Colorado Public School Academic Performance

A Study of Financing versus Student Performance

Jarryd Allison
CSPB 4502: Data Mining
University of Colorado
Boulder, CO 80309
allisonj@colorado.edu.com

Abstract

This report seeks to identify any existing correlation between school finances, specifically spending per student, and the mathematical academic performance of school. The data used is publicly available, and includes performance by grade, as well as spend per student and the total budget. Contingency analysis was used, along with other data mining techniques, to uncover a weak but perpetually present negative correlation between spend per student and the mean scale score achieved in each grade, from 2019 to 2023. Because results included 2019, this negative correlation cannot be solely due to the COVID school shutdowns, though those undoubtedly played a part in the correlation spike and poor performances seen in 2020/2021. The report concludes with applications for parents, teachers, and legislatures, as well as other researchers, to continue to investigate and integrate additional data to better understand variables affecting educational performance, and close the gap between the low and high performers. From the analysis, money alone tends to have a slightly negative impact, therefore monetary and other solutions must be presented in order for true change to occur and the negative trends to reverse significantly.

Introduction

Public education in the United States has come under increased scrutiny of late, due to consistently poor trends in academic performance. This has hit the Denver metropolitan area as well, with numerous articles bemoaning the state of public education[1,2].

The most recent data from US News for Denver area schools shows that while the operating budget has grown considerably over the past decade[3], test scores continue to decline or at least fail to improve. Proficiency scores for reading hover in the low 30s to upper 40 percentages, but math remains stuck below 30% proficiency across all age groups[4].

The combination of low test scores but accompanied with a growing budget seems to be counterintuitive, and demands closer inspection to better understand the various relationships that exist within the available data. For example, Denver does allocate funding differently based on perceived and measured needs of students[5]. Could a focused look at financing across Denver public schools reveal additional measures to better allocate funding, or provide links between funding and increased performance? The aim of this research topic is to answer this critical question.

Additionally, there may be other underlying correlations and interesting findings that reveal themselves through the conduct of this research. Additional findings will be included in the final report.

Specifically, I wish to address the following questions:

1. Does a correlation exist between school funding and mathematics performance? If so, what is this correlation, and what are its most important variables?
2. How has funding and mathematics performance changed given the dataset from 2019 - 2023? It is noted that funding data is not readily accessible prior to 2019, which altered the study's range to the range of funding information available.
3. What variables appear to influence mathematical performance the most?
4. What variables appear to influence funding the most?

CCS Concepts

Social and professional topics → Professional Topics

Keywords

Education, finance, data mining, analysis

ACM Reference format:

Jarryd Allison. 2024. Colorado Public School Academic Performance: A Study of Financing versus Student Performance. In *. ACM, New York, NY, USA, 3 pages.* https://doi.org/10.1145/XXXX

Related Work

There exists a large number of articles and research that describe the issues that Denver Public Schools have with regards to test scores and low performance[1,2,6], and financial data exists to describe the growth in the budget[3], but very little reporting and data mining exists to justify increases, at least from what appears to be publicly available at the time of the writing of this paper.

Aside from the sources listed, no analysis exists that intends to identify patterns and provide a holistic view of the budget versus student performance, and possibly help inform public policy to level set expectations, understand expenditures, and identify programs that may increase performance and educational outcomes in Denver public schools. There do exist compilation reports describing funding and performance, though these fail to describe significant patterns and a deeper analysis regarding the data[9].

Data Set

The datasets identified for the scope of this research are as follows:

1. **Colorado Measures of Academic Success (CMAS) - Mathematics, English Language Arts, Science and Social Studies Data and Results (2014 - 2023)**

The CMAS performance dataset[7] is compiled and provided by the Colorado Department of Education annually. It is provided in multiple formats, to include .csv format, and is available for download. Limitations are myriad in this dataset. While it does present the most comprehensive dataset available for this information, it only gives mean scale score and the standard deviation, while omitting high/low scores and more accurate representations of the scores across each school as a whole. While valuable metrics can still be gleaned from this, they would be better served with more scoring information. This report attempted to quantify these restrictions by using the scores and standard deviations to determine if any interesting trends emerge. Additionally, this data set includes scores for English Language Arts (ELA), Spanish Language Arts (SLA) and Mathematics. This report looked singularly at mathematics performance across Colorado schools. The most pressing reason for the data mining to focus on this aspect of educational performance is because of aggregated reports[4], the mathematical performance is much poorer than ELA/SLA performance.

