The Influence of Virtual Tools on Assessing Education Progress: A Detailed Study in NYC High Schools.

Adeola Odunewu

Introduction

The role of education is of utmost importance in the growth and development of individuals and society at large. It is necessary that a well-structured and adaptable system that caters to the evolving needs of society is put in place. To ensure and improve the quality of education, as well as comprehend the factors that affect academic performance, the analysis of datasets is indispensable. This study delve into the variations in performance among public high schools in New York, specifically focusing on the influence of enrollment on SAT examination results. Furthermore, it seeks to examine whether factors such as the quality of learning environment have an impact on academic achievement. To gain a comprehensive understanding of the data, an interactive virtual analysis system will be utilized, harnessing both human cognitive strengths and computational analysis techniques.

Understanding the relationship between student enrollment and academic performance is crucial for shaping effective educational policies and practices (Heer, J., & Shneiderman, B. 2012). This study specifically aims to investigate if schools with lower student enrollment exhibit higher academic performance compared to those with larger populations. The academic success of students is often assessed through the SAT, a widely recognized measure of academic aptitude that significantly influences college admissions. By examining public high schools with the highest cumulative SAT scores, this study seeks to gain valuable insights into the factors contributing to their academic excellence. Through an analysis of both these high-performing schools and the correlation between student enrollment and academic performance, a comprehensive understanding of the dynamics and characteristics associated with top-performing public high schools will be achieved. These design choices, combined with the interactivity and intuitive layout of the Tableau dashboard, enhance the user's ability to explore the data and find answers to the research questions (Mackinlay, J. 2023). The visuals are accompanied by relevant labels, color choices, and appropriate legends, following best practices for clear communication and interpretation of the information. Additionally, the use of filters and calculated fields in the dashboard allows users to dynamically adjust the view and focus on specific subsets of the data, aiding in a more comprehensive understanding of the research findings.

The findings from this research endeavor will encourage collaboration among educators, administrators, and policymakers, empowering them to enhance educational strategies, allocate resources efficiently, and ultimately improve academic outcomes for students. Additionally, to facilitate this analysis, the study will utilize Tableau, a powerful data visualization tool, to present the insights in a clear and accessible manner. This understanding will enable educators and policymakers to make informed decisions regarding school size and resource allocation, ultimately optimizing student achievement.

Data

This dataset consists of a row for every accredited high school in New York City with its department ID number, school name, borough, building code, street address, latitude/longitude coordinates, phone number, start and end times, student enrollment withrace breakdown, and average scores on each SAT test section for the 2014-2015 school year.

Reference: https://www.kaggle.com/datasets/nycopendata/high-schools>

Metadata

	Name	Description	Domain
	SCHOOL ID	School identification	Nominal
	SCHOOL NAME	Name of the school	Nominal
	BOROUGH	Town within the Local Government	Nominal
	BUILDING CODE	Building Unique Number	Integer
	STREET ADDRESS	Street Address of the School	Nominal
	CITY	Name of the city	Nominal
	STATE	Name of the state	Nominal
	ZIP CODE	Area code	Integer
	LATITUDE	Coordinate of the Schools	Interval
	LONGITUDE	Coordinate of the Schools	Interval
	PHONE NUMBER	Telephone Number	Integer
	START TIME	Examination Starting time	Nominal
	END TIME	Examination ending time	Nominal
	STUDENT ENROLMENT	Numbers of student enrol in each School	Integer
	PERCENT WHITE	Percentage of white student	Integer(%)
	PERCENTAGE BLACK	Percentage of Black student	Integer(%)
	PECENTAGE HISPANIC	Percentage of Hispanic student	Integer(%)
	PECENTAGE ASIAN	Percentage of Asian	Integer(%)

		student	
	AVERAGE SCORE (SAT MATH)	Examination score Math	Ordinal
	AVERAGE SCORE (SAT READING)	Examination score Reading	Ordinal
	AVERAGE SCORE (SAT WRITING)	Examination score Writing	Ordinal
	PERCENTAGE TESTED	Percentage of public high school students that participate in the test.	Integer(%)

Calculated Field/Data

	Name	Description	Domain
	NUMBER OF SCHOOL	The length of the overall schools name was calculated	Integer
	SAT SCORE	Math, Reading, and Writing SAT score were sum up to give overall score	Integer
	CUMMULATIVE SAT SCORES	Total combined scores obtained by a student across all sections of the SAT	Integer

Problem Description and Research Question

This study aims to investigate two key aspects related to public high schools and academic performance. Firstly, it seeks to identify public high schools with the highest cumulative SAT scores and analyze the characteristics and factors associated with their academic excellence. Secondly, the study examines the relationship between student enrollment and academic performance, exploring whether schools with lower student enrollment exhibit better academic outcomes compared to those with larger populations. By addressing these questions, the study aims to provide valuable insights into the determinants of academic success and inform educational strategies for improving overall academic outcomes in public high schools.

