Analysing City School Data using Pandas and Jupyter Notebook

Given Background

You are the new Chief Data Scientist for your local government area. In this capacity, you'll be helping the school board and mayor make strategic decisions regarding future school budgets and priorities. As a first task, you've been asked to analyse the area-wide standardised test results. You'll be given access to every student's maths and reading scores, as well as various information on the schools they attend. Your task is to aggregate the data to showcase obvious trends in school performance.

Objectives

1. **To create a** summary report **that includes the data given below(1-9), prepared using Pandas and Jupyter Notebook.**
2. **To analyse the data to showcase obvious trends in school performance and add the** conclusions **to the report.**

Required Output Data includes:

1. Local Government Area (LGA) Summary
2. School Summary
3. Highest-Performing Schools (by % Overall Passing)
4. Lowest-Performing Schools (by % Overall Passing)
5. Maths Scores by Year
6. Reading Scores by Year
7. Scores by School Spending
8. Scores by School Size
9. Scores by School Type

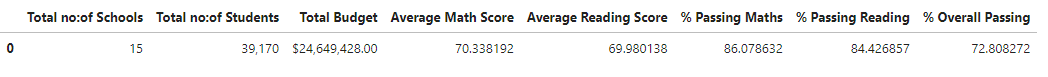
# SUMMARY REPORT

The objective is to analyse the trends in school performance based on various information derived from the given dataset. The dataset provides data including the details of students in different schools under the local self-government, which includes their student ID, some personal information including name and gender, math scores, reading scores, year/grade budget per student etc. The analysis was done to compare average math and reading scores, average math and reading pass per cent and overall pass per cent of students in different schools. The comparison was also made based on the type of schools (independent/ government), the size of the school and the average budget that was allocated per student.

1. Local Government Area (LGA) Summary

To create a high-level snapshot of the local government area's critical metrics in a DataFrame including the following:

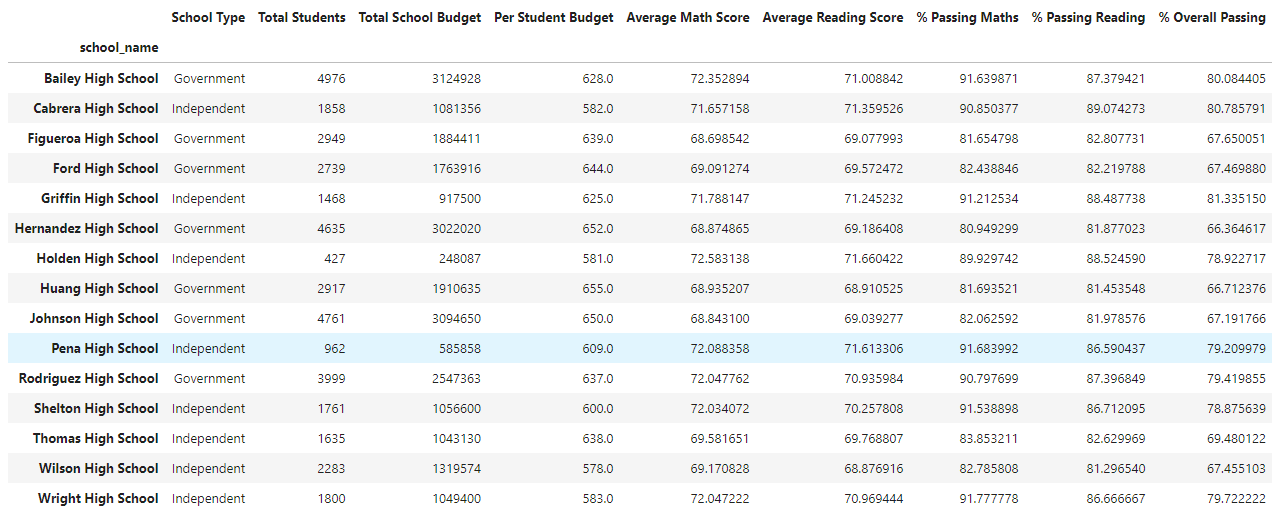
* Total number of unique schools
* Total Students
* Total budget
* Average maths score
* Average reading score
* % passing maths (the percentage of students who passed maths)
* % passing reading (the percentage of students who passed reading)
* % overall passing (the percentage of students who passed maths AND reading)