This dataset is quite rich in attribute features, with 31 available in the 2023 data set. While many of the data attributes could generate interesting patterns, this study was most concerned with the following:

- **District Code**: This was used to iterate through the data set and scrape for financial data.
- **School Code**: This was used to iterate through the data set and scrape for financial data.
- **Grade**: Individual grades were cleaned and organized on each school row, simplifying data visualization and allowing for a more detailed examination of performance by grade.
- **Number of Valid Scores**: A large number of schools (reported below) did not meet the state threshold for reporting, which per the data set appears to be fewer than 16 tests taken. This, and its impact on the results, is discussed later.
- **Mean Scale Score**: The mean scale score was used to calculate the effect of spend on the average mean scale score performance of schools by grade, as well as to support visualizations of mean scale score performance.
- **Percent met or exceeded expectations**: For 2023, this was simply under the column header '2023', which indicated the reported percentage of students that met or exceeded the state minimum required scores for testing.

2. **Colorado Department of Education Financial Transparency Office**

The Colorado Department of Education also offers Financial Transparency, both in the form of an online tool[8] as well as comprehensive quarterly financial statements[3]. These datasets are not as well structured as the performance data, and will need to be heavily processed in order to correctly align them with the performance dataset described above. This processing will be done with the creation of a web scraper. Initially, 2023 data will be utilized, but, given additional time, historic data through 2019 is present, giving a much richer data set that will help reveal insights to the correlations between performance and funding. The largest issue with this dataset is the omission of reporting prior to 2019. This could be gleaned by parsing PDFs also available, but that would be well outside the scope of this course.

This data set was also relatively rich in information, but limited to the past four years of data. The two main data points that were used in this data mining were:

- **Spend Per Student**: This reported number is calculated for each school and reported in a consistent html format, allowing for data scraping for each school.
- **Total Spending**: This reported number is calculated for each school and reported in a consistent html format, allowing for data scraping for each school.

Main Techniques Applied

The work involved in creating this report involved a careful process of cleaning, preprocessing, and integration of multiple datasets to better understand the problem set. This differs from the literature survey conducted simply because the depth of data mining is lacking. The literature survey indeed studies the test performance dataset, and also arrays anecdotal evidence atop it to determine whether performance is good or bad.

But the literature review is severely lacking in objective data mining techniques to truly develop a deeper understanding, through pattern exploration and data mining. Additionally, layering in financial data also appears to be completely ignored by the majority of the literature reviewed for this project.

The following work was completed in the conduct of crafting this report, including data cleaning, data preprocessing, and data integration.

1. **Data Cleaning**

As stated in the Data Set section, much of the data exists in excel documents, specific to individual school years, with omissions and various discrepancies for multiple attributes across the data sets. Additionally, in the course of cleaning this data, it was discovered that a reporting threshold appears to exist for Colorado schools. If a school can report fewer than 16 valid test scores, they are excused from providing data. This presented an issue when creating the final data structure as an output of this process.

The data cleaning step was intended to first infer missing data, using either the mean, median, to apply the central tendency, or from determining the most likely values from surrounding data points. However, conducting the data mining on the data set would become more inconsistent when comparing it to existing analysis. Additionally, this would unintentionally skew the data. Instead, the data mining conducted extended to identifying how many schools did not meet the threshold, which is discussed later.

Additionally, throughout the conduct of cleaning the data, it was also discovered that the csv format shifted from year to year. For example, years 2023 and 2022 collected an aggregate “Met or Exceeded Expectations” value simply under the column header of that year (e.g., for 2023 this was simply “2023”). However, prior to 2022, this header was under the title “Percent Met or Exceeded Expectations.” Cleaning and collating these values to the proper locations was paramount to ensuring the data was trustworthy, and involved a significant amount of manual and automated processes to ensure the data was accurately transcribed to a single master csv for data mining.

