**Py City Schools Module 4 Challenge**

**Overview of the Project**

**Purpose:**

The purpose of this analysis is to assist Maria and the school board in utilizing student and school data to determine school sizes, budgets, and performance so they can use that information to make informed decisions regarding funding and resource allocation. Additionally, we were tasked with removing invalid data from Thomas High School and identifying any impact this had on the data.

**Results**

* **Generate the School District Summary**

Based on the analysis at the school district level, the school district appears to have a significantly fewer number of students passing reading than they are math. This should indicate to the school board that greater resources should be allocated to the hiring of math teachers, providing them and their students with adequate teaching/learning materials, and possibly investing in tutoring services for students who are struggling with course materials. See the image below for a breakdown of the school district data.

The removal of Thomas High School 9th graders did not have any significant impact on the school district summary from our analysis. This may likely be since Thomas High School only has 461 9th grade students that were removed from the overall analysis. This number is not too significant in the overall picture.

* **Generate the School Summary**

Breaking the data down by school we can see that some of the largest schools have the lowest overall passing grades in the district. This will become more apparent in the section covering Group Scores by School size. Overall, we see an overview of school sizes, budget per student and passing percentages by category. Thomas High School did see slightly lower passing grade percentages in all categories after 9th grade was removed, however, this difference is insignificant.

Table

Description automatically generated

* **High and Low Performing Schools**

Using the following code, we took a look at the highest and lowest performing schools from the dataset (ascending was set to True for the lowest schools): cleaned\_top\_schools = per\_school\_summary\_df.sort\_values(["% Overall Passing"], ascending=False). As we can see from the chart of highest performing schools, Calbrera, Thomas, Griffin, Wilson, and Pena high schools performed the best. The removed grades from Thomas High School did not impact its overall standing ion this list.

A screenshot of a computer

Description automatically generated

Below we can see the list of lowest performing high schools: Rodriguez, Figueroa, Huang, Hernandez, and Johnson. These schools had an overall passing grade almost 40% lower than those on top.

* **Average Math and Reading Scores by Grade**

In this first image we can see the average math scores by grade for each high school. Something that stands out from the module analysis is that the grade 9 math score for Thomas High School is listed as ‘NaN’. This is the same for the reading scores. We can see that, outside of a few outliers, students tend to perform at about the same level across grades in both math and reading. Table

Description automatically generated

Figure Math

Table

Description automatically generated

Figure Reading

* **Group Scores by School Spending per Student**

Looking at the budget per student we see some interesting results. It appears that schools with the highest spending per student perform worse in all categories when compared to schools with less spending per student. This could be due to a number of factors, but we do not see those from the analysis here. This may be a matter for the school board to investigate what these schools are spending their money on.

Table

Description automatically generated with medium confidence

* **Group Scores by School Size**

As previously mentioned, overall grades are worse in larger schools which is apparent from the following table. As we can see, schools grouped into the large category have significantly lower passing percentages than schools of small and medium size. This should indicate to the district that more staff, resources, and funding are needed in these schools to provide students with a better education.

Graphical user interface

Description automatically generated with medium confidence

This dataset was not impacted by the updated Thomas High School scores. Again, the school’s rather small number of impacted data points don’t sway the data too much.

* **Group Scores by School Type**

Here we see a breakdown of the data by school type. Charter schools perform significantly better than District schools. This may be due to a variety of factors including budget per student and school size.

Graphical user interface, table

Description automatically generated with medium confidence

**Summary**

The most noticeable change between the module and updated school district analysis can be seen in the reading and math scores by grade for each school. There we can see the ‘NaN’ value listed for Thomas HS grade 9. Another change could be seen in the school summary data. Thomas HS had slightly lower % averages for math, reading, and overall passing after 9th grade was removed. Another change was the total student count of the district after the Thomas HS 9th grade values had been removed. 39,170 students in total was reduced to 38,709. To do this I first had to isolate the 9th grade Thomas HS students using the loc method (see details here: <https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.loc.html>). From there I assigned the 9th grade students to a variable, generated a count, and subtracted that number from the total number of students. A final change that was noticed in the data was the average scores based on school size. A tenth of a percent difference can be noted in the overall passing score of medium schools between the original module version and the new version. This is because Thomas HS fell into the medium sized school category.