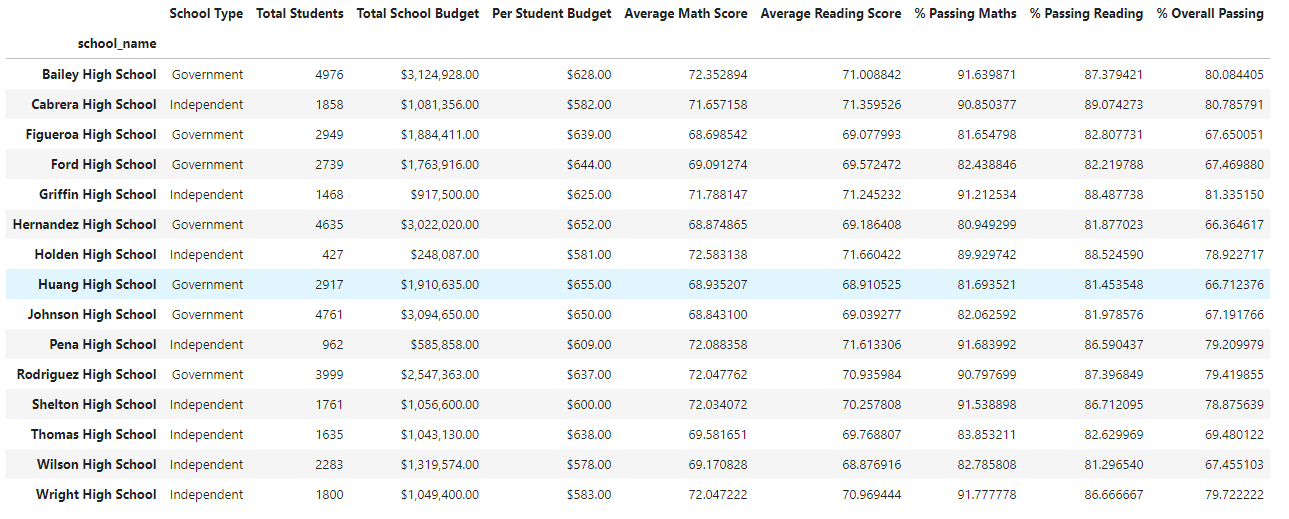
2. School Summary

To create a DataFrame that summarises key metrics about each school including the following:

* School name
* School type
* Total Students
* Total school budget
* Per student budget
* Average maths score
* Average reading score
* % passing maths (the percentage of students who passed maths)
* % passing reading (the percentage of students who passed reading)
* % overall passing (the percentage of students who passed maths AND reading)

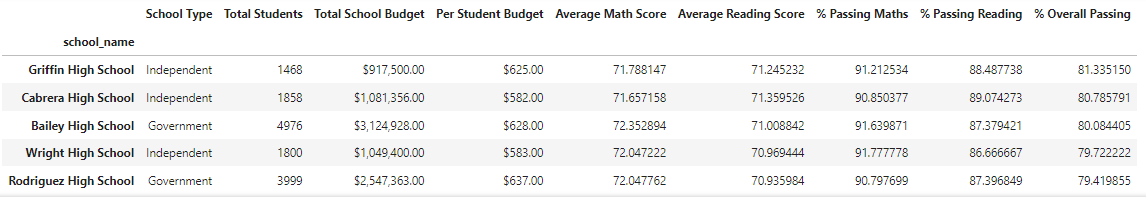


Formatted and cleaned:



