Journal of Applied Data Sciences
Vol. 5, No. 2, May 2024, pp. 316-325

An Unsupervised Learning and EDA Approach for Specialized High School Admissions

Adi Suryaputra Paramita[1,*], Arief Ramadhan[2]

[1] *Information Systems Department, School of Information Technology, Universitas Ciputra Surabaya, Indonesia*

[2] *School of Computing, Telkom University, Bandung,40257, Indonesia*

(Received: February 26, 2024; Revised: March 1, 2024; Accepted: March 13, 2024; Available online: May 17, 2024)

Abstract

This research investigates disparities in access and representation within specialized high school admissions processes, focusing on public middle schools in New York City. Leveraging a dataset by a non-profit organization dedicated to increasing diversity in specialized high school admissions, the study employs exploratory data analysis and unsupervised learning techniques to identify schools with high levels of underrepresentation and academic potential. The analysis reveals significant disparities in access to specialized high schools, with certain demographic groups and schools facing barriers to entry. Through k-means clustering, schools are categorized based on their academic performance and demographic composition, enabling targeted intervention strategies to address disparities in access and representation. The research proposes general use towards education, including on-campus interventions, awareness campaigns, and regional information sessions, aimed at fostering equitable access to specialized high school programs. This study contributes to the broader discourse on educational equity and offers valuable insights for policymakers, educators, and researchers seeking to promote diversity and inclusion within educational systems.

Keywords: Exploratory data analysis, Unsupervised learning, Demographic composition, Specialized high schools

1. Introduction

In contemporary education systems, ensuring equitable access to specialized educational opportunities remains a critical concern [1], [2], [3]. The landscape of high school admissions often reflects broader societal disparities, with certain demographic groups facing barriers to entry into prestigious programs [4], [5], [6]. Addressing these disparities requires a nuanced understanding of the factors influencing students' access and success within specialized high school environments [7], [8], [9]. This study seeks to explore the complex interplay between school demographics, academic performance, and access to specialized high schools, proposing a methodology to identify and address inequities in admissions processes.

The proposed methodology centers on a comprehensive analysis of middle school data, leveraging both demographic indicators and academic performance metrics [10], [11], [12]. By examining the compositional differences between feeder and non-feeder middle schools, the study aims to identify patterns of underrepresentation among specific demographic groups [13]. Utilizing unsupervised learning techniques such as k-means clustering, the research seeks to categorize middle schools based on their academic performance levels and potential as feeders for specialized high schools. This approach enables a nuanced understanding of the factors contributing to disparities in access to specialized educational opportunities [14], [15].

One of the key innovations of this study lies in its emphasis on unsupervised learning methods and exploratory data analysis, diverging from traditional predictive modeling approaches. By prioritizing exploratory analysis, the research aims to uncover hidden patterns and insights within the data, avoiding the perpetuation of existing biases encoded in predictive models. Furthermore, the study proposes targeted interventions based on the intersection of demographic

***Corresponding author: Adi Suryaputra Paramita (adi.suryaputra@ciputra.ac.id)**
DOI: https://doi.org/10.47738/jads.v5i2.178

ISSN 2723-6471

composition and academic performance, offering a novel framework for addressing inequities in specialized high school admissions processes.

The primary objective of this research is to provide actionable insights for stakeholders involved in educational equity initiatives. By identifying schools with significant disparities in access to specialized high schools, the study aims to inform targeted intervention strategies aimed at increasing diversity and representation within these institutions. Additionally, the research seeks to contribute to the broader discourse on educational equity, highlighting the importance of data-driven approaches in identifying and addressing systemic disparities in access to specialized educational opportunities. Through these objectives, the study aims to catalyze positive change in the landscape of high school admissions, fostering a more equitable and inclusive educational environment for all students.

2. Literature Review

Numerous studies have delved into the complexities surrounding access and representation within specialized high school admissions, offering valuable insights into the factors influencing students' pathways to these prestigious institutions. For example, Lovett [16] conducted a comprehensive analysis of demographic trends in specialized high school admissions, highlighting disparities in acceptance rates among different racial and socioeconomic groups. Their study underscored the need for targeted interventions to address these disparities, emphasizing the importance of early intervention programs and outreach initiatives in fostering diversity within specialized high school populations [17], [18].

In a similar Chamorro et al. [19] explored the role of middle school academic preparation in influencing students' likelihood of admission to specialized high schools. Through a longitudinal analysis of academic performance data, they identified key predictors of success on specialized high school entrance exams, shedding light on the importance of early academic enrichment programs and access to resources in middle school settings. Their findings underscored the need for interventions aimed at enhancing academic readiness among underrepresented student populations, particularly in schools with limited access to educational resources [20].