1. The public high school's with the highest SAT score.
To answer the above question this study will explore;
I. The public high schools with the highest cumulative SAT scores.
II. The group of students with the highest cumulative SAT percentage.

2. Do schools with lower student enrollment perform better.
To answer the above question this study will explore;
I. What is the aggregate performance for each Borough
II. What is the relationship between student enrollment and academic performance variations.

Requirement

The response to the first question there will be a need to clean the data before going ahead with visualising the variables.

To clean the SAT_scores dataset using R, we will perform the following data preprocessing steps:
Cleaning missing values: Any missing values in the SAT score dataset will be addressed.
Scaling missing columns: If any columns contain missing values, we will scale the data appropriately.
Handling duplicates: Any duplicate entries in the dataset will be handled.
Handling categorical variables: Categorical variables in the SAT_score dataset will be processed accordingly.

Implementation

Question One

The public high schools with the highest cumulative SAT scores.

To calculate the cumulative SAT score for each school, a calculated field will be created. This can be achieved by multiplying the average SAT score by the number of students who took the SAT.

To create the calculated field, follow these steps:
1. Click on the tab located on the left, close to the search bar.
2. In the drop-down menu, select "Create Calculated Field" to open the calculated field editor.
3. Enter a name for the calculated field, e.g., "Cumulative_SAT_scores".
4. In the formula field, enter the calculation: [Average SAT score] * [Number of students who took the SAT].
5. Ensure that the calculation is aggregated as desired. To do this, click on the drop-down arrow next to the "SAT score" field in the calculated field editor, and select "Aggregation" > "Average".
6. Click "OK" to create the calculated field.

Next, arrange the fields in the Tableau worksheet as follows:
- Drag the "School.Names" field to the Columns shelf.
- Drag the "Cumulative_SAT_scores" field to the Rows shelf.
- Choose the "Vertical Bars" option to display the plot on Tableau.

To sort the bars in descending order based on the cumulative SAT scores, follow these steps:
1. Click on the drop-down arrow on the Cumulative_SAT_scores axis to open the options menu.
2. Select "Measure" and choose "Sum".
3. In the top panel, select "Descending" to sort the bars in descending order based on the cumulative SAT scores.

The group of students from public high schools who achieved the highest overall SAT scores.

1. Right-click on the fields "Percentage.Black," "Percentage.White," "Percentage.Hispanic," and "Percentage.Asian" one by one.
2. Select "Convert to Dimension" for each of these fields.
3. Drag the fields "Percentage.Black," "Percentage.White," "Percentage.Hispanic," and "Percentage.Asian" to the Columns shelf.
4. Drag the "Cumulative_SAT_scores" field to the Rows shelf.
5. Apply a filter on the "Sat_Scores" field to only include scores above or below a definite threshold, as required.
6. Drag the "School.Name" field to the Detail shelf on Marks.
7. Choose the "Side-by-Side Bars" option to display the plot in Tableau.

Question Two.

What is the aggregate performance for each Borough.

1. Drag the "Borough" field to Marks and assign it to the Label.
2. Drag the "Sat_score" field to Marks and assign it to the Color.
3. Drag the "Sat_scores" field to Size.
4. Choose the "Tree Map" option to display the plot in Tableau.
5. Drag the "Cumulative_Sat_scores" and "Number_of_School" fields to Label.

What is the relationship between student enrollment and academic performance variations.

1. Drag the "Longitude" field to the Columns shelf.
2. Drag the "Latitude" field to the Rows shelf.
3. Drag the "City" field to the Marks shelf, and assign it to both the Color and Label.
4. Drag the "Borough" field to the Marks shelf and assign it to Label.
5. Calculate the "Number of school" by using the COUNT() function with the "School.Name" field. This will return the count of schools.
6. Drag the calculated field "Number of school" to the Label.
7. Choose the "Symbol Map" option to display the plot in Tableau.