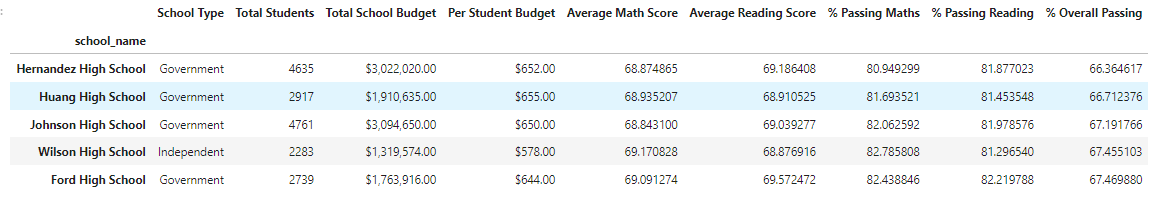
3. Highest-Performing Schools (by % Overall Passing)

To sort the schools by % Overall Passing in descending order and display the top 5 rows and save the results in a DataFrame called "top\_schools".



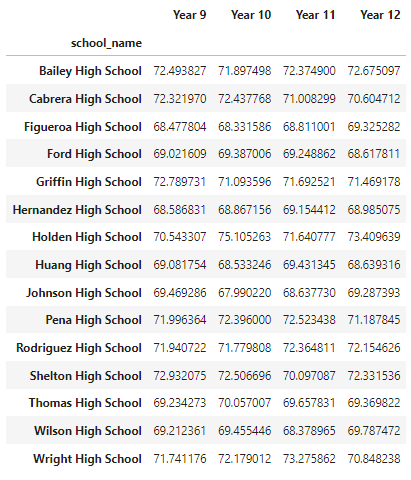
4. Lowest-Performing Schools (by % Overall Passing)

To sort the schools by % Overall Passing in ascending order and display the top 5 rows and save the results in a DataFrame called "bottom\_schools".

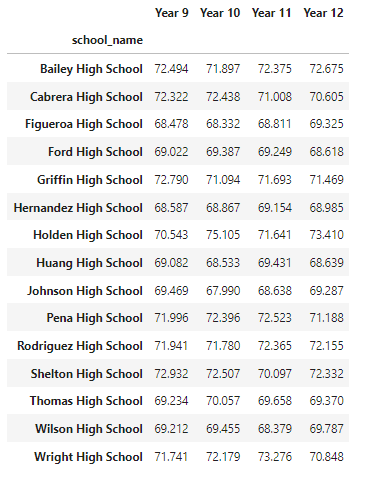


5. Maths Scores by Year

To create a dataframe that lists the average maths score for students of each year level (9, 10, 11, 12) at each school.

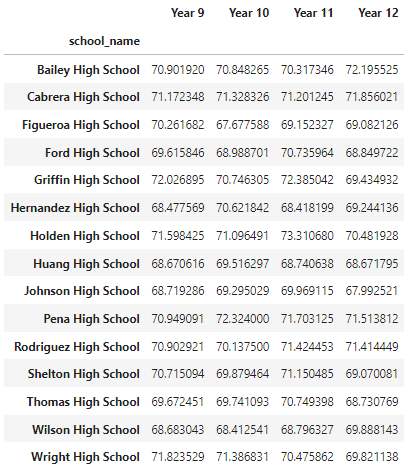


Formatted and cleaned:

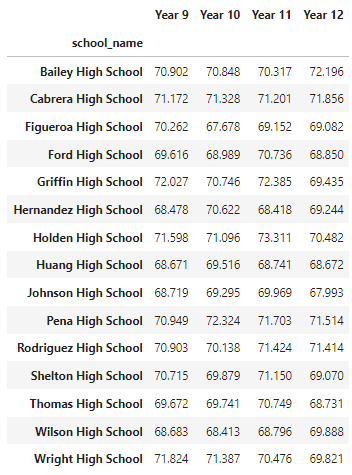


6. Reading Scores by Year

To create a DataFrame that lists the average reading score for students of each year level (9, 10, 11, 12) at each school.



