

1 PyCity Schools Analysis

- As a whole, schools with higher budgets, did not yield better test results. By contrast, schools with higher spending per student actually (\$645 - 675) underperformed compared to schools with smaller budgets (\$585 per student).
- As a whole, smaller and medium sized schools dramatically out-performed large sized schools on passing math performances (89-91% passing vs 67%).
- As a whole, charter schools out-performed the public district schools across all metrics. However, more analysis will be required to glean if the effect is due to school practices or the fact that charter schools tend to serve smaller student populations per school.

Note: Instructions have been included for each segment. You do not have to follow them exactly, but they are included to help you think through the steps.

```
In [ ]: # Dependencies and Setup
import pandas as pd
import numpy as np

# File to Load (Remember to Change These)
school_data_to_load = "data/schools_complete.csv"
student_data_to_load = "data/students_complete.csv"

# Read School and Student Data File and store into Pandas Data Frames
school_data = pd.read_csv(school_data_to_load)
student_data = pd.read_csv(student_data_to_load)

# Combine the data into a single dataset
school_data_complete = pd.merge(student_data, school_data, how="left", on=["school_id"])
school_data_complete.head()

# school_data_complete.count()
```

executed in 2.59s, finished 23:08:31 2022-08-05

1.1 District Summary

- Calculate the total number of schools
- Calculate the total number of students
- Calculate the total budget
- Calculate the average math score
- Calculate the average reading score
- Calculate the overall passing rate (overall average score), i.e. (avg. math score + avg. reading score)/2
- Calculate the percentage of students with a passing math score (70 or greater)
- Calculate the percentage of students with a passing reading score (70 or greater)
- Create a dataframe to hold the above results
- Optional: give the displayed data cleaner formatting

```
In [ ]: # Create a District Summary
```

```
In [ ]: # Total number of schools
```

```
In [ ]: # Total number of students
```

```
In [ ]: # Total budget
```

```
In [ ]: # Average math score
```

```
In [ ]: # Average reading score
```

```
In [ ]: # Overall average score
```

```
In [ ]: # Percentage of passing math (70 or greater)
```

1.2 School Summary

- Create an overview table that summarizes key metrics about each school, including:
 - School Name
 - School Type
 - Total Students
 - Total School Budget

- Per Student Budget
- Average Math Score
- Average Reading Score
- % Passing Math
- % Passing Reading
- Overall Passing Rate (Average of the above two)
- Create a dataframe to hold the above results

1.2.1 Top Performing Schools (By Passing Rate)

- Sort and display the top five schools in overall passing rate

```
In [ ]: # Sort and display the top five schools in overall passing rate
```

```
In [ ]: # Calculate total school budget
```

```
In [ ]: # Calculate per student budget
```

```
In [ ]: # Calculate the avg math and reading score
```

1.2.1.1 Find the passing rate for math and reading (above 70 points)

```
In [ ]: # Find the total counts of math result

# Find the counts for math result in each school that pass 70 or higher

# Calculate the math passing rate
```

```
In [ ]: # Find the total counts of read result

# Find the counts for read result in each school that pass 70 or higher

# Calculate the read passing rate
```

```
In [ ]: # Calculate the overall passing rate (average of the math and reading passing rate)
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

1.2.2 Bottom Performing Schools (By Passing Rate)

- Sort and display the five worst-performing schools

```
In [ ]: # Sort and display the worst five schools in overall passing rate
```

1.3 Math Scores by Grade

- Create a table that lists the average Reading Score for students of each grade level (9th, 10th, 11th, 12th) at each school.
 - Create a pandas series for each grade. Hint: use a conditional statement.
 - Group each series by school
 - Combine the series into a dataframe
 - Optional: give the displayed data cleaner formatting

```
In [ ]: # Create table that lists the average math score for each school of each grade level
```

```
In [ ]: # Calculate the average math score for 9th grade in each school
```

```
In [ ]: # Calculate the average math score for 10th grade in each school
```

```
In [ ]: # Calculate the average math score for 11th grade in each school
```

```
In [ ]: # Calculate the average math score for 12th grade in each school
```

1.3.1 Reading Score by Grade

- Perform the same operations as above for reading scores

```
In [ ]: # Create table that lists the average reading score for each school of each grade
```

```
In [ ]: # Calculate the average reading score for 9th grade in each school
```

```
In [ ]: # Calculate the average reading score for 10th grade in each school
```

```
In [ ]: # Calculate the average reading score for 11th grade in each school
```

```
In [ ]: # Calculate the average reading score for 12th grade in each school
```

1.4 Scores by School Spending

- Create a table that breaks down school performances based on average Spending Ranges (Per Student). Use 4 reasonable bins to group school spending. Include in the table each of the following:
 - Average Math Score
 - Average Reading Score
 - % Passing Math
 - % Passing Reading
 - Overall Passing Rate (Average of the above two)

```
In [ ]: # Sample bins. Feel free to create your own bins.  
spending_bins = [0, 585, 615, 645, 675]  
group_names = ["<$585", "$585-615", "$615-645", "$645-675"]
```

```
In [ ]: # Create a new column to show budget per student in each row
```

```
In [ ]: # Create a new column to define the spending ranges per student
```

```
In [ ]: # Calculate the average math score within each spending range
```

```
In [ ]: # Calculate the percentage passing rate for math in each spending range
```

```
In [ ]: # Calculate the percentage passing rate for reading in each spending range
```

```
In [ ]: # Calculate the percentage overall passing rate in each spending range
```

1.4.1 Scores by School Size

- Perform the same operations as above, based on school size.

```
In [ ]: # Sample bins. Feel free to create your own bins.  
size_bins = [0, 1000, 2000, 5000]  
group_names = ["Small (<1000)", "Medium (1000-2000)", "Large (2000-5000)"]
```

```
In [ ]: # Create a new column for the bin groups
```

Look for the total count of test scores that pass 70% or higher

```
In [ ]: # math_pass_size
```

```
In [ ]: # read_pass_size
```

```
In [ ]: # Calculate the overall passing rate for different school size
```

1.4.2 Scores by School Type

- Perform the same operations as above, based on school type.

```
In [ ]: # Create bins and groups, school type {'Charter', 'District'}
```

Find counts of the passing 70 or higher score for the both test

```
In [ ]: # math pass size
```

```
In [ ]: # reading pass size
```

In []: *# Calculate the overall passing rate*