

Kimberly Bonilla
Data Analytics
Module 4 Challenge
March 27, 2024

Module 4 Challenge

Background

In this assignment, I have been tasked with working with two datasets named `schools_complete.csv` and `students_complete.csv` which have been consolidated into a single, unified dataframe. I will proceed to conduct an in-depth analysis of the district-wide standardized test result, focusing on students' performance in math and reading. Additionally, I will explore information regarding the schools they are enrolled in.

District Summary

I successfully calculated key metrics including the total number of unique schools, total number of students, total budget, average math score, average reading score, percentage of students passing math, percentage of students passing reading, and overall percentage of students passing. Upon completing these calculations, I was able to observe we had 15 unique schools, a total of 39,170 students, and a total budget of \$24,649,428. Furthermore, upon analysis, it was observed that the average math score, recorded at 78.99%, was slightly lower than the average reading score, which stood at 81.88%. Additionally, the percentage of students passing math, at 74.98%, was slightly lower than the percentage passing reading, which stood at 85.81%. Furthermore, the overall passing percentage was recorded at 65.17% which is fairly low.

School Summary

I successfully computed key metrics, encompassing the school name, school type, total students, total school budget, per student budget, average math score, average reading score, percentage passing math, percentage passing reading, and the percentage overall passing. Subsequently, I generated a DataFrame named "school_summary" to depict the key metrics regarding each school. Upon reviewing the new DataFrame "school_summary," it became evident that district schools boasted larger student populations compared to charter schools. Additionally, there was a notable discrepancy in budgets, with charter schools having significantly lower budgets in contrast to district schools. Charter high schools exhibited higher average math scores in comparison to district high schools. For instance, Cabrera High School achieved an average math score of 83.06, whereas Bailey High School attained a score of 77.05. Upon observation, it became evident that whether attending a charter or district school, there was a little discrepancy in average reading scores, with both types of schools scoring above 80

percent. Overall, charter schools had higher performance rates than district schools. For example, Griffin High School(charter) boasted an impressive overall passing percentage of 90.60%, whereas Hernandez High School(district) lagged at 53.55%.

Highest-Performing Schools (by % Overall Passing)

I sorted the schools in the 'school_summary' DataFrame by their overall passing percentage in descending order and displayed the top 5 rows. The results were then saved in a new DataFrame named 'top_schools'. From this dataset, I noted the top-performing schools with a percentage overall passing were all charter Schools. For instance, Cabrera High School achieved an overall passing percentage of 91.33%, Thomas High School attained 90.95%, and so forth. I have included an image below.

	School Type	Total Students	Total school budget	Per student budget	Average math score	Average reading score	Percentage Passing math	Percentage Passing reading	Percentage Overall Passing
Cabrera High School	Charter	4761	\$1,081,356.00	\$582.00	83.06	83.98	94.13	97.04	91.33
Thomas High School	Charter	1468	\$1,043,130.00	\$638.00	83.42	83.85	93.27	97.31	90.95
Griffin High School	Charter	2949	\$917,500.00	\$625.00	83.35	83.82	93.39	97.14	90.60
Wilson High School	Charter	962	\$1,319,574.00	\$578.00	83.27	83.99	93.87	96.54	90.58
Pena High School	Charter	1800	\$585,858.00	\$609.00	83.84	84.04	94.59	95.95	90.54

Lowest-Performing Schools (by % Overall Passing)

Upon sorting the 'school_summary' in ascending order, I observed that the top 5 rows comprised district schools, all exhibiting lower percentages of overall passing For instance, Rodriguez High School achieved an overall passing percentage of 52.99%, Figueroa High School attained 53.20%, and so forth. I have included an image below.

	School Type	Total Students	Total school budget	Per student budget	Average math score	Average reading score	Percentage Passing math	Percentage Passing reading	Percentage Overall Passing
Rodriguez High School	District	1761	\$2,547,363.00	\$637.00	76.84	80.74	66.37	80.22	52.99
Figueroa High School	District	4635	\$1,884,411.00	\$639.00	76.71	81.16	65.99	80.74	53.20
Huang High School	District	2283	\$1,910,635.00	\$655.00	76.63	81.18	65.68	81.32	53.51
Hernandez High School	District	2917	\$3,022,020.00	\$652.00	77.29	80.93	66.75	80.86	53.53
Johnson High School	District	1858	\$3,094,650.00	\$650.00	77.07	80.97	66.06	81.22	53.54

Math Scores and Reading Scores by Grade

I computed the necessary computations to generate a DataFrame detailing the average math and reading scores for students across each grade level (9th, 10th, 11th, 12th) at each school. Reading scores were higher than average math scores. I have included both images below.