Building upon these foundational studies, Czibula et al. [10] employed machine learning algorithms to identify patterns of underrepresentation among middle schools in specialized high school admissions. By analyzing demographic data and academic performance metrics, they developed a predictive model capable of identifying schools with significant disparities in access to specialized high schools. Their study highlighted the potential of data-driven approaches in informing targeted intervention strategies aimed at increasing diversity and representation within specialized high school populations [11], [21], [22].

Furthermore, Mahindru and Sangal [23] conducted a qualitative study exploring the experiences of underrepresented students in specialized high school settings. Through in-depth interviews and focus group discussions, they examined the socio-cultural factors influencing students' sense of belonging and academic success within these institutions. Their findings underscored the importance of fostering inclusive environments and providing targeted support services for underrepresented student populations in specialized high schools [24], [25].

In summary, previous research on similar concepts has highlighted the multifaceted nature of disparities in specialized high school admissions, emphasizing the need for targeted interventions aimed at increasing diversity and representation within these institutions. By leveraging data-driven approaches, such as machine learning algorithms and qualitative analyses, researchers have provided valuable insights into the factors influencing students' pathways to specialized high schools and identified promising strategies for fostering inclusivity and equity within these educational settings.

3. Method

3.1. Dataset Analysis

The dataset utilized for this research is sourced from Kaggle and provided by PASSNYC, a non-profit organization dedicated to increasing diversity in students applying for and receiving placements at specialized high schools (SPHS) in New York City. This dataset comprises comprehensive information on public middle schools in New York City, including demographic data, academic performance metrics, and indicators of socioeconomic status. The relevance of this dataset to the research lies in its alignment with the central focus of the study, which seeks to explore disparities in access and representation within specialized high school admissions processes.

The dataset's suitability for this research is multifaceted. Firstly, it offers rich insights into the demographic composition and academic performance of middle schools in New York City, allowing for a nuanced analysis of factors influencing students' pathways to specialized high schools. Additionally, the dataset includes indicators of socioeconomic status, providing valuable context for understanding disparities in access to specialized high school programs. Moreover, the dataset's provision of school-level data enables a comprehensive examination of the educational landscape, facilitating the identification of patterns and trends that inform targeted intervention strategies.

The relevance of this dataset extends beyond the context of New York City and holds significant implications for educational policies and practices worldwide, including in Indonesia. While the specific nuances of educational systems may vary between countries, the overarching challenges of promoting diversity, equity, and inclusion within specialized high school admissions processes are universal. By leveraging insights from the PASSNYC dataset, policymakers, educators, and researchers in Indonesia can gain valuable insights into effective strategies for addressing disparities in access to specialized high school programs and fostering equitable educational opportunities for all students. Thus, the dataset's high relevance and applicability to the Indonesian case underscore its importance in informing evidence-based decision-making and fostering positive change in educational systems globally.

3.2. Research Steps

The research step as shown in Figure 1, process begins with an executive summary and exploratory data analysis (EDA), offering a succinct overview of dataset analysis and preliminary findings. This summary serves as a precursor to subsequent steps aimed at investigating which public middle schools are more likely to have students underrepresented at SPHS.

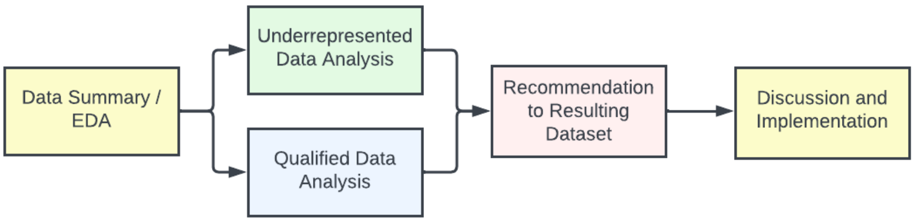

Figure 1. Research Steps

The analysis unfolds by examining the compositional differences between feeder and non-feeder schools, utilizing statistical methods to assess demographic and socioeconomic disparities. Additionally, an Underrepresentation Score is calculated for each middle school, delineating the extent of underrepresentation among specific groups. Subsequently, middle schools with the highest underrepresentation scores are pinpointed, spotlighting areas necessitating targeted interventions to alleviate disparities in SPHS access.

Further exploration focuses on identifying public middle schools more likely to have students qualified for SPHS. Through K-means clustering, schools are categorized based on academic performance and pertinent features. This step aims to unveil schools with a heightened likelihood of producing academically qualified SPHS candidates. Subsequent analysis of the resulting clusters elucidates patterns and trends in academic performance levels among middle schools, providing insights into factors influencing students' readiness for SPHS admissions.