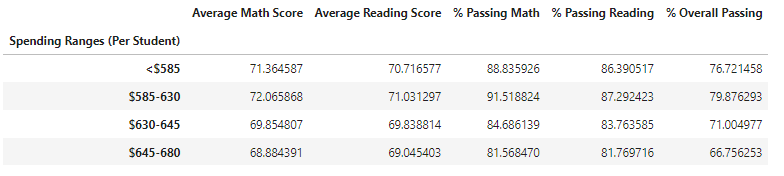
Formatted and cleaned:



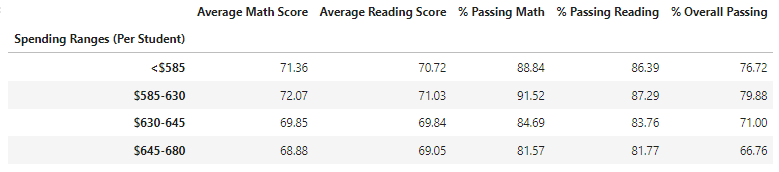
7. Scores by School Spending

To create a table that breaks down school performance based on average spending ranges (per student) and Includes the following metrics:

* Average maths score
* Average reading score
* % passing maths (the percentage of students who passed maths)
* % passing reading (the percentage of students who passed reading)
* % overall passing (the percentage of students who passed maths AND reading)



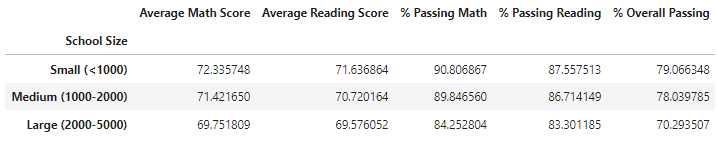
Formatted and cleaned:



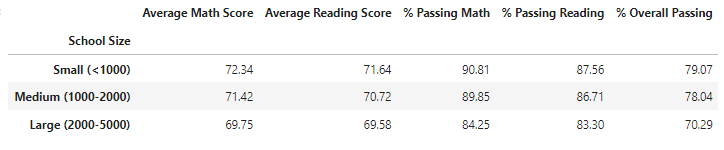
8. Scores by School Size

To create a DataFrame called size\_summary that breaks down school performance based on school size (small, medium, or large) and Includes the following metrics:

* Average maths score
* Average reading score
* % passing maths (the percentage of students who passed maths)
* % passing reading (the percentage of students who passed reading)
* % overall passing (the percentage of students who passed maths AND reading)



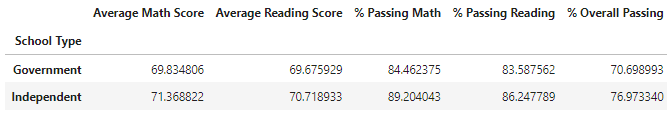
Formatted and cleaned:



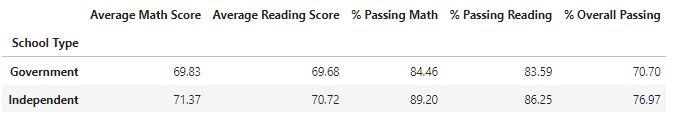
9. Scores by School Type

To create a new DataFrame called type\_summary that shows school performance based on the "School Type" and Includes the following metrics:

* Average maths score
* Average reading score
* % passing maths (the percentage of students who passed maths)
* % passing reading (the percentage of students who passed reading)
* % overall passing (the percentage of students who passed maths AND reading)

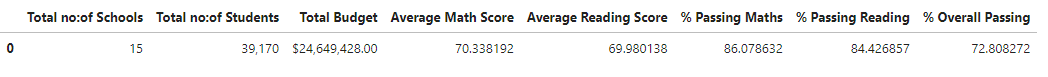


Formatted and cleaned:

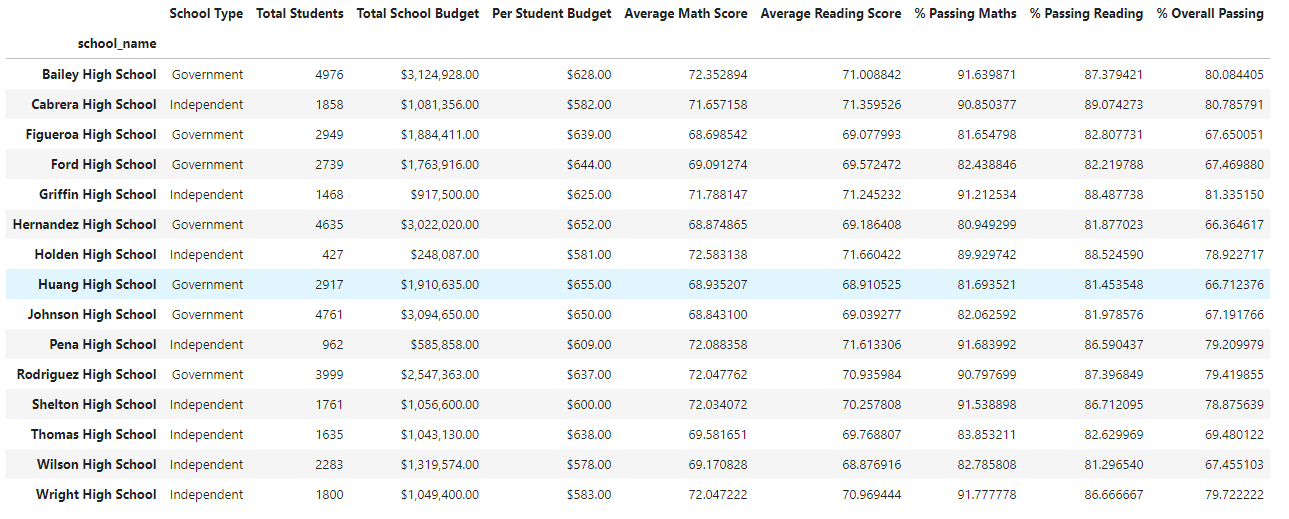


# CONCLUSIONS

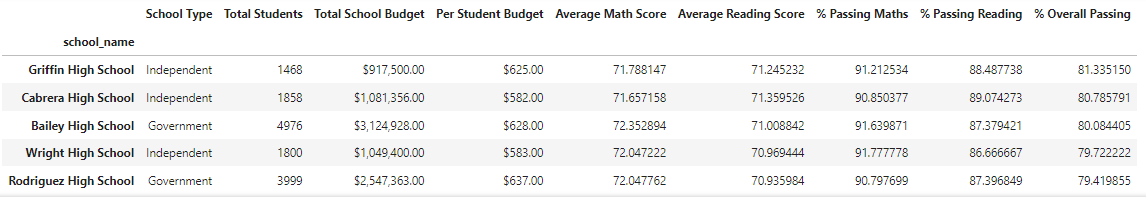
1. The number of students passing either maths or science is high (80+) compared to the students passing both (70+) when looking at the data for the whole local self-government jurisdiction area.



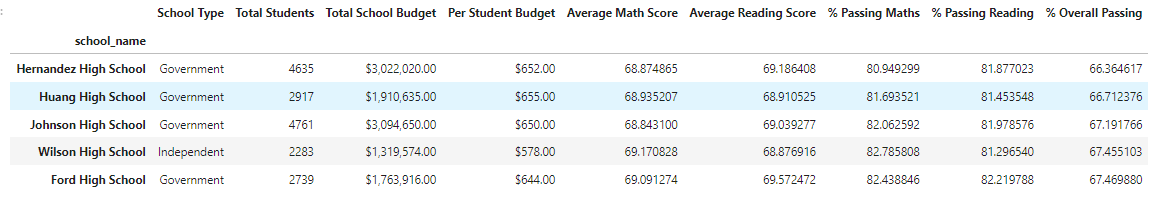
1. A similar trend is observed while analysing school-wise data. Looking at the math and reading pass per cent and overall pass per cent in different schools, The number of students passing either maths or science is high (80+) compared to the students passing both (70+) irrespective of them being small, large, independent or government-run.