Finally, this step also included the cleaning of data related to the budget data collected by scraping the public information[3] available as discussed above. The results of the initial pull included the entire HTML tree of the website. This was analyzed to pull the specific “Spend Per Student” and “Total Budget” attributes required for the later stages of data mining. The results were spot checked to ensure accuracy, and then the scraper run for the full 2023 dataset. Previous data sets were not run and compared, because the information was organized per the 2023 list of schools, and the web scraper returned data for the previous four years, which allowed data mining on the change in spend data over the time period where performance was also available.

2. **Data Preprocessing**

Data preprocessing first involved data reduction, given the size of the dataset and the scope of the project. Multiple attributes simply did not apply, and were removed. For example, all the available data sets per year included District and Colorado-wide roll-ups of the results. However, this data mining was meant to go further, relying on the school-level data and even into specific grades to determine if a correlation existed between spend and mathematical performance. Therefore, the District and School level information was simply removed. Additionally, multiple attributes, as discussed above, were not

considered in the final analysis, although future data mining could rely heavily on other attributes not accounted for in this report.

Duplicates were not an anticipated issue, as the data is structured by schools in the Denver area, meaning that duplicates were highly unlikely, but the data sets were screened to ensure duplicates did not exist.

This step also included preprocessing each of the available csv's by school year into pandas dataframes in python, as well as the financial data. This prepared each data set to be integrated into a single csv prior to analysis.

3. **Data Integration**

This step proved both the most difficult and most important in preparing for the data mining of the available data. Using the 2023 performance csv as a base, each csv was read into a pandas dataframe. Each data frame then removed all but the school level (omitting the district and state roll-up) and the mathematics performance data. Because the 2023 data set was used to scrape the financial data, the 2023 data set was extended to include the spend data for each of the previous four years. Using this as a base, each of the previous four years performance data and mean scale scores were added to the 2023 data set, creating new columns for each grade and finding the specific values for each grade in a given year.

The end result was saved as a csv that could be read in for data mining. This extended the original data attributes from 31 to 88, including performance at each school level for all grades, and grades three through eight, though not all schools had data for each grade. Once this step was complete, the data had been fully integrated, and the pattern analysis could begin. This transformation of the data could be described as creating a data warehouse for each historic year, with a relatively straightforward process that could easily ingest and support additional years, and given more time and access, more historic data. While a data cube was not explicitly created for this data mining, multiple subsets of aggregated data were investigated and created during the analysis, which could have been aggregated into a data cube.

4. **Classification**

The main data mining involved classification techniques to determine if any correlation exists between spend per student and the mean scale score of students in each school, by grade. The spend per student was the best attribute to measure, as it is by definition independent of the number of students at each school, and can be compared to the aggregated scores.

To conduct this analysis, contingency tables were created for each grade, for each year. First, the data set subset of the year and grade were selected, and the NaN's removed (where mean scales weren't reported or were 0). Then, the average Mean Scale and average Spend Per Student were calculated. The resulting contingency table was built as shown in Figure 1.

```
2019 Grade 3 Contingency Table
            >= Avg SPS  < Avg SPS  Row Sum
>= Avg MSS         200        279      479
< Avg MSS          293        184      477
Col Sum            493        463      956
Mean Spend Per Student:  12053
Mean Mean Scale Score:  739
Lift Calculation:  0.8097
Correlation:  negative correlation
X^2 Calc:  37.032
Support:  20.92
Confidence:  41.75 %
```

Figure 1: Example contingency table for 2019 Grade 3.

As shown in Figure 1, the lift was also calculated, along with the x^2 correlation analysis applied to determine the correlation between spend and score performance. Additionally, the support and confidence were also calculated, to determine the strength of the correlation.

Finally, numerous other analytical techniques were applied, including scatter plots, 10-bar graphs, and line charts, which identified interesting patterns and trends in the data.