The culmination of the research culminates in recommendations to PASSNYC, tailored to address identified disparities and enhance access to specialized high schools. These recommendations encompass targeted interventions, such as on-campus interventions at schools with high academic potential (Cluster A), awareness campaigns at schools with notable education quality (Cluster B), and regional information sessions and workshops aimed at all schools, particularly those clustered in areas with pronounced underrepresentation. Through these targeted strategies, the research endeavors to foster equitable access and representation within specialized high school admissions.

4. Result and Discussion

4.1. Exploratory Data Analysis

EDA presented in this study delves into the dynamics of the SPHS in New York City, particularly focusing on issues of diversity and academic performance among middle school students. By examining school-level data, the analysis seeks to offer actionable insights to PASSNYC, an organization committed to enhancing diversity among SPHS applicants and recipients. Unlike traditional approaches that rely on biased data or predictive models, this analysis adopts a novel perspective, leveraging exploratory analysis and unsupervised learning techniques. By dissecting the demographic composition and academic performance of middle schools, the study aims to identify schools that historically underrepresent certain demographics while also harboring academically qualified students likely to aspire to SPHS admission. Through this intersectional lens, the analysis pinpoints middle schools where interventions from PASSNYC could be most effective. The recommendations advocate for targeted strategies such as on-campus interventions at five specific middle schools, a widespread awareness campaign across 48 schools, and the organization of regional information sessions and workshops at three key locations. These initiatives are strategically designed to address the multifaceted challenges surrounding access and representation within the SPHS system, thereby fostering a more equitable and inclusive educational landscape in New York City.

4.2. Underrepresented Data Analysis

The specialized high schools in New York City are currently under scrutiny due to their perceived lack of diversity, particularly in terms of gender, race, and socioeconomic background. Corcoran and Baker-Smith's research utilizes sequential logistic models to highlight the underrepresentation of specific social groups, namely Black, Hispanic, female, and low-income students, across various stages of the SPHS application process. This section utilizes demographic data at the school level to explore the diversity issue within SPHS, with a focus on two main inquiries: whether feeder schools, which contribute significantly to SPHS offers, exhibit distinct compositions compared to other schools, and which public schools' students are most likely to face underrepresentation in the SPHS application process.

4.2.1. Compositional difference between feeder and non-feeder schools

In a typical year, approximately 25,000 students in New York City take the Specialized High Schools Admissions Test (SHSAT) and apply for placement in SPHS, constituting about one-third of the total eighth-grade cohort in the city. However, the distribution of offers among students from various middle schools is unequal, with the top 10 middle schools accounting for a quarter of all offers in 2018. To assess whether these feeder schools, which concentrate offers, differ in composition from other schools, density plots were utilized to analyze three key metrics: the percentage of female students, the percentage of Black and Hispanic students, and the Economic Need Index (ENI), calculated based on factors such as temporary housing as shown in Figure 2.

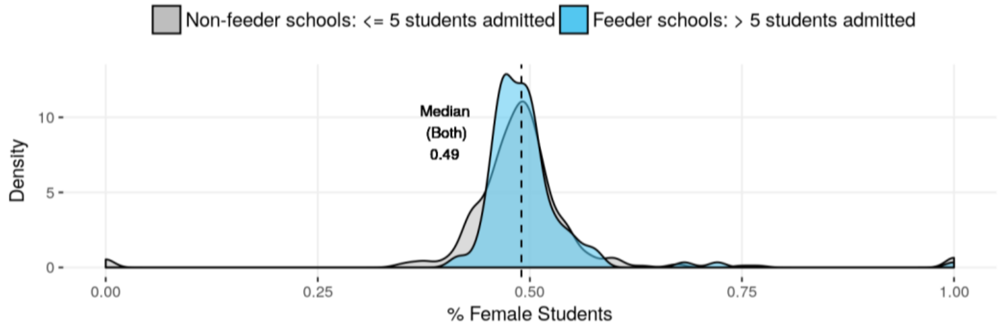

Figure 2. Distribution of % Female Students

The analysis revealed that feeder schools and non-feeder schools exhibited similar distributions in terms of gender ratio, with both centered around 49%. Despite gender disparities observed at various stages of the application process, the similarity in gender ratio across schools complicates the assessment of middle school impact using solely school-level data. Conversely, significant disparities were observed in the racial composition and economic need level between

ISSN 2723-6471

feeder and non-feeder schools, as depicted in Figures 3 and 4. Notably, Asian, Caucasian, and high-income students were not only more likely to attend SPHS but also originated from middle schools with a higher proportion of Asian and Caucasian students and lower economic need.