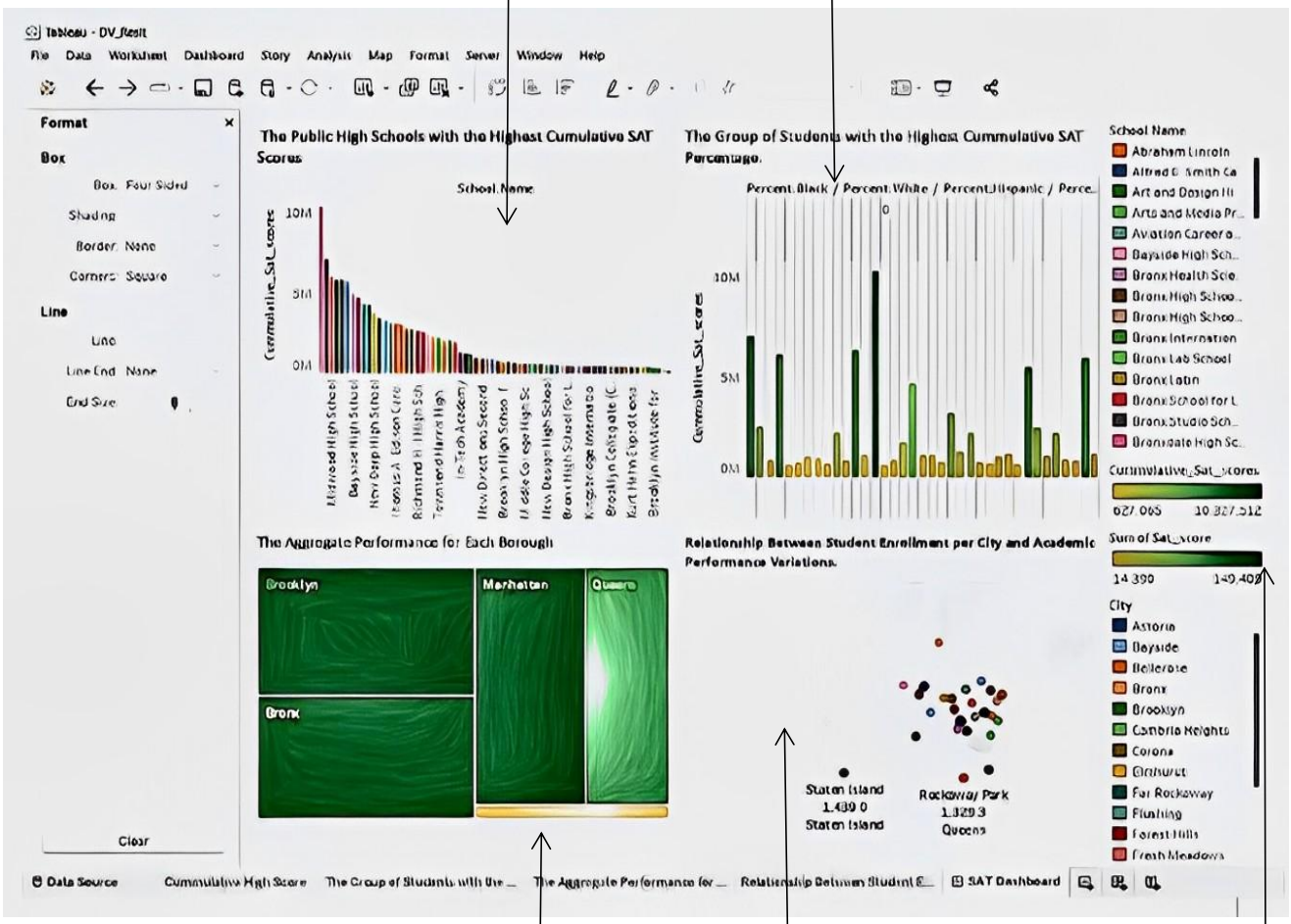
These following steps involves modify the visualization in Tableau to include the specified fields, arrange them in the appropriate shelves, and calculate the number of schools using the COUNT() function. This will enable us to display the data effectively with longitude and latitude as the spatial coordinates and the City, Borough, and Number of school as the visual attributes.

Paper Landscape

Objective: this study objective is to provide insights into the characteristics and factors associated with high SAT scores in public high schools and investigate the relationship between student enrollment and academic performance. The findings will help inform educational strategies, resource allocation, and decision-making processes to enhance academic outcomes for students (Spence, R. 2014).

Vertical Bars is use for display, calculated field is created to calculate the cumulative SAT score for each. Q1(i) .

The group of students from public high schools who achieved the highest overall SAT scores will be visualize using Side-by-side Bars
Q1 (ii)



The aggregate performance for each Borough, Tree Map is used to display the plot. Q2 (i)

The relationship between student enrollment and academic performance variations is visualized using Symbol Map. Q2(ii)

Global filter details
This indicate the color and the cleaning of unwanted noise.