Evaluation Methods

While the body of literature is relatively small, the Common Sense Institute report[9] compiles a list of generalized information regarding performance and finances across Colorado schools. Additionally, historic[10] and more current research[11] seems to be mixed, with previous studies finding no correlation between funding and performance, but more focused studies finding a strong correlation. To date, there does not seem to be any Denver-focused study on this topic. However, the existing research can help measure and provide additional techniques to better understand the relationships that may exist.

Finally, the data set itself provides data that can be used to confirm initial results and ensure that the underlying functions are correctly being utilized across each row. For example, 2023 reporting includes changes from 2022 and 2019 (2021 and 2020 data was mysteriously absent), which can be used to confirm findings year over year in some small regards.

Overall, the findings in this report were supported by the literature that was researched prior to the application of the aforementioned data mining techniques.

Tools

To conduct this research, the following tools were utilized:

1. Jupyter notebooks were heavily utilized to run various data mining techniques and produce images to support the findings.
2. Python was used extensively for data mining, as it is adept at formatting CSVs to run advanced data mining techniques against the underlying data.
3. Numpy/pandas was used in python, as these libraries contain multiple powerful features to simplify the data processing steps. Specifically, pandas was used to create dataframes, which are much easier to utilize when conducting data mining on datasets.
4. BeautifulSoup, a python library used to help parse HTML data, was used to scrape public information regarding Colorado school financial data.
5. Matplotlib is a popular python plotting library that was used extensively to create data visualizations in support of the research goals of this project.

Key Results

The first results of note fostered a deep underlying suspicion of the data results routinely reported in aggregations of Colorado academic

performance[4], which is already underwhelming. Because Colorado allows schools with fewer than 16 valid scores to not report, it is entirely possible that the results are far worse, or better, than the aggregated state performance that is reported. As Figure 2 shows, the number of schools omitting scoring data, while improving slightly for 2023, is tremendous, representing nearly 3000 schools in 2023 alone. Estimating on average 10 scores per school, this total would represent close to 9% of the total valid scores across the state, and certainly a much larger percentage at the District level. It is unknown why the test results are not required to be reported, since it is obvious that the data exists at the school level from the data sets. This omission of data was frustrating, but the data mining continued.

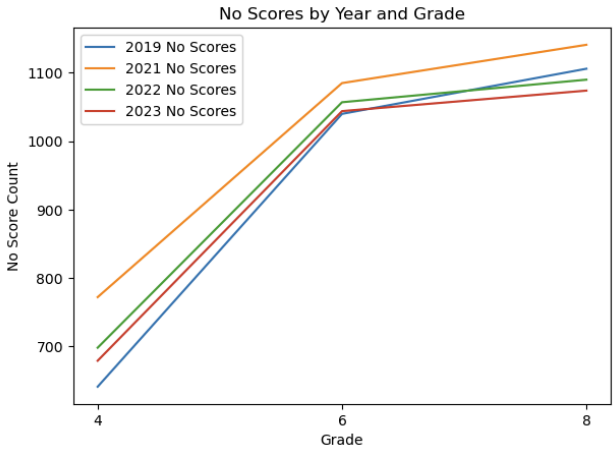

Figure 2: No scores required to be reported, by grade, for each year since 2019.

In spite of this omission, data mining was able to continue, investigating interesting patterns across the data sets available. Most important to the premise of this report was the correlation analysis via contingency tables. This was conducted across all school years, for all grades, where data was available. This revealed an interesting result, showing that while small for many instances, the actual correlation x^2 is always above 1, and therefore a correlation does exist, however small. Additionally, in every case, the observed value was less than the expected value, meaning that for every year data was available for mathematics performance and spend per student, the two were negatively correlated. The results of the x^2 calculation are seen in Figure 3.

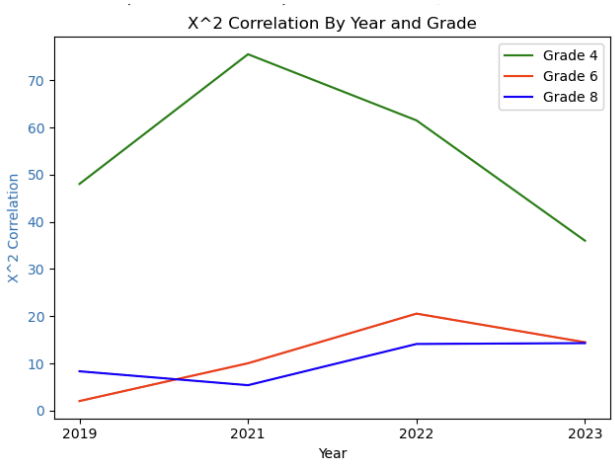

Figure 3: x^2 correlation calculation per year by grade.