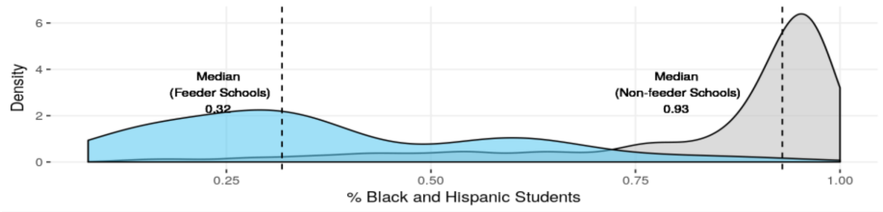

Figure 3. Distribution of % Black and Hispanic Students

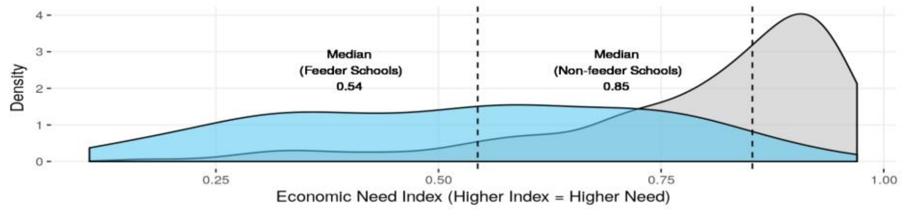

Figure 4. Distribution of Economic Need Index

These compositional disparities, along with the widespread distribution of the latter two metrics, underscore the socioeconomic and racial segregation inherent in New York City's public school system. Such disparities are perpetuated by social prejudices and discriminatory policies, further contributing to societal divisions. In the subsequent section, the ENI and racial composition metrics will be utilized to calculate a scaled score, aiding in the measurement of the likelihood that students from a particular middle school will be underrepresented at SPHS.

4.2.2. Underrepresentation Score

The scatterplot depicted Figure 5 below illustrates a strong positive correlation ($r = 0.77$, $p < 0.01$) between the Economic Need Index and the percentage of Black and Hispanic students. Feeder schools, indicated by larger-sized blue points, generally exhibit low-to-medium economic need and a lower proportion of Black or Hispanic students, while a notable cluster of non-feeder schools (gray points) appears in the upper right corner of the plot. These non-feeder schools predominantly consist of low-income Hispanic or Black students who are often underrepresented at SPHS.

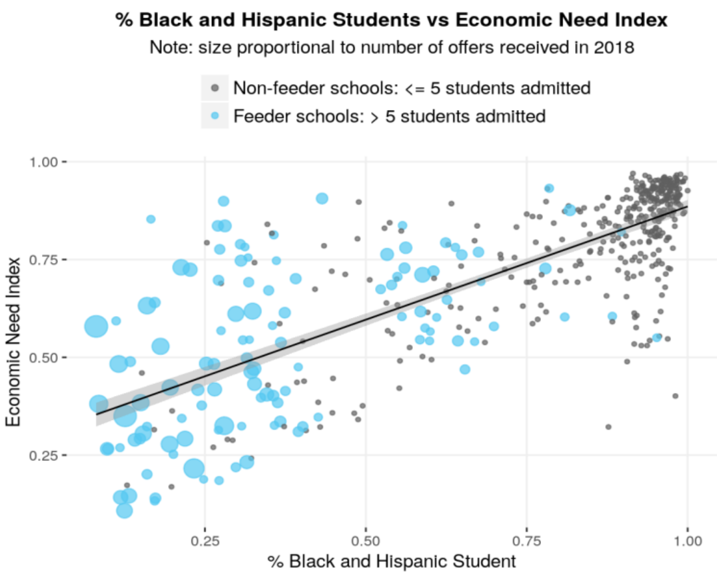

Figure 5. Positive correlation between Economic Need Index and % Black and Hispanic students

To facilitate PASSNYC's prioritization of outreach efforts based on the varying needs of students across different schools, an Underrepresentation Score will be introduced. This score aims to quantify the likelihood of students from a particular school experiencing underrepresentation during the SPHS application process. Calculated as the scaled Euclidean distance between each school's data point and the origin, the Underrepresentation Score ranges from 0 to 1. A score closer to 1 indicates a higher level of underrepresentation at SPHS. For a comprehensive list of middle schools with their corresponding Underrepresentation Scores, please refer to Table 1 in the appendix.

4.2.3. Middle schools with the highest underrepresentation scores

In order to maximize the impact of PASSNYC's efforts, it is crucial to pinpoint the middle schools most likely to be underrepresented at SPHS and strategically allocate resources accordingly. As such, the subsequent analysis will concentrate on the top 25% of middle schools exhibiting the highest Underrepresentation Scores, comprising 121 out of the 482 middle schools with available demographic data as shown in Figure 6. Notably, none of these 121 schools had more than five students admitted to SPHS in 2018. The provided table details information on these 121 schools, arranged in descending order based on their Underrepresentation Scores. Refer to the appendix for a comprehensive list of middle schools alongside their corresponding Underrepresentation Scores. Remarkably, the top 25% of schools with the highest Underrepresentation Scores span across 20 out of the 32 school districts in New York City, with districts showcasing higher index scores generally harboring a greater proportion of underrepresented middle schools.