Dashboard

Result

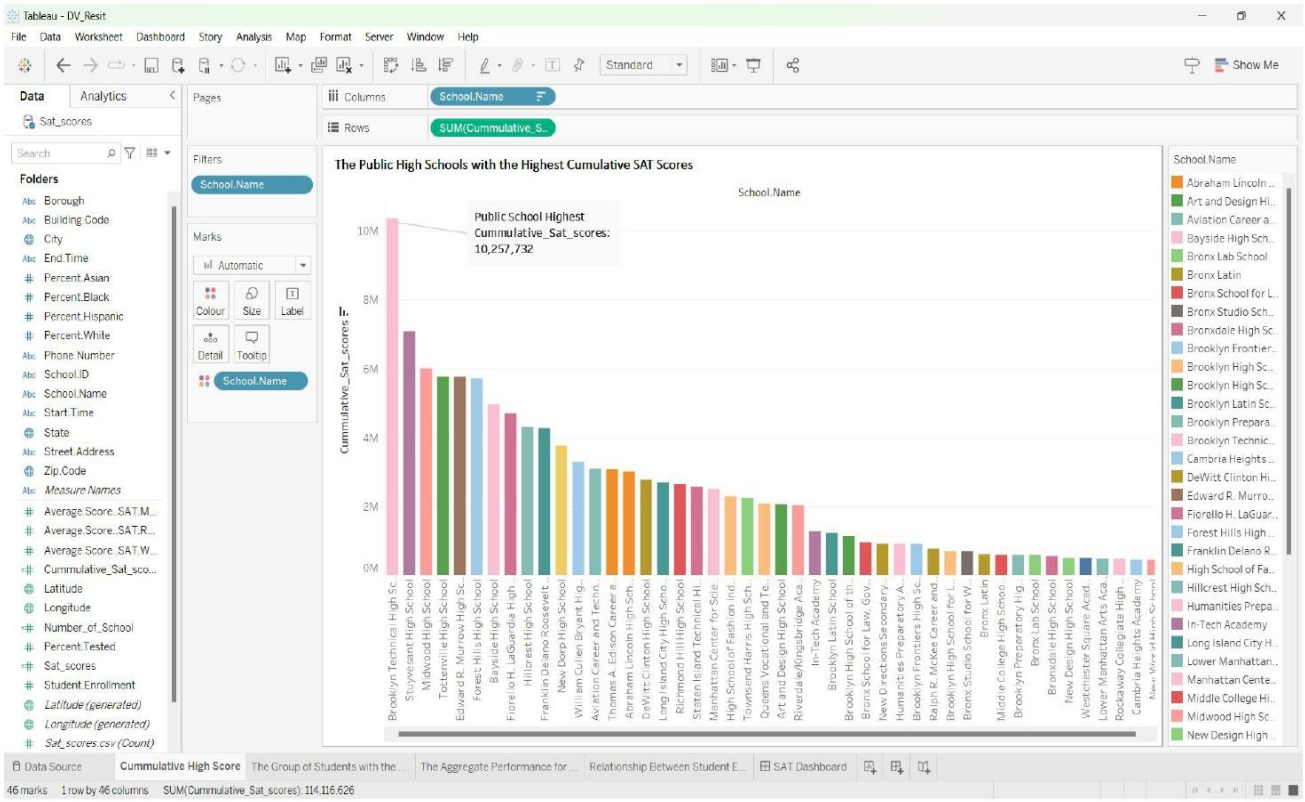
The user will be able to visualize to gain insights into the characteristics and factors associated with high SAT scores in public high schools and investigate the relationship between student enrollment and academic performance. across the city and Borough with the period of 2014-2015.

Implementation

Sheet One

To determine the cumulative SAT score average SAT score of a schools is multiply by the total number of students who took the SAT at that school. By considering the case of Brooklyn Technical High School, the calculated cumulative SAT_score is 10,257,732. The resulting value represents the cumulative SAT score for Brooklyn Technical High School, indicating the overall academic performance of the students who took the SAT at that institution. By calculating the cumulative SAT score for each school in a similar manner, we can compare and evaluate their academic performance relative to one another. This information provides insights into the overall proficiency and achievement levels of students in different schools based on their SAT scores.

