PyCity Schools Analysis

- Analysis #1 :The highest performing schools are usually chater schools
- Analysis #2 : The bottom performing schools tend to be district schools
- Analysis #3 : The school size affect the students grades. The larger the school size the lower the students reading and math score. charter schools tend to be smaller and higher reading and math scores

---

```
# Dependencies and Setup
import pandas as pd
from pathlib import Path

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

# Read School and Student Data File and store into Pandas DataFrames
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",
school_data_complete.head()
```

Out[26]:

udent ID	student_name	gender	grade	school_name	reading_score	math_score	School ID	typ
0	Paul Bradley	M	9th	Huang High School	66	79	0	Distric
1	Victor Smith	M	12th	Huang High School	94	61	0	Distric
2	Kevin Rodriguez	M	12th	Huang High School	90	60	0	Distric
3	Dr. Richard Scott	M	12th	Huang High School	67	58	0	Distric
4	Bonnie Ray	F	9th	Huang High School	97	84	0	Distric

In [27]:

```
1 school_data.head()
```

Out[27]:

	School ID	school_name	type	size	budget
0	0	Huang High School	District	2917	1910635
1	1	Figueroa High School	District	2949	1884411
2	2	Shelton High School	Charter	1761	1056600
3	3	Hernandez High School	District	4635	3022020
4	4	Griffin High School	Charter	1468	917500

In [28]:
```
1 student_data.head()
```

Out[28]:

	Student ID	student_name	gender	grade	school_name	reading_score	math_score
0	0	Paul Bradley	M	9th	Huang High School	66	79
1	1	Victor Smith	M	12th	Huang High School	94	61
2	2	Kevin Rodriguez	M	12th	Huang High School	90	60
3	3	Dr. Richard Scott	M	12th	Huang High School	67	58
4	4	Bonnie Ray	F	9th	Huang High School	97	84

In [29]:
```
1 school_data_complete.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 39170 entries, 0 to 39169
Data columns (total 11 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   Student ID     39170 non-null  int64
 1   student_name   39170 non-null  object
 2   gender         39170 non-null  object
 3   grade          39170 non-null  object
 4   school_name    39170 non-null  object
 5   reading_score  39170 non-null  int64
 6   math_score     39170 non-null  int64
 7   School ID      39170 non-null  int64
 8   type           39170 non-null  object
 9   size           39170 non-null  int64
```

`

District Summary

In [30]:
```
# Calculate the total number of unique schools
school_count = school_data_complete.school_name.nunique()
school_count
```

Out[30]: 15

In [31]:
```
# Calculate the total number of students
student_count = student_data["Student ID"].count()
student_count
```

Out[31]: 39170

In [32]:
```
# Calculate the total budget
total_budget = school_data.budget.sum()
total_budget
```

Out[32]: 24649428

In [33]:
```
# Calculate the average (mean) math score
average_math_score = student_data["math_score"].mean()
average_math_score
```

Out[33]: 78.98537145774827

In [34]:
```
# Calculate the average (mean) reading score
average_reading_score = student_data["reading_score"].mean()
average_reading_score
```

Out[34]: 81.87784018381414

In [35]:

```
# Use the following to calculate the percentage of students who passed
passing_math_count = school_data_complete[(school_data_complete["math_
passing_math_percentage = passing_math_count / float(student_count) *
passing_math_percentage
```

Out[35]: 74.9808526933878

In [36]:

```
# Calculate the percentage of students who passed reading (hint: look
passing_reading_count = school_data_complete[(school_data_complete["re
passing_reading_percentage = passing_reading_count / float(student_cou
passing_reading_percentage
```

Out[36]: 85.80546336482001

In [37]:

```
# Use the following to calculate the percentage of students that passe
passing_math_reading_count = school_data_complete[
    (school_data_complete["math_score"] >= 70) & (school_data_complete
].count()["student_name"]
overall_passing_rate = passing_math_reading_count /  float(student_cou
overall_passing_rate
```

Out[37]: 65.17232575950983

In [38]:

```
# Create a high-level snapshot of the district's key metrics in a Data
district_summary = pd.DataFrame(
    {
        "Total Schools": [school_count],
        "Total Students": [student_count],
        "Total Budget": [total_budget],
        "Average Math Score": [average_math_score],
        "Average Reading Score": [average_reading_score],
        "% Passing Math": [passing_math_percentage],
        "% Passing Reading": [passing_reading_percentage],
        "% Overall Passing": [overall_passing_rate]
    }
)

# Formatting
district_summary["Total Students"] = district_summary["Total Students"
district_summary["Total Budget"] = district_summary["Total Budget"].ma

# Display the DataFrame
district_summary
```

Out[38]:

	Total Schools	Total Students	Total Budget	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	% Overall Passing
0	15	39,170	$24,649,428.00	78.985371	81.87784	74.980853	85.805463	65.172326

School Summary

In [39]:
```
# Use the code provided to select all of the school types
school_types = school_data.set_index(["school_name"])["type"]
school_types
```