Figure 4 shows the contingency table and resulting analysis computed for 2023, Grade 8. Across each of the generated contingency tables, it is quite evident that more schools are achieving average or higher than average Mean Scale Scores with a below average spend per student. While perhaps surprising, this does answer the stated question at the beginning of this report. Both a correlation exists, and it is negative, implying that in aggregate, monetary solutions tend to have a weak negative effect on the performance of students in the Colorado school system, when considering the reported mean scale scores for participating schools.

```
2023 Grade 8 Contingency Table
            >= Avg SPS  < Avg SPS  Row Sum
>= Avg MSS          97         143      240
< Avg MSS          164         124      288
Col Sum            261         267      528
Mean Spend Per Student:  14734
Mean Mean Scale Score:  730
Lift Calculation:  0.8176
Correlation:  negative correlation
X^2 Calc:  14.3059
Support:  18.37 %
Confidence:  40.42 %
```

Figure 4: 2023 Grade 8 Contingency Table.

This result is indeed surprising, and even more surprising when viewing the meets/exceeds expectations percentage for schools who were above average mean scale score, and below average spend, versus schools with an above average spend and a below average mean scale score, as shown in Figure 5.

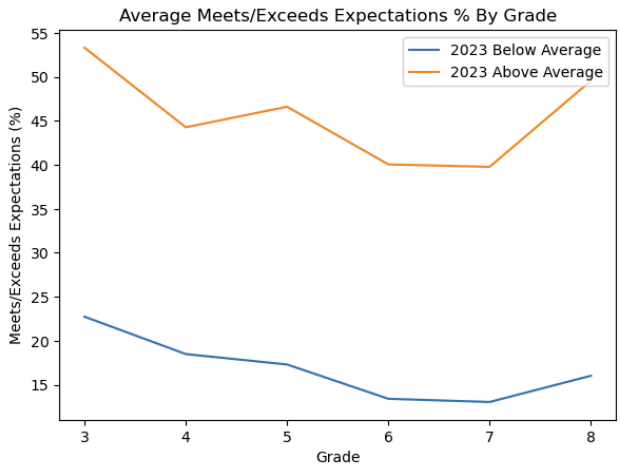

Figure 5: Plot of 2023 percent met/exceeded expectations above average MSS, below average SPS, versus below average MSS and above average SPS.

However, would other data visualizations and data mining techniques reveal an understanding of this, or other patterns, that might help prompt ways to improve performance? Looking at the Mean Scale Scores by year proved fruitful. 2021 was an odd reporting year, wherein only grades 4, 6, and 8 were reported. So, two plots for Mean Scale Score were generated, as seen in Figures 6 and 7.

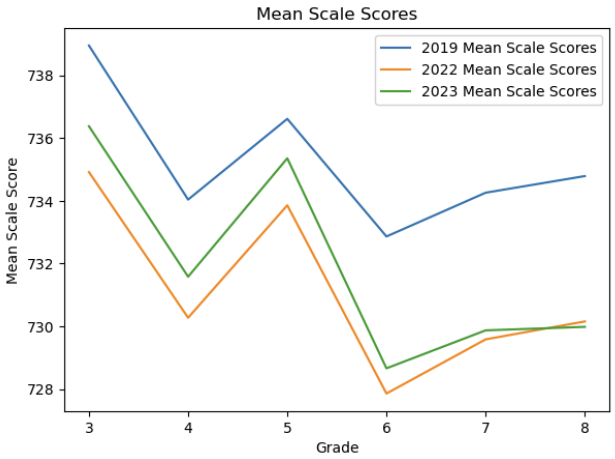

Figure 6: Mean Scale Scores by grade for 2019, 2022, and 2023 for grades 3-8.