Note: Essential component or principal component depict schools that participate in the test



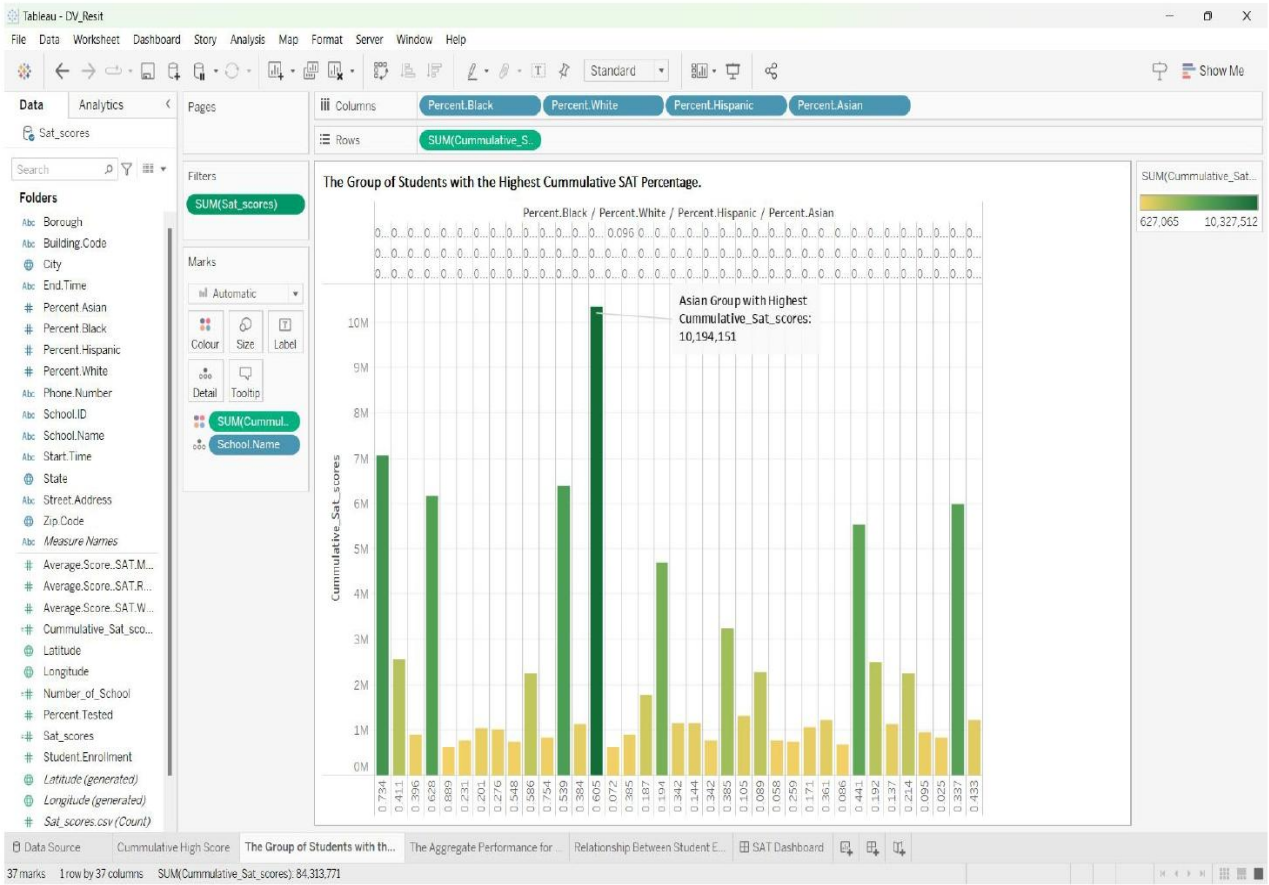
Vertical-Bars for School names and aggregate score

Sheet Two

To determine the group of students with the highest cumulative SAT Percentage, the visualization compares the percentages of different groups (Black, White, Hispanic, Asian) in relation to their cumulative SAT scores. By converting the percentage fields to dimensions and placing them on the Columns shelf, Each group's percentage is represented by a separate column in the visualization. In this context, the calculation suggests that the Asian group has the highest cumulative percentage among the racial or ethnic groups being compared in the visualization.

However, the cumulative percentage is calculated as the proportion of cumulative SAT scores contributed by each group. The "Cumulative_SAT_scores" field represents the total cumulative SAT scores for each school.