Out[39]:
```
school_name
Huang High School        District
Figueroa High School     District
Shelton High School       Charter
Hernandez High School    District
Griffin High School       Charter
Wilson High School        Charter
Cabrera High School       Charter
Bailey High School       District
Holden High School        Charter
Pena High School          Charter
Wright High School        Charter
Rodriguez High School    District
Johnson High School      District
Ford High School         District
Thomas High School        Charter
Name: type, dtype: object
```

In [45]:
```
# Calculate the total student count per school
per_school_counts = school_data.set_index(["school_name"])["size"]
per_school_counts
```

Out[45]:
```
school_name
Huang High School        2917
Figueroa High School     2949
Shelton High School      1761
Hernandez High School    4635
```

In [46]:

```
# Calculate the total school budget and per capita spending per school
per_school_budget = school_data.set_index(["school_name"])["budget"]
per_school_budget
```

Out[46]:

```
school_name
Huang High School        1910635
Figueroa High School     1884411
Shelton High School      1056600
Hernandez High School    3022020
Griffin High School       917500
Wilson High School       1319574
Cabrera High School      1081356
Bailey High School       3124928
Holden High School        248087
Pena High School          585858
Wright High School       1049400
Rodriguez High School    2547363
Johnson High School      3094650
Ford High School         1763916
Thomas High School       1043130
Name: budget, dtype: int64
```

In [47]:

```
per_school_capita = per_school_budget/per_school_counts
per_school_capita
```

Out[47]:

```
school_name
Huang High School        655.0
Figueroa High School     639.0
Shelton High School      600.0
Hernandez High School    652.0
Griffin High School      625.0
Wilson High School       578.0
Cabrera High School      582.0
```

In [48]:

```
# Calculate the average test scores per school
per_school_math = school_data_complete.groupby("school_name")
per_school_math = per_school_math["math_score"].mean()
per_school_math
```

Out[48]:

```
school_name
Bailey High School       77.048432
Cabrera High School      83.061895
Figueroa High School     76.711767
Ford High School         77.102592
Griffin High School      83.351499
Hernandez High School    77.289752
Holden High School       83.803279
Huang High School        76.629414
Johnson High School      77.072464
Pena High School         83.839917
Rodriguez High School    76.842711
Shelton High School      83.359455
Thomas High School       83.418349
Wilson High School       83.274201
Wright High School       83.682222
Name: math_score, dtype: float64
```

In [49]:

```
per_school_reading = school_data_complete.groupby("school_name")
per_school_reading = per_school_reading["reading_score"].mean()
per_school_reading
```

Out[49]:

```
school_name
Bailey High School       81.033963
Cabrera High School      83.975780
Figueroa High School     81.158020
```

```
Name: reading_score, dtype: float64
```

In [55]:

```
# Calculate the number of students per school with math scores of 70 o
students_passing_math = school_data_complete.loc[school_data_complete.
#students_passing_math

school_students_passing_math = students_passing_math.groupby("school_r
school_students_passing_math
```

Out[55]:

```
school_name
Bailey High School       3318
Cabrera High School      1749
Figueroa High School     1946
Ford High School         1871
Griffin High School      1371
Hernandez High School    3094
Holden High School        395
Huang High School        1916
Johnson High School      3145
Pena High School          910
Rodriguez High School    2654
Shelton High School      1653
Thomas High School       1525
Wilson High School       2143
Wright High School       1680
dtype: int64
```

In [56]:

```
# Calculate the number of students per school with reading scores of 7
students_passing_reading = school_data_complete.loc[school_data_comple
#students_passing_reading

school_students_passing_reading = students_passing_reading.groupby("sc
school_students_passing_reading
```

In [57]:

```
# Use the provided code to calculate the number of students per school
students_passing_math_and_reading = school_data_complete.loc[
    (school_data_complete["reading_score"] >= 70) & (school_data_compl
]
school_students_passing_math_and_reading = students_passing_math_and_r
school_students_passing_math_and_reading
```

Out[57]:

```
school_name
Bailey High School       2719
Cabrera High School      1697
Figueroa High School     1569
Ford High School         1487
Griffin High School      1330
Hernandez High School    2481
Holden High School        381
Huang High School        1561
Johnson High School      2549
Pena High School          871
Rodriguez High School    2119
Shelton High School      1583
Thomas High School       1487
Wilson High School       2068
Wright High School       1626
dtype: int64
```

In [59]:

```
# Use the provided code to calculate the passing rates
per_school_passing_math = school_students_passing_math / per_school_
per_school_passing_reading = school_students_passing_reading / per_s
overall_passing_rate = school_students_passing_math_and_reading / pe
```

In [61]:

```
#to Check if count is correct
per_school_passing_math
```

Out[61]:

```
school_name
Bailey High School       66.680064
Cabrera High School      94.133477
Figueroa High School     65.988471
Ford High School         68.309602
Griffin High School      93.392371
Hernandez High School    66.752967
Holden High School       92.505855
Huang High School        65.683922
Johnson High School      66.057551
Pena High School         94.594595
Rodriguez High School    66.366592
Shelton High School      93.867121
Thomas High School       93.272171
Wilson High School       93.867718
Wright High School       93.333333
dtype: float64
```