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.

```
# import pandas
import pandas as pd

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

# Preview the data
schools.head()

# Question 1: Which NYC schools have the best math results? The best math results
are at least 80% of the *maximum possible score of 800* for math. Save your results
in a pandas DataFrame called best_math_schools, including "school_name" and
"average_math" columns, sorted by "average_math" in descending order.
schools['math_pct']= schools['average_math']/800 #created column that seperated the
math scores and made them a decimal percentage (0-1)
best_math_schools_1 = schools.sort_values('average_math', ascending= False) #ordered
the values of teh avergae math scores in deceding order (highest to lowest)
best_math_schools_2 = best_math_schools_1[best_math_schools_1['math_pct']>=0.80]
#created a subset where only the data of the schools who's average was 0.8 (80%) or
greater was included
best_math_schools = best_math_schools_2.loc[:, ["school_name","average_math"]]
#ceated dataset that included all schools with equal or highter than 80% on average
math and selected teh coliumns i wanted to be shown
print(best_math_schools)

# Question 2: What are the top 10 performing schools based on the combined SAT
scores? Save your results as a pandas DataFrame called top_10_schools containing the
"school_name" and a new column named "total_SAT", with results ordered by
"total_SAT" in descending order ("total_SAT" being the sum of math, reading, and
writing scores).
schools["total_SAT"]=
schools['average_math']+schools['average_reading']+schools['average_writing'] #added
the three score averages for each school to create a new row that showed the overal
total SAT average for each.
top_10_schools = schools.sort_values("total_SAT", ascending=False)
[["school_name","total_SAT"]].head(10) #last ordered the schools in descending order
of their total SAT and selected the top 10 schools using .head. Made sure to only
include columns I wanted in the argument after sort_values
print(top_10_schools)

#Question 3: Which single borough has the largest standard deviation in the combined
SAT score?Save your results as a pandas DataFrame called largest_std_dev.The
DataFrame should contain one row, with: "borough" - the name of the NYC borough with
the largest standard deviation of "total_SAT". "num_schools" - the number of schools
in the borough. "average_SAT" - the mean of "total_SAT". "std_SAT" - the standard
deviation of "total_SAT". Round all numeric values to two decimal places.
```

```
schools['total_SAT']=
schools['average_math']+schools['average_reading']+schools['average_writing'] #added
average score of three scores to get total SAT of each school
boroughs = schools.groupby("borough")["total_SAT"].agg(["count", "mean",
"std"]).round(2) #grouped the data by boroughs and selected the count, mean, and
standard deviation for each boroughs Total SAT
largest_std_dev = boroughs[boroughs["std"] == boroughs["std"].max()] #created data
set and selected the borough with the highest standard deviation. Dataset includes
only the borough name, count, mean and std of the borough with the highest std
largest_std_dev = largest_std_dev.rename(columns={"count": "num_schools", "mean":
"average_SAT", "std": "std_SAT"}) #used rename function to change the name of each
column to the requested names.
print(largest_std_dev)
```

```
                                        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
280                    Brooklyn Technical High School       1896
45                      Eleanor Roosevelt High School       1889
68   High School for Mathematics, Science, and Engi...       1889
           num_schools  average_SAT  std_SAT
borough
Manhattan           89      1340.13   230.29
```