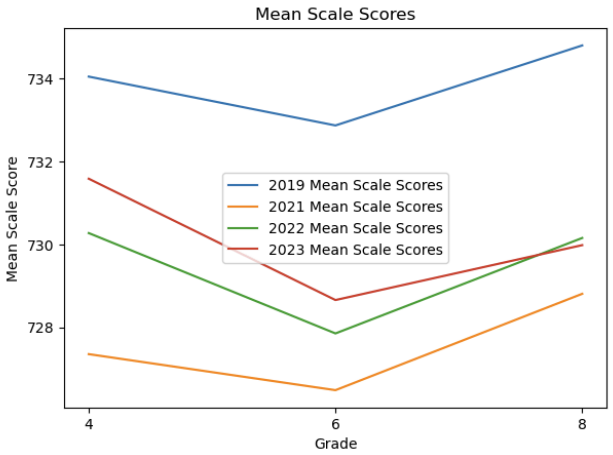

Figure 7: Mean Scale Scores by grade for 2019, 2021, 2022, and 2023 for grades 4, 6, and 8.

These linear plots show a disturbing trend between grades 3 and 4, and 5 and 6, before leveling out at relatively unacceptable levels after grade 7. Figure 6 reveals even more significantly the poor performance by grade. While the state requirements for meeting or exceeding expectations based on the score alone is unknown, the reported data on these

percentages also shows poor trends and a surprising result. Even though our analysis above shows a negative correlation, we can see a positive improvement in the percentage of students who meet or exceed expectations, as shown in Figure 8.

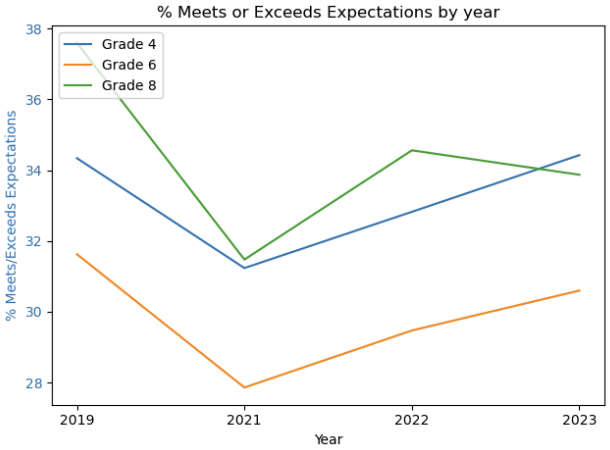

Figure 8: Percentage of students meeting/exceeding expectations by year and grade.

No doubt some of this improvement was due to improvements in educational conditions post-COVID, but now consider the same plot overlaid with the total school budget, which follows a similar trend as the spend per student and spend per school as seen in Figure 9.

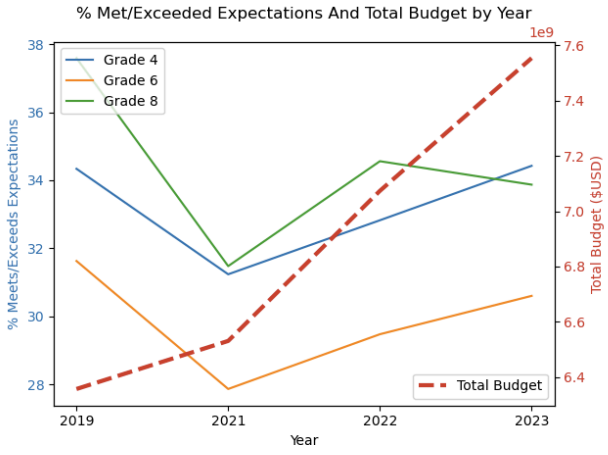

Figure 9: % Met/exceeded expectations by year and total budget by year.

The percentage change in improvement year over year for mathematical performance never exceeds the same percentage change in budget, and, in the case of grade 8 performance, the difference is quite stark.

