PROJECT: EXPLORING NYC PUBLIC SCHOOL TEST RESULT SCORES

Photo by Jannis Lucas on Unsplash.

Every year, American high school students take SATs, which are standardized tests intended to measure literacy, numeracy, and writing skills. There are three sections - reading, math, and writing, each with a **maximum score of 800 points**. These tests are extremely important for students and colleges, as they play a pivotal role in the admissions process.

Analyzing the performance of schools is important for a variety of stakeholders, including policy and education professionals, researchers, government, and even parents considering which school their children should attend.

You have been provided with a dataset called `schools.csv`, which is previewed below.

You have been tasked with answering three key questions about New York City (NYC) public school SAT performance.

```
# Re-run this cell
import pandas as pd

# Read in the data
schools = pd.read_csv("schools.csv")

# Preview the data
print(schools.head(3))

# QUESTION = Which NYC schools have the best math results?
# Sort the schools by average_math in descending order and get the school names
df = pd.DataFrame(schools)
best_math_schools = df[df["average_math"] >= 0.8 * 800]
best_math_schools = best_math_schools[["school_name",
"average_math"]].sort_values(by="average_math", ascending=False)

#NYC schools that have the best math results
print(best_math_schools)

# QUESTION = What are the top 7 performing schools based on the combined SAT scores?
# Sum across colums to get the total SAT scores
schools["total_SAT"]=
df[["average_math","average_reading","average_writing"]].sum(axis=1)

# Top ten Schools
top_10_schools = schools[["school_name", "total_SAT"]].sort_values(by="total_SAT",
ascending=False).head(7)
print(top_10_schools)

#QUESTION = Which single borough has the largest standard deviation in the combined
SAT score?
# Group by borough and calculate the standard deviation of total_SAT
grouped = df.groupby("borough").agg(
    num_schools=("school_name", "count"),
    average_SAT=("total_SAT", "mean"),
    std_SAT=("total_SAT", "std")
).reset_index()

# Get borough with highest standard deviation in total_SAT
largest_std_dev = grouped.sort_values(by="std_SAT",
ascending=False).round(2).head(1).reset_index(drop=True)

print(largest_std_dev)
```

```
                                     school_name  ... percent_tested
0  New Explorations into Science, Technology and ...  ...            NaN
1                              Essex Street Academy  ...           78.9
```

```
2                      Lower Manhattan Arts Academy  ...           65.1

[3 rows x 7 columns]
                                        school_name  average_math
88                            Stuyvesant High School           754
170                     Bronx High School of Science           714
93                Staten Island Technical High School           711
365  Queens High School for the Sciences at York Co...           701
68   High School for Mathematics, Science, and Engi...           683
280                    Brooklyn Technical High School           682
333                      Townsend Harris High School           680
174  High School of American Studies at Lehman College           669
0    New Explorations into Science, Technology and ...           657
45                      Eleanor Roosevelt High School           641
                                        school_name  total_SAT
88                            Stuyvesant High School       2144
170                     Bronx High School of Science       2041
93                Staten Island Technical High School       2041
174  High School of American Studies at Lehman College       2013
333                      Townsend Harris High School       1981
365  Queens High School for the Sciences at York Co...       1947
5                      Bard High School Early College       1914
     borough  num_schools  average_SAT  std_SAT
0  Manhattan           89      1340.13   230.29
```