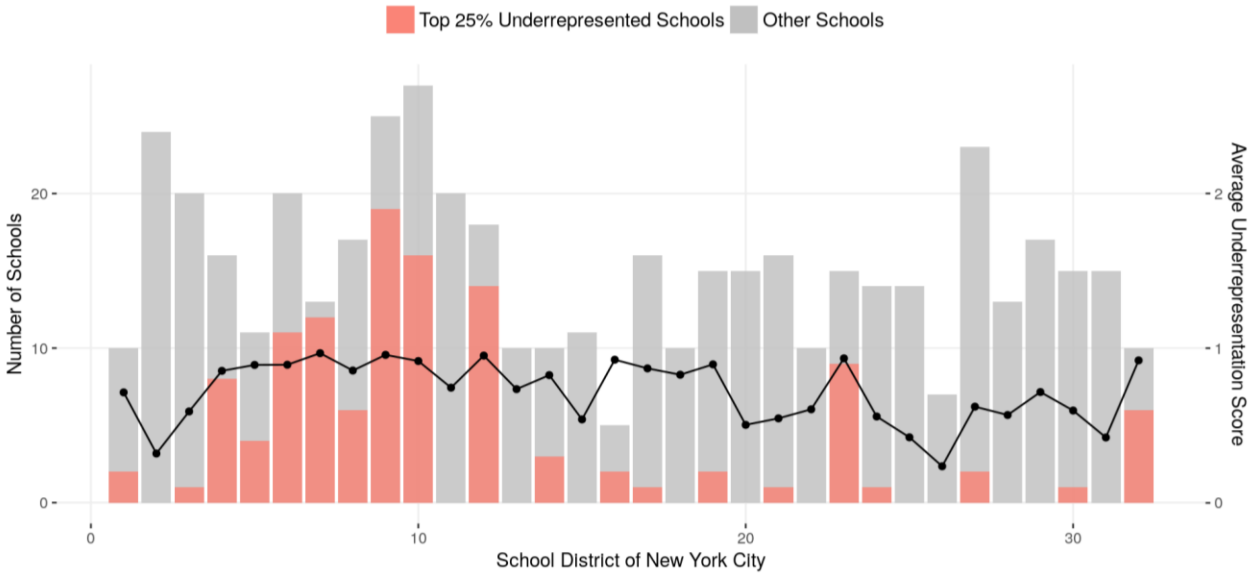

Figure 6. School Breakdown and Underrepresentation Score by District

4.3. Analysis Qualified Data Analysis

SPHS in New York City have a rich tradition of catering to the educational needs of students with exceptional academic or artistic abilities. Admission to eight out of nine SPHS is primarily determined by performance on the Specialized SHSAT, which evaluates proficiency in English and Math. Given the pivotal role of SHSAT in securing placements in these prestigious institutions, it is presumed that students from middle schools with robust academic performance are more inclined to seek admission to SPHS. Consequently, this section will concentrate on academic performance, specifically aiming to address one key question through the utilization of k-means clustering: which non-feeder schools exhibit education quality and academic performance levels akin to those of feeder schools?

4.3.1. K-means Clustering and Features Used

In this section, we utilize the k-means clustering algorithm, an unsupervised learning technique, to investigate the academic characteristics of non-feeder schools in comparison to feeder schools. Building upon the conclusion drawn in the previous section that feeder schools exhibit distinct compositions from non-feeder schools, our focus shifts to identifying non-demographic features that could render non-feeder schools academically similar to feeder schools. The analysis encompasses various indicators, including New York State annual test results such as the percentage of eighth

ISSN 2723-6471

graders scoring at each level in Math and English Language Arts (ELA) tests, as well as the average scale scores for these subjects. Additionally, data from the New York City Department of Education's annual quality review, encompassing measures of rigorous instruction, collaborative teaching practices, supportive environments, effective school leadership, strong family-community ties, and trust, are examined. Other factors under consideration include average student attendance rates, chronic absenteeism rates, pupil-to-teacher ratios, and average class sizes. By leveraging k-means clustering, we aim to identify non-feeder schools that demonstrate academic similarities to feeder schools, thereby shedding light on potential pathways to enhancing academic preparation and opportunities for students from underrepresented backgrounds.