Now, this analysis shows a clear, albeit weak, negative correlation between spend per student and mathematical performance. One last chart is provided, which shows a scatterplot of performance vs. school spend in Figure 10.

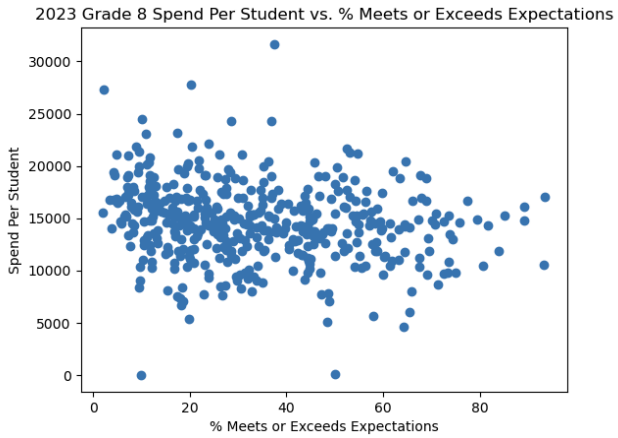

Figure 10: Scatterplot of performance versus spend for 2023 Grade 8 students.

The scatterplot clearly shows the majority of schools well below the 40% mark but also containing the majority of the high spend per student. This is in keeping with the trend shown in previous images and descriptions of the data mining conducted for this report.

Now, the results are not strongly convincing that *reducing* spend would immediately improve scores. In fact, further analysis should be conducted to determine if, where spend decreased, did scores improve. That analysis was not conducted as part of this report. But instead, this report at least shows that more care should be taken to apply funding where it makes sense and can have the greatest impact to mathematical performance for the largest number of students. In the conduct of this

analysis, bar charts by year were created, and interesting anomalies were noted, though none more egregious than that seen for the data from 2021.

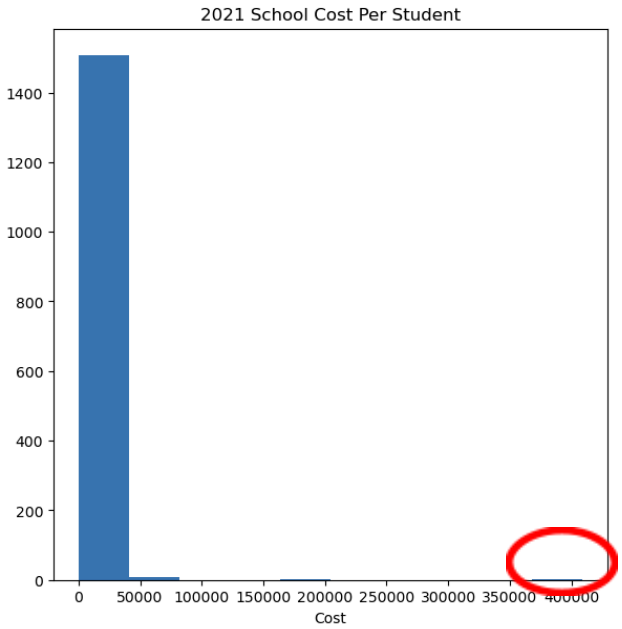

Figure 11: 10 bar chart for Spend Per Student in 2021.

As Figure 11 shows, the data for the 2021 school year was incredibly skewed by a single, massive outlier. From the state's own financial reporting[3], we see that the data actually supports this incredible result. Figure 12 is directly from the state's financial reporting site:

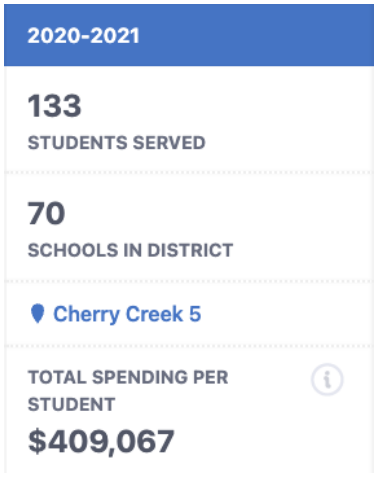

Figure 12: State reporting for the largest spend per student.

