Spending:**

The analysis categorized schools based on their per-student budgets into spending ranges. The categories were: “<$585, $585-630, $630-645, and $645-680”.

Charter schools with lower per-student budgets exhibited higher average scores and passing rates. This indicates that cost-effective education may lead to better student outcomes.

District schools with higher budgets did not necessarily translate to better performance, suggesting that additional spending does not always result in improved outcomes.

**Recommendations for Strategic Improvement**

**Examine School Type Dynamics:**

Conduct an in-depth exploration of the practices and teaching methodologies employed in charter schools. Evaluate their effectiveness and consider the feasibility of implementing successful strategies within district schools.

**Addressing School Size Challenges:**

Deliberate on the possibility of disaggregating larger schools into more manageable smaller units. Alternatively, strategically allocate additional resources to enhance student-teacher interaction, especially in larger schools.

**Enhancing Budget Optimization:**

Direct efforts towards meticulous optimization of per-student budgets to improve overall student performance. Prioritize investments in cost-effective strategies that ensure the delivery of quality education, maximizing the impact of financial resources.

**Conclusion**

The analysis unveils profound insights into school performance, considering factors such as school type, size, and spending. Charter schools exhibit excellence, smaller schools consistently outperform larger counterparts, and the relationship between lower per-student budgets and superior outcomes challenges conventional assumptions. These collective findings not only offer a foundation for strategic decision-making but also underscore the intricate interplay of diverse factors influencing academic performance in the district. As the district aspires to elevate overall educational outcomes, a meticulous exploration into successful models, allocation strategies, and educational practices becomes imperative for informed and effective decision-making.

**References**

Mockaroo (2022). Data generated by [Mockaroo, LLC Links to an external site.](https://mockaroo.com/). Realistic Data Generator. Data for this dataset was generated by edX Boot Camps LLC and is intended for educational purposes only.

Pandas Documentation (2023). Retrieved from <https://pandas.pydata.org/pandas-docs/stable/index.html>. Retrieved on November 07, 2023.

**Report Prepared By:** Godswill Anyasor