4.3.2. K-Means Clusters

Utilizing the elbow method based on within-cluster sum of squares, the analysis identified three as the optimal number of clusters. Employing the k-means algorithm and utilizing the 20 listed features, 472 middle schools with available non-demographic data were partitioned into three distinct clusters showcasing varying levels of academic performance, as depicted in Figure 7 the subsequent scatterplot. Notably, Cluster A emerged with over 60% of its middle schools categorized as feeder schools, indicating that these institutions had more than five students accepted by SPHS in 2018. Furthermore, Cluster A exhibited the highest mean scores in state-wide ELA and Math examinations, both recognized as robust predictors for students' performance in the SHSAT. Leveraging the cluster labels assigned to each middle school based on non-demographic features, the analysis proposes an inference regarding the likelihood of a given middle school to have students qualified for SPHS.

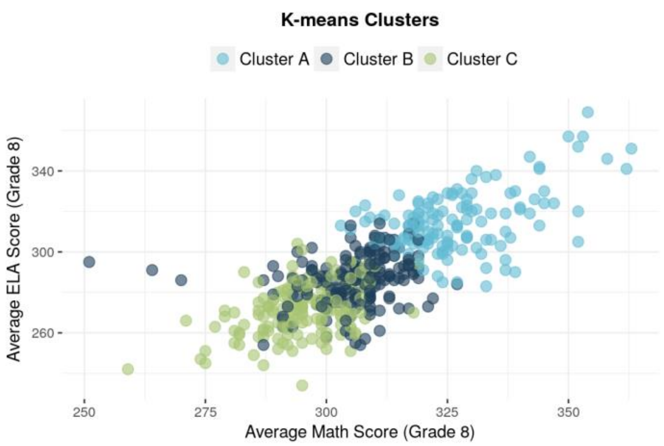

Figure 7. K-Means Clusters Result

Note: Cluster A is deemed the most probable; Cluster B somewhat likely; and Cluster C least likely to have students qualified for SPHS admission.

4.4. Recommendation to Resulting Dataset

4.4.1. Recommendation 1: On-campus intervention at 5 schools in Cluster A

In Cluster A, which comprises middle schools most likely to have students qualified for SPHS, 5 schools from the top quantile of the Underrepresentation Score are identified as shown in Figure 8. Despite their potential, none of these schools had more than 5 students accepted by SPHS in 2018. Thus, it is crucial to prioritize on-campus interventions at these schools to ensure that students benefit from services provided by PASSNYC and its partners. The recommended interventions include organizing on-campus information sessions to engage parents and students, assigning volunteers to address questions and assist with the application process, and facilitating access to test preparation resources through recruitment and matching processes. Additionally, on-campus test preparation workshops or tutoring sessions should be organized to enhance students' readiness for the SHSAT.

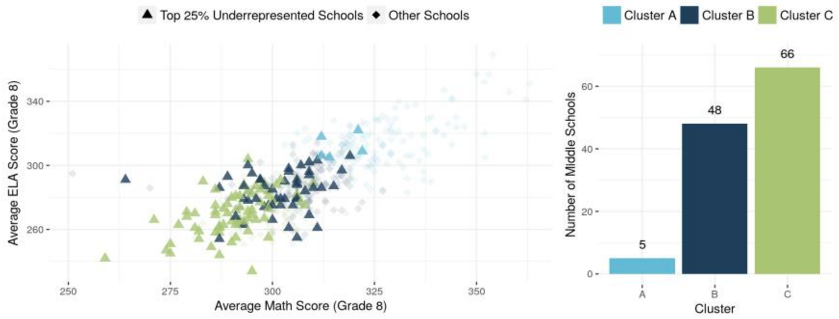

Figure 8. K-means Clusters for Top 25% Middle Schools with the Highest Underrepresentation Scores

4.4.2. Recommendation 2: Awareness campaign at 48 schools in Cluster B

Cluster B encompasses 48 middle schools with notable education quality and academic potential. To increase awareness about the SHSAT and SPHS within this cluster, an awareness campaign is recommended. This campaign could involve distributing informational flyers, sending physical mails or emails to teachers, parents, and students, and conducting outreach activities to disseminate essential information about the application process and available resources.

4.4.3. Recommendation 3: Regional information sessions and workshops at 3 locations

An analysis of the geographic distribution of public middle schools in New York City highlights three prominent locations: Harlem, Bronx, and Brooklyn (Broadway Junction), which are characterized by a high proportion of Black and Hispanic residents as shown in Figure 9. Given this clustering and the historical impact of residential segregation on education and household income, it is proposed that PASSNYC organize regional information sessions and workshops at these locations. These sessions would aim to boost awareness of the SHSAT and SPHS, assign volunteers to provide guidance and support to parents and students, and facilitate access to regional test preparation workshops tailored to the needs of the communities. This targeted approach ensures that resources are effectively allocated to address the specific challenges faced by schools in these geographic clusters.