It remains possible that this data is erroneous, but it carries weight as it is hosted by the state's own reporting branch. Additionally, this outlier reported fewer than 16 tests for this year, meaning that it too was exempt from reporting the number of students that met or exceeded expectations, and we cannot even surmise if this ridiculous sum improved scores.

Applications

This data is surprising, seemingly counterintuitive, and above all depressing considering the performance of hundreds of thousands of children in the state education system of Colorado. Despite over $7B (USD) spent in 2023, performance remains low, with no solid plan in site to either appropriate funds more accurately to where it may improve scores, or indeed any cohesive plan in the legislature to address this shocking shortfall in academic performance. This report, then, is applicable in a number of different venues, all within the public sphere.

1. **Articles and awareness**

It is the intention of the author to publish these findings in a series of substack articles or twitter articles. The spending represented in the above findings comes directly from state taxpayers, many of whom know or have participants currently in the education system. Increasing public awareness of the performance of public education in the state can result in real change, by prompting legislatures to review expenditures, correct erroneous reporting, conduct investigations, and acting on this data to help improve student outcomes.

2. **Further analysis**

This report may spur additional researchers, and even this author, to continue to conduct research, find more information, and better understand the underlying patterns in academic

performance. No doubt additional overlaying of census, location, average income, ethnicity, and other such data sources would more richly identify the various underlying variables contributing to the educational decline in the state.

3. **Legislative action**

Should the data be collected into a convincing format, it could also drive legislative plans to improve the educational outcomes. The example in Figure 12 is but one instance where a close eye must be kept on how funds are appropriated and spent. No doubt underfunded schools would benefit immensely from a second review of how money is allocated.

4. **School initiatives**

PTA organizations and individual schools should review the top performers who are able to spend less per student, but achieve above average performances from their students. Specifically, Figure 5 identifies a massive gap that must be closed to ensure students are both learning and granted more opportunities as an aggregate.

REFERENCES

[1] Melanie Asmar, 2022. More Colorado schools and districts earn low state ratings. https://www.chalkbeat.org/colorado/2022/9/8/23343341/colorado-school-performance-framework-ratings-2022/.

[2] Alan Gottlieb, 2022. Performance plummets in Denver's schools. https://gazette.com/opinion/denver-columns/column-performance-plummets-in-denver-s-schools/article_54f8631a-b4dd-11ec-ac74-c36f0357135b.html

[3] Colorado Financial Services Reports, accessed 2024 for years 2014-2024. https://financialservices.dpsk12.org/o/financialservices/page/financial-transparency

[4] US News. 2024. Overview of Denver Public Schools. https://www.usnews.com/education/k12/colorado/districts/school-district-no-1-in-the-county-of-denver-and-state-of-c-112125

[5] Melanie Asmar, 2022. The $3,500-per-student difference between two Denver schools. https://www.chalkbeat.org/colorado/2022/4/28/23045997/denver-student-based-budgeting-smith-carson-elementary/.

[6] Ari Armstrong, 2023. Denver Public Schools celebrates lowering achievement bar. https://pagetwo.completecolorado.com/2023/09/26/armstrong-denver-public-schools-celebrates-lowering-the-bar/

[7] CMAS - Mathematics, English Language Arts, Science and Social Studies Data and Results, 2024. https://www.cde.state.co.us/assessment/cmas-dataandresults

[8] Financial Transparency for Colorado Schools, 2024. https://www.cde.state.co.us/schoolview/financialtransparency/homepage

[9] Common Sense Institute, 2022. Dollars and Data: A look at PK-12 Funding and Performance in Colorado. https://commonsenseinstituteco.org/dollars-and-data-a-look-at-pk-12-funding-and-performance-in-colorado/

[10] Lawrence O. Picus, 1995. Does Money Matter in Education? A Policymaker's Guide. https://nces.ed.gov/pubs97/web/97536-2.asp

[11] C. Kirabo Jackson, Rucker C. Johnson, Claudia Persico, 2015. The effects of school spending on Educational and Economic Outcomes: Evidence from School Finance Reforms. https://www.nber.org/system/files/working_papers/w20847/w20847.pdf