Analyzing NYC Public School SAT Scores: Insights for Policy Makers, Researchers, and Parents

Overview

Every year, school test results impact the college admissions fate of millions of students. In this project, standardized test performance data from NY C's public schools will be used to identify the schools with top math results, analyze how performance varies by borough, and find the city's top ten performing schools!

Skills involved: data cleaning, data analysis, data visualization, and insights generation.

Problem Statement

This project aims to properly answer these questions:

a) Which NYC schools have the best math results?

b) What are the top 10 performing schools based on combined SAT scores?

c) Which borough has the largest standard deviation in SAT scores, and what insights can we derive from it?

Dataset

File name: schools.csv

Description: Includes SAT performance data of NYC public schools (as image below).

	school_name	borough	building_code	average_math	average_reading	average_writing	percent_tested
0	New Explorations into Science, Technology and ...	Manhattan	M022	657	601	601	NaN
1	Essex Street Academy	Manhattan	M445	395	411	387	78.9
2	Lower Manhattan Arts Academy	Manhattan	M445	418	428	415	65.1
3	High School for Dual Language and Asian Studies	Manhattan	M445	613	453	463	95.9
4	Henry Street School for International Studies	Manhattan	M056	410	406	381	59.7

Source: Publicly available.

Size and structure: Share the number of rows, columns, and any key details about the data.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 375 entries, 0 to 374
Data columns (total 7 columns):
 #   Column           Non-Null Count   Dtype
---  ------           --------------   -----
 0   school_name      375 non-null     object
 1   borough          375 non-null     object
 2   building_code    375 non-null     object
 3   average_math     375 non-null     int64
 4   average_reading  375 non-null     int64
 5   average_writing  375 non-null     int64
 6   percent_tested   355 non-null     float64
dtypes: float64(1), int64(3), object(3)
memory usage: 20.6+ KB
None
```

Approach and Methodology

Data Cleaning:

The data in the csv file was already clean and fully populated.

Exploratory Data Analysis (EDA):

To better understand the data some methods were applied to the data frame. Subsetting and sorting were essential to manipulate data and extract the information needed.

Analysis:

The results were obtained from the data frame manipulation, using methods like groupby() and sort_values() as well as visualization methods like .iloc[].

The combined SAT Score and thus the Top 10 Schools were calculated through the sum() method. The mean() and the standard deviation - std_dev() - were used in order to calculate the Borough with the largest Standard Deviation.

Insights:

The last question opens up space for discussions on how to reduce the discrepancy between school scores, which reflect the quality of education.

Results

Provide summaries and key findings for each question:

Best Math Results: List of schools that meet the criteria, with their math scores.

```
                                           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
```

Top 10 Schools by Combined SAT: The top 10 schools and their combined SAT scores.

```
                                           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
```

Borough with Largest Standard Deviation: Name of the borough, number of schools, average SAT score, and standard deviation.

```
            num_schools  average_SAT  std_SAT  total_SAT
borough
Manhattan            89      1340.13   230.29     119272
```

Technologies Used

Languages: Python

IDE: Jupyter Notebook on Google Colab

Libraries: Pandas, NumPy and Matplotlib.

GitHub repo: https://github.com/gustavo-plc/NYC-Schools

Skills Demonstrated

Data preprocessing.

Exploratory data analysis and visualization.

Statistical analysis.

Problem-solving and insights generation.

Reproducibility with clean code and well-documented processes.

Future Improvements

Additional datasets, like schools budget and academic background of teachers, can be combined with the existing data to generate further analysis and insights.

Machine learning could predict future test performance, classify schools by their results to identify the top ten, and analyze performance variations across boroughs to identify influential factors.

Interactive visualizations or dashboards to improve visualization.

Contact

Feel free to reach out to me through my LinkedIn profile or GitHub repository for any inquiries, feedback, or further discussion.

LinkedIn: https://www.linkedin.com/in/gustavo-plc/

GitHub: https://github.com/gustavo-plc

I welcome the opportunity to connect with recruiters and others in the industry to discuss potential opportunities or to receive constructive feedback. Let's stay in touch!