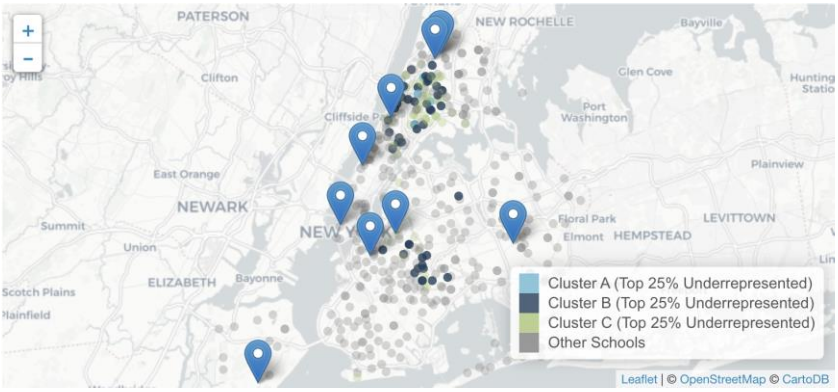

Figure 9. Locations of Middle Schools

Note: Circles = Public Middle Schools; Markers = Specialized High Schools

5. Conclusion

This research delineates the intricacies of the educational landscape in New York City, focusing on the compositional disparities between feeder and non-feeder middle schools as well as the levels of representation among students in SPHS programs. Data analysis indicates that while there are several middle schools that produce academically qualified students for SPHS, significant disparities persist in the representation of certain social groups, such as Black, Hispanic, female, and low-income students. These findings underscore the importance of targeted interventions to enhance accessibility and representation in SPHS, with recommendations for awareness campaigns in specific schools, regional information sessions, and direct assistance in schools with high academic potential but low representation in SPHS.

ISSN 2723-6471

Thus, the research provides valuable insights for organizations like PASSNYC in designing more effective intervention strategies to address disparities in access and representation in these prestigious higher education institutions.

Within the context of available data limitations, this study also emphasizes the necessity of a holistic approach in understanding the educational challenges in New York City. The findings highlight the importance of considering factors such as demographic composition, educational quality, and geographic location in designing effective intervention strategies. By providing specific and measured recommendations, this research offers a deep insight into the potential improvements that can be achieved through collaboration between non-profit organizations, educational institutions, and local governments. In conclusion, this research not only provides a better understanding of the dynamics of education in New York City but also lays a strong foundation for collaborative efforts aimed at enhancing equality of access and representation in prestigious higher education institutions in the city.

6. Declarations

6.1. Author Contributions

Conceptualization: A.S.P. and A.R.; Methodology: A.R.; Software: A.S.P.; Validation: A.S.P., A.R.; Formal Analysis: A.S.P., A.R.; Investigation: A.S.P.; Resources: A.R.; Data Curation: A.R.; Writing Original Draft Preparation: A.S.P. and A.R.; Writing Review and Editing: A.R. and A.S.P.; Visualization: A.S.P.; All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] A. E. Khedr and N. Yaseen, "Predicting stock market behavior using data mining technique and news sentiment analysis," *Int. J. Intell. Syst. Appl.*, vol. 9, no. 7, p. 22, 2017.

[2] X. Liu and J. Walsh, "Study on development strategies of fresh agricultural products e-commerce in China," *Int. Bus. Res.*, vol. 12, no. 8, pp. 61–70, 2019.

[3] R. Dwi Sihono, "Jurnal Internasional: The Effect of Transformational Leadership, Workload, and Compensation on Employee Performance.," *Adv. Soc. Sci. Res. J.*, vol. 9, no. 11, pp. 311–324, 2022.

[4] A. R. Novaliyan et al., "Bimbingan dan Konseling Mahasiswa yang Berbasis Sistem Pakar dengan Menggunakan Metode Faktor Kepastian," *J. Eng. Technol. Appl. Sci.*, vol. 3, no. 2, pp. 21–34, 2021, doi: 10.36079/lamintang.jetas-0302.234.

[5] X. Lu, Y. Zhu, Y. Xu, and J. Yu, "Learning from multiple dynamic graphs of student and course interactions for student grade predictions," *Neurocomputing*, vol. 431, pp. 23–33, 2021, doi: https://doi.org/10.1016/j.neucom.2020.12.023.

[6] N. Hsueh, "An Alliance-Based Term Project in Software Quality Courses : a Lesson," *IJIIS Int. J. Informatics Inf. Syst.*, vol. 1, no. 1, pp. 28–35, 2018.

[7] J. Littenberg-Tobias and J. Reich, "Evaluating access, quality, and equity in online learning: A case study of a MOOC-based

ISSN 2723-6471

blended professional degree program," *Internet High. Educ.*, vol. 47, p. 100759, 2020, doi: 10.1016/j.iheduc.2020.100759.