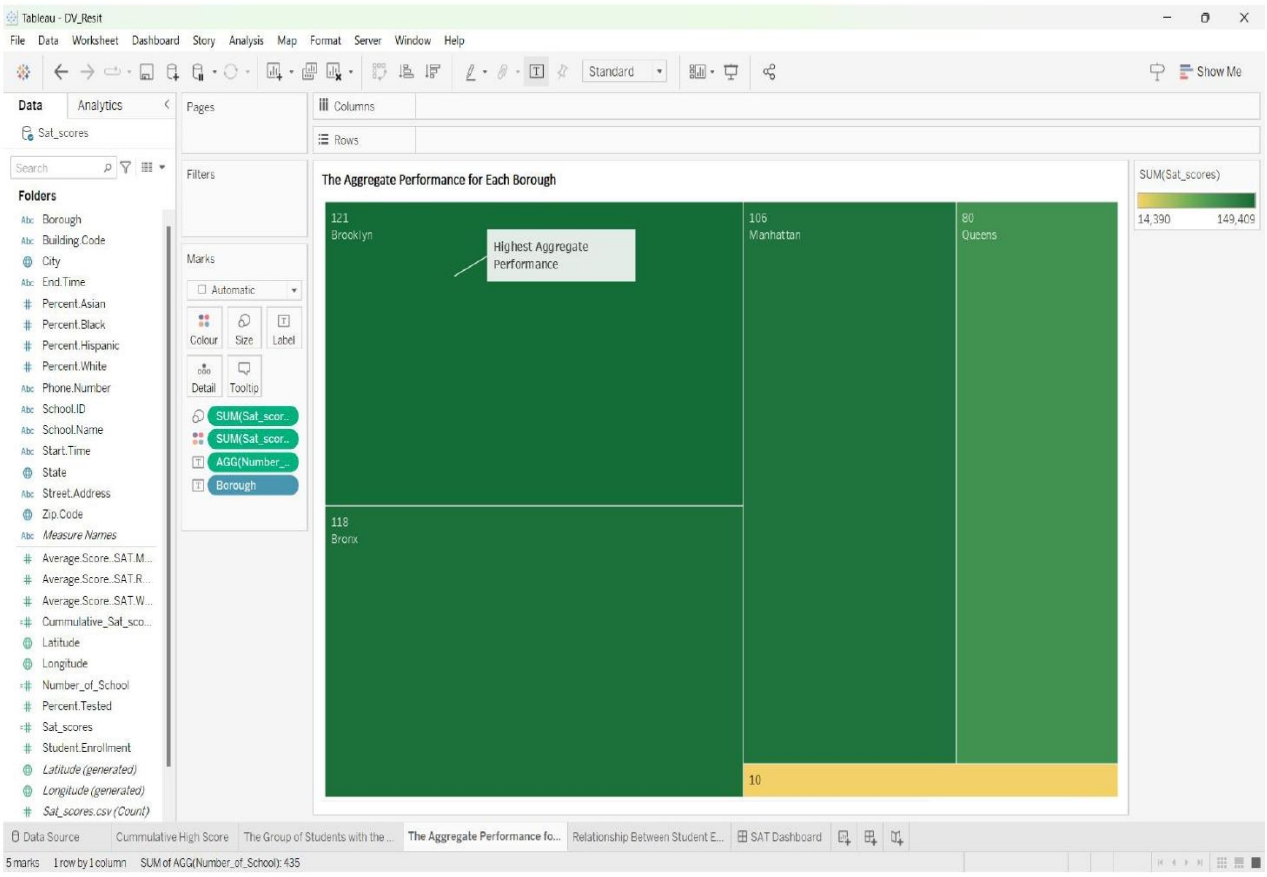
By summing up the SAT scores for all schools, then obtain the average and multiply by the number of student enrol cumulative SAT scores was extracted. The group percentage was available from the original data, as they were place at the appropriate shelf to arrive at the cumulative SAT scores contributed by the Asian group. The resulting value, 0.605%, in Brooklyn Technical High School, 0.734 in Stuyvesant High School respectively. This represents the proportion or percentage of the overall cumulative SAT scores attributed to the Asian group.