1. There are more independent schools in the top 5 (best-performing schools) list than government schools and more government schools in the bottom 5 (worst-performing schools) when analysed based on overall pass per cent. And also government schools in both the top 5 and bottom 5 categories receive more funding than their independent counterparts.

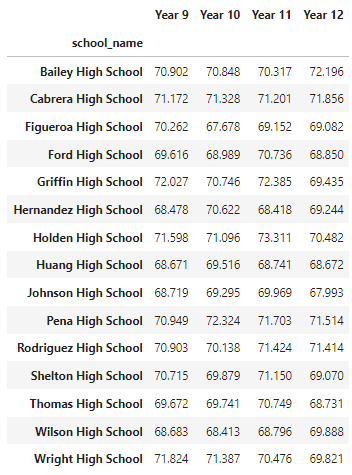
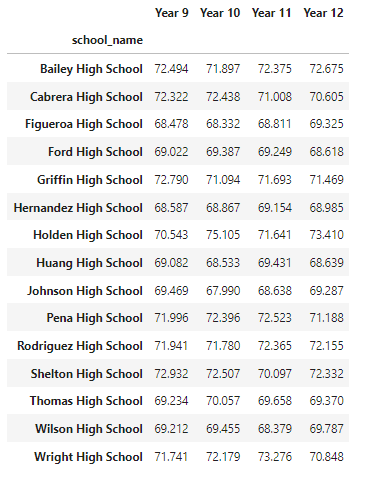
Top Performing Schools (By % Overall Passing)

Bottom Performing Schools (By % Overall Passing)

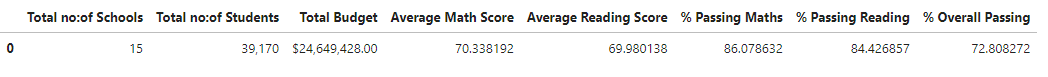


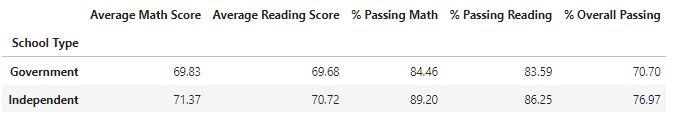
1. Reading and maths scores are almost similar among different class years.

Reading scores: Maths scores:

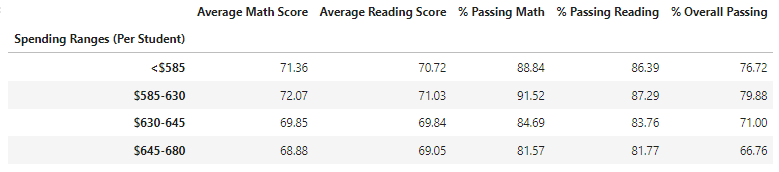


1. Comparing the average overall passing score of independent schools and government schools to the area's average, Independent schools have a higher overall passing rate of 76.97 compared to 70.70 and 72.80 for government schools and the area average.
2. Similarly, the performance of independent schools is above the area average and that of government schools is below the area average when analysing various matrices like average math and reading scores, and pass per cent in math and reading.

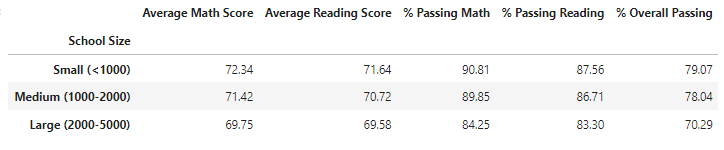




1. There appears to be no direct correlation between the overall passing rate and the average budget allocated per student.



1. The overall passing rate, the average math and reading scores and the average math and reading pass per cent decrease as the school size increase.



1. Independent schools are seen to perform better than government schools when analysed based on the average math and reading scores, average math and reading pass per cent and overall pass per cent.