[8] X. Griffin, D. Griffin, A. Berry, and D. Hunter, "Cancellation of elective surgery – any improvement after ten years?," *Bull. R. Coll. Surg. Engl.*, vol. 88, no. 1, pp. 28–30, 2006, doi: 10.1308/147363506x78082.

[9] Z. Zvizdic et al., "The predictors of perforated appendicitis in the pediatric emergency department: A retrospective observational cohort study," *Am. J. Emerg. Med.*, vol. 49, pp. 249–252, 2021, doi: 10.1016/j.ajem.2021.06.028.

[10] G. Czibula, A. Mihai, and L. M. Crivei, "S PRAR: A novel relational association rule mining classification model applied for academic performance prediction," *Procedia Comput. Sci.*, vol. 159, pp. 20–29, 2019, doi: https://doi.org/10.1016/j.procs.2019.09.156.

[11] F. Greco and A. Polli, "Emotional Text Mining: Customer profiling in brand management," *Int. J. Inf. Manage.*, vol. 51, no. December 2018, pp. 1–8, 2020, doi: 10.1016/j.ijinfomgt.2019.04.007.

[12] K. Fiok, W. Karwowski, E. Gutierrez, and M. Wilamowski, "Analysis of sentiment in tweets addressed to a single domain-specific Twitter account: Comparison of model performance and explainability of predictions," *Expert Syst. Appl.*, vol. 186, no. July 2020, p. 115771, 2021, doi: 10.1016/j.eswa.2021.115771.

[13] W. C. Lin, S. W. Ke, and C. F. Tsai, "Top 10 data mining techniques in business applications: a brief survey," *Kybernetes*, vol. 46, no. 7, pp. 1158–1170, 2017, doi: 10.1108/K-10-2016-0302.

[14] N. Shafiabady, L. H. Lee, R. Rajkumar, V. P. Kallimani, N. A. Akram, and D. Isa, "Using unsupervised clustering approach to train the Support Vector Machine for text classification," *Neurocomputing*, vol. 211, pp. 4–10, 2016.

[15] S. Sarfraz, N. Murray, V. Sharma, A. Diba, L. Van Gool, and R. Stiefelhagen, "Temporally-weighted hierarchical clustering for unsupervised action segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 11225–11234.

[16] B. J. Lovett, "Disability identification and educational accommodations: Lessons from the 2019 admissions scandal," *Educ. Res.*, vol. 49, no. 2, pp. 125–129, 2020.

[17] S. Gilsbach et al., "Increase in admission rates and symptom severity of childhood and adolescent anorexia nervosa in Europe during the COVID-19 pandemic: data from specialized eating disorder units in different European countries," *Child Adolesc. Psychiatry Ment. Health*, vol. 16, no. 1, p. 46, 2022.

[18] K. Hextrum, Special admission: How college sports recruitment favors white suburban athletes. Rutgers University Press, 2021.

[19] K. Chamorro, M. Fors, F. X. Mora, and M. Pulgar-s, "Biomarkers of severe COVID-19 pneumonia on admission using data-mining powered by common laboratory blood tests-datasets," no. January, 2020.

[20] R. Chetty, D. J. Deming, and J. N. Friedman, "Diversifying society's leaders? The causal effects of admission to highly selective private colleges," *National Bureau of Economic Research*, 2023.

[21] S. H. Duncan, "Myspace Is Also Their Space: Ideas for Keeping Children Safe from Sexual Predators on Social-Networking Sites.," *Kentucky Law J.*, vol. 96, no. 4, pp. 527–577, 2008.

[22] T. Y. Cheng, N. Li, C. J. Conselice, A. Aragón-Salamanca, S. Dye, and R. B. Metcalf, "Identifying strong lenses with unsupervised machine learning using convolutional autoencoder," *arXiv*, vol. 17, no. April, pp. 1–17, 2019, doi: 10.1093/mnras/staa1015.

[23] A. Mahindru and A. L. Sangal, "SemiDroid: a behavioral malware detector based on unsupervised machine learning techniques using feature selection approaches," *Int. J. Mach. Learn. Cybern.*, vol. 12, no. 5, pp. 1369–1411, 2021, doi: 10.1007/s13042-020-01238-9.

[24] P. Gutierrez, A. Cordier, T. Caldeira, and T. Sautory, "Data augmentation and pre-trained networks for extremely low data regimes unsupervised visual inspection," *Opt. Metrol.*, 2021.

[25] P. O. O Pinheiro, A. Almahairi, R. Benmalek, F. Golemo, and A. C. Courville, "Unsupervised learning of dense visual representations," *Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 4489–4500, 2020.