Side-by-side bars for Overall Score SAT test and Enrolment

Sheet Three

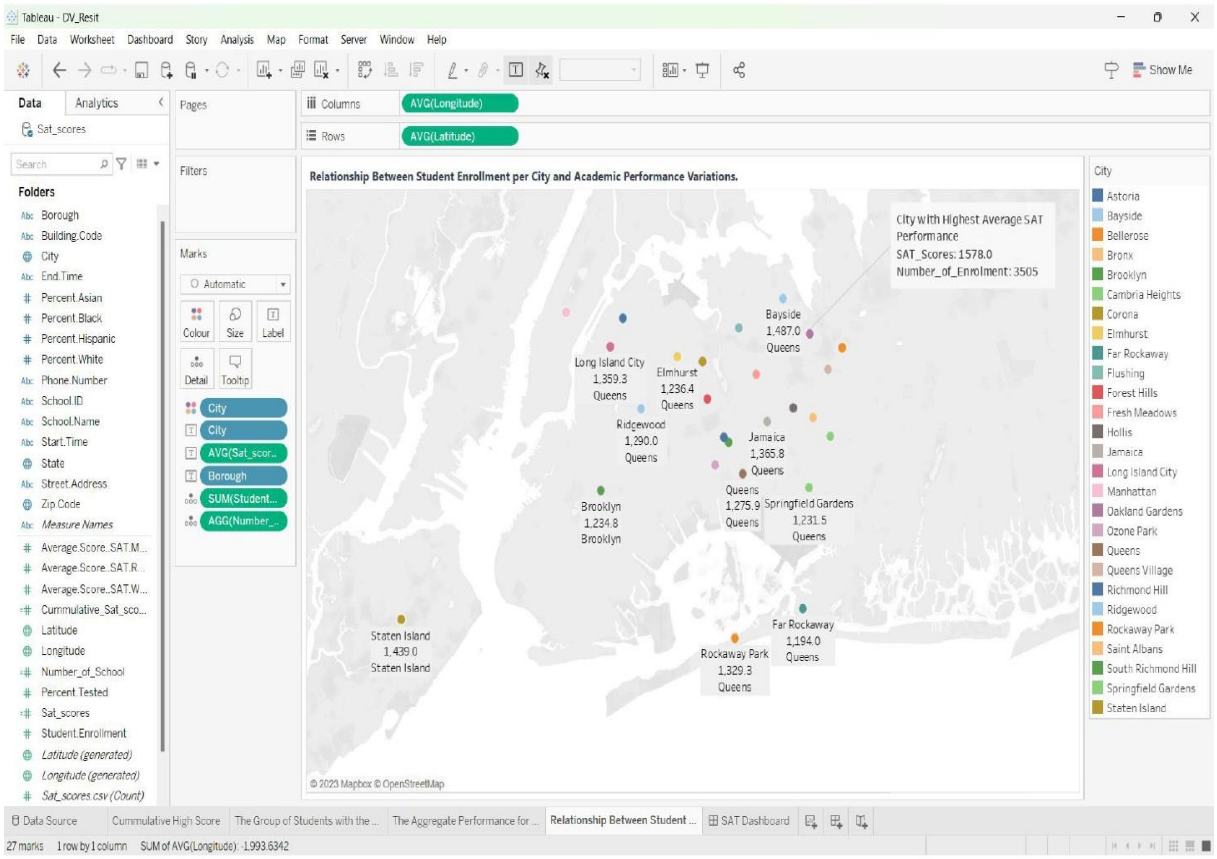
The Treemap visualization is utilized to identify the Borough with the highest overall score and the number of schools in each Borough. The Overall Score measure is added to the Marks section and tagged with size to display the relative size of each Borough on the map. The Sat_score measure is also added to the colour section with two colour thresholds selected to show the map. Additionally, the Sat_scores is added to size on the Mark shell to display the details on the Treemap. The number of schools is calculated using the COUNT() function, which account for the number of schools in each participate in the test, and Borough measured is added to show each Borough on the map.



Tree map to show aggregate performance of each Borough

Sheet Four

To determine the relationship between student enrollment per city and academic performance variations.. The longitude and latitude fields determine the spatial coordinates of each school. The city field is used to color and label the points on the map, while the borough field provides additional labeling. The calculated field "Number of school" counts the schools in each location using the COUNT() function. By selecting the Symbol Map option, the resulting visualization displays the schools as symbols on the map, with city names, borough names, and the count of schools. This visualization allows for exploration of the spatial distribution of schools and analysis of school counts across different cities and boroughs. In the case of Oakland Gardens, which has a student enrollment of 3,505 and only one school, the SAT score is reported as 1,578.0. On the other hand, Brooklyn, with a significantly higher enrollment of 84,577 and 121 schools, has a lower SAT score of 1,234.8. Based on this specific examples provided in this research, a comparison between two schools illustrates the trend that schools with lower student enrollment tend to perform better compared to their counterpart.



Symbol Maps for Borough and Overall Scores.

Discussion

Data visualization is essential for comprehending statistical data, as it uncovers patterns, trends, and relationships. William Playfair, a pioneer in data visualization during the late 18th century, introduced influential techniques in his book "Commercial and Political Atlas" (1786). He invented the line graph to display time-series data, the bar chart for comparing categories or quantities, and popularized the use of pie charts for representing proportions. Jacques Bertin, a notable figure in the mid-20th century, published "Semiology of Graphics" (1967), which introduced the concept of visual variables. He categorized visual elements and emphasized the importance of retinal variables in creating effective visualizations. Bertin's work highlighted the need for a systematic approach to data visualization, considering both visual properties and data structure. Together, Playfair and Bertin's contributions laid the groundwork for modern data visualization techniques, enabling us to comprehend complex information through visual patterns, trends, and relationships. In this research, Tableau, a robust data visualization tool, played a pivotal role in generating valuable insights. The dashboard crafted in Tableau serves as a concise summary of the study's findings, presenting the data in a visually engaging manner.