Average math score			Average reading score		
school_name	grade		school_name	grade	
Bailey High School	10th	77.00	Bailey High School	10th	80.91
	11th	77.52		11th	80.95
	12th	76.49		12th	80.91
	9th	77.08		9th	81.30
Cabrera High School	10th	83.15	Cabrera High School	10th	84.25
	11th	82.77		11th	83.79
	12th	83.28		12th	84.29
	9th	83.09		9th	83.68
Figueroa High School	10th	76.54	Figueroa High School	10th	81.41
	11th	76.88		11th	80.64
	12th	77.15		12th	81.38
	9th	76.40		9th	81.20
Ford High School	10th	77.67	Ford High School	10th	81.26
	11th	76.92		11th	80.40
	12th	76.18		12th	80.66
	9th	77.36		9th	80.63
Griffin High School	10th	84.23	Griffin High School	10th	83.71
	11th	83.84		11th	84.29
	12th	83.36		12th	84.01
	9th	82.04		9th	83.37
Hernandez High School	10th	77.34	Hernandez High School	10th	80.66
	11th	77.14		11th	81.40

Scores by School Spending

Using the previously established school_summary dataframe, I generated a table that dissects school performance according to average spending ranges per student. This was achieved by applying the provided code to define four bins with sensible cutoff values for categorizing school spending.

```
spending_bins = [0, 585, 630, 645, 680]
labels = ["<$585", "$585-630", "$630-645", "$645-680"]
```

Afterward, I utilized pd.cut to categorize spending into bins and employed the following code to compute mean scores per spending range.

```

spending_math_scores = school_spending_df.groupby(["Spending
Ranges (Per Student)"])[["Average Math Score"].mean()

spending_reading_scores = school_spending_df.groupby(["Spending
Ranges (Per Student)"])[["Average Reading Score"].mean()

spending_passing_math = school_spending_df.groupby(["Spending
Ranges (Per Student)"])[["% Passing Math"].mean()

spending_passing_reading = school_spending_df.groupby(["Spending
Ranges (Per Student)"])[["% Passing Reading"].mean()

overall_passing_spending = school_spending_df.groupby(["Spending
Ranges (Per Student)"])[["% Overall Passing"].mean()

```

Then, I incorporated the following metric to construct a DataFrame called `spending_summary`. From this dataset, it can be inferred that schools with lower spending ranges, characterized as charter schools, exhibited higher performance rates compared to those with higher spending ranges, namely districts. As demonstrated in the chart below, schools with a spending range of <\$585 achieved an overall passing percentage of 90.37%, while those with a range of \$586-\$630 attained 81.42%, and so forth. Hence, it is evident that schools with higher spending ranges yielded lower passing results.

Spending Ranges(Per Student)	Average math score	Average reading score	Percentage Passing math	Percentage Passing reading	Percentage Overall Passing
<\$585	83.45	83.93	93.46	96.61	90.37
\$585-630	81.90	83.16	87.13	92.72	81.42
\$630-645	78.52	81.62	73.48	84.39	62.86
\$645-680	77.00	81.03	66.16	81.13	53.53

Scores by School Type

Using the previously established `school_summary` dataframe, I generated a table that dissects school performance according to the total student ranges. This was achieved by applying the provided code to define four bins with sensible cutoff values for categorizing school spending.

```

size_bins = [0, 1000, 2000, 5000]
labels = ["Small (<1000)", "Medium (1000-2000)", "Large (2000-5000)"]

```

Afterward, I utilized `pd.cut` to categorize spending into bins and employed the following code to compute mean scores per student range.

```
#Create a DataFrame called size_summary that breaks down school performance
#based on school size (small, medium, or large).

size_math_scores = school_spending_df.groupby(["Total Student Ranges"])["Average math
score"].mean()

size_reading_scores = school_spending_df.groupby(["Total Student Ranges"])["Average reading
score"].mean()

size_passing_math = school_spending_df.groupby(["Total Student Ranges"])["Percentage
Passing math"].mean()

size_passing_reading = school_spending_df.groupby(["Total Student Ranges"])["Percentage
Passing reading"].mean()

overall_passing_size = school_spending_df.groupby(["Total Student Ranges"])["Percentage
Overall Passing"].mean()
```

Total Student Ranges	Average math score	Average reading score	Percentage Passing math	Percentage Passing reading	Percentage Overall Passing
Small (<1000)	83.82	83.93	93.55	96.10	89.89
Medium (1000-2000)	83.37	83.87	93.60	96.79	90.62
Large (2000-5000)	77.75	81.34	69.96	82.77	58.28

Consequently, a smaller student population tends to yield higher overall passing scores.