Sheet One: Calculate cumulative SAT scores for each school by multiplying average SAT score by total number of students who took the SAT. Example: Brooklyn Technical High School has cumulative SAT score of 10,257,732, indicating overall academic performance. Compare scores to assess proficiency and achievement levels.

Sheet Two: Visualize and compare cumulative SAT scores and percentages of different racial/ethnic groups (Black, White, Hispanic, Asian). Asian group shows highest cumulative percentage, reflecting significant contribution to overall SAT scores.

Sheet Three: Use Treemap visualization to identify borough with highest overall score and number of schools. Block sizes represent scores, color indicates SAT score. COUNT() function calculates school count per borough. Analyze borough-wise performance and school distribution.

Sheet Four: Explore relationship between student enrollment per city and academic performance. Plot schools on map using longitude and latitude. Color and labeling based on city and borough. COUNT() function calculates school count. Example: Oakland Gardens (1 school, 3,505 students) vs. Brooklyn (121 schools, 84,577 students). Lower student enrollment correlates with better academic performance.

The Tableau implementation process was simplified by its user-friendly functionality. I utilized calculated fields to aggregate additional values, and the drag and drop feature facilitated an interactive visual approach. While considering the composition of my dashboard, I contemplated adding more than the four sheet. However, I realized that overcrowding the dashboard would hinder the analysis's purpose. Therefore, I decided to include only four sheets, ensuring that users could easily access the information without excessive scrolling. As emphasized by Few (2006), minimizing errors involves using color sparingly, employing it only when necessary, and arranging data in a neat and organized manner within the dashboard. Following these principles ensures that the dashboard is visually coherent and facilitates better understanding of the information presented.

Findings

The study examined the cumulative SAT scores of public high schools to identify the ones with the highest academic performance. Brooklyn Technical High School stood out with a remarkable cumulative SAT score of 10,257,732, indicating strong achievement levels among its students. Additionally, the research investigated relationship between student enrollment and academic

performance variation. The findings revealed a trend supporting this hypothesis that schools with lower student enrollment demonstrate better academic performance.. For instance, Oakland Gardens, with only one school and 3,505 enrolled students, achieved a higher SAT score of 1,578.0 compared to Brooklyn, which had 121 schools and 84,577 students, resulting in a SAT score of 1,234.8. These examples emphasize that schools with lower student enrollment tend to perform better academically. Analyzing group contributions, identifying top-scoring boroughs, and investigating student enrollment's impact on academic performance. Reveal trends and patterns in academic achievement.

Challenges

Challenges in this research include obtaining reliable and comprehensive data on SAT scores and student enrollment, ensuring accuracy and completeness. Analyzing complex datasets requires expertise to identify patterns and draw accurate conclusions. Findings may be specific to selected schools, limiting generalizability. Contextual factors like school culture and community support may also affect academic performance. Tableau poses challenges, including properly labeling cities on the symbol map without overcrowding. Symbol map visualization must accurately represent school distribution. Design considerations and interactive features can address these challenges. Lastly, careful data handling, rigorous analysis, and thoughtful visualization design are essential to ensure valid and reliable research findings.

Conclusion,

Completing this module on data visualization has enhanced my skills and knowledge in efficiently using Tableau, loading data, and creating informative visualizations. I now understand the importance of selecting relevant information and turning raw data into meaningful insights. Excited to apply this knowledge in future projects. Data visualization is a powerful tool that enhances understanding, communication, and decision-making with complex data. It reveals patterns, trends, and relationships, identifies outliers and errors, and supports data-driven choices across various fields. By presenting data visually, it bridges the gap between data and decision-makers, enabling informed choices and driving positive change. It also engages a wider audience, making it essential for effective data communication.

Reference

Few, S. (2006). Information Dashboard Design. O'Reilly Media, Inc.

Spence, R. (2014). Information Visualization: An Introduction. Springer International Publishing AG. Available from: ProQuest Ebook Central.

Mackinlay, J. (2023). The History of Data Visualization-From Crave Drawing to Tableau. TABLEAU SOFTWARE, LLC, A SALESFORCE COMPANY.

Heer, J., & Shneiderman, B. (2012). Interactive dynamics for visual analysis. Communications of the ACM, 55(4), 45-54.

Website
https://www.tableau.com/whitepapers/designing-great-visualizations.

https://www.kaggle.com/datasets/nycopendata/high-schools