UNP JOURNAL MANAGEMENT CENTER

math.ppj.unp.ac.id
p-ISSN 2716-0726
e–ISSN 2716-0734

Indexed

Article History
Vol. 4, No. 2, 2025

Subject Areas:
Exploratory Data Analysis, Cluster Analysis, Second Subject

Keywords:
Admission pathway, exploratory study, hierarchical density based clustering, public junior high school, student admission

Author for correspondence:
Dodi Vionanda
Email: dodi_vionanda@fmipa.unp.ac.id

DOI:
https://doi.org/10.24036/rmj.v4i2.106

How to Cite:
Vionanda, D., Wood, R., A., & Susrifalah, A. (2025). Do Prestigious Schools Still Exist in Padang? An Exploratory Study on State Junior High School Admission 2025 in Padang. *Rangkiang Mathematics Journal*, 4(2), 119-131. https://doi.org/10.24036/rmj.v4i2.106

Do Prestigious Schools Still Exist in Padang? An Exploratory Study on State Junior High School Admission 2025 in Padang

Dodi Vionanda[1], Raihan Attaya Wood[1] and Amelia Susrifalah[1]

[1] Department of Statistics, Universitas Negeri Padang

Abstract- In this study, we perform an exploratory study of New Student Admission datasets for public Junior High School in Padang in 2025. We utilized tables, barplots, and boxplots to present information contained in datasets and we carried out cluster analysis using HDBSCAN algorithm. For this study we made use of admitted students' datasets for each admission pathway of all state Junior High Schools in Padang in 2025. We carried out this study to investigate the emergence of prestigious schools among public Junior High School in Padang amid the implementation on zoning system. Our study reveals the presence of group of prestigious schools along with group of schools that admitted students mostly live nearby the schools. Hence, it is recommended for Padang Municipal government to improve the quality of schools that are not considered as prestigious schools since there are many schools that admitted students mostly live nearby the school.

1. Introduction

Student admission process in Indonesia, known as Penerimaan Peserta Didik Baru (PPDB, New Student Admission) is an annual event in an academic institution at the beginning of an academic year. Based on Primary and Secondary Indonesian Ministerial Act No. 3 of 2025, there are four admission paths for public schools in Indonesia namely domicile, affirmative, achievement and relocating paths. The affirmative path includes impoverished family and disability student paths. The achievement path entails school report card, academic achievement and non-academic achievement paths. The relocating path consists parent relocating and teacher's children paths. Meanwhile, for domicile

path there is only one path (Kementerian Pendidikan Dasar dan Menengah, 2025).

The implementation of this admission system aims to ensure access equality to educational service for students, to bring them closer to their family environment and to eliminate discrimination in schools (Salim & Nora, 2022). This admission system also provides an access to quality education for low-performing students coming from disadvantaged families who live nearby a favorite public school (Mahyani et al, 2019).

However, the goal of PPDB remains only partially attained. Sulistyosari et al, (2023) reported that the above policy has not restrained parents from enrolling their children into schools designated as favorite schools. Likewise, Sutadi et al, (2025) also found that schools either were located within urban areas or those with better reputation are more preferable. In addition, Salim & Nora, (2022) stated that some parents manipulated the system to gain school admission for their children.

The same phenomenon also occurred in countries with school zoning policy in student admissions. For instance, in Denmark, Bjerre-Nielsen et al, (2023) proposed that the applicants manipulate eligibility criteria and change addresses before the high school application deadline. Meanwhile, in the UK, distance to higher-performing primary school affects the prices of nearby homes. Whereas in the US, He & Giuliano (2018) put forward that parents tend to choose top-achieving schools for their children even though they should leave their neighborhood schools. In addition, Gibbons & Machin (2003) and Black & Machin (2011) have demonstrated how proximity to high-performing schools affects housing prices in the UK, while Reardon et al. (2019) and He & Giuliano (2018) have discussed the persistence of educational stratification in the US despite zoning reforms.

Moving on now to consider quantitative studies on PPDB. Some of these studies made use of datasets of the distance between schools and students' homes and students' scores. Among others, Nurhidayat & Juliane (2024) performed research study on 847 students who enrolled SMKN 1 Kedawaung, Cirebon, in 2023. They carried out clustering algorithm on dataset of the distance between SMKN 1 Kedawaung, Cirebon, and students' domiciles and reported the presence of three clusters. Meanwhile, Iskandar & Ernawati (2023) found that most of the admitted students of SMAN 1 Tanah Putih, Tanjung Melawan, Rokan Hilir, in 2020, lived within a distance of 5 km. However, the studies described above were conducted at the school level, focusing mainly on how the admission policy is implemented within individual educational institutions. There appears to be a lack of research on PPDB datasets at municipal-level. In this regard, this study presents the results from implementing exploratory study on both students' home and school distance datasets for all admission paths and student score for school report card admission path at Padang municipal-level.

2. Methods

(a) New Student Enrollment

PPDB with zoning system has been implemented since 2017 in Indonesia based on Education and Culture Ministerial Act No. 17 of 2017. This act has been implemented in order to create equal access to education (Sutadi et al, 2025). Education and Culture Ministerial Acts that regulates PPDB has been revised annually. The latest one is Primary and Secondary Indonesian Ministerial Act No. 3 of 2025. According to this act, there are four admission paths for public schools in Indonesia namely domicile, affirmative, relocating, and achievement paths. The first three pathways involve distance data between students' home and the admitted school. Meanwhile, the last one takes into account both distance data between students' home and the admitted school and score of school report. Hence, for this study, we shall consider datasets on these two quantities.

(b) Datasets

In this study, we utilized datasets of PPDB 2025 for 45 Public Junior High Schools (SMP Negeri) in Padang which previously available at https://psb.diknaspadang.id/home/smp/. Data cleaning stage was

not required in this study since we did not find any missing value as well as duplicate entry. In general, there are 7.888 students were admitted to the entire SMP Negeri in Padang in 2025. These datasets consist of information on students' report scores for report score, academic achievement, and non-academic achievement admission paths and information on distance between students' home and admitted schools for the entire admission paths. Hence, we only deal with numerical datasets. The details of these variables are described in Table 1. Nevertheless, not all of these variables are included in the final analysis.

Table 1. Variable description

Admission Pathway	Variables	Variable Name
Domicile Path		
Domicile	Distance between students' home and admitted schools	*Dist_dom* (x_1)
Achievement Path		
School report card	Score of report card	*Score_school_report* (x_2)
	Distance between students' home and admitted schools	*Dist_school_report* (x_3)
Academic achievement	Score of report card	*Score_acad_achiev* (x_4)
	Distance between students' home and admitted schools	*Dist_acad_achiev* (x_5)
Non-academic achievement	Score of report card	*Score_non_acad_achiev* (x_6)
	Distance between students' home and admitted schools	*Dist_non_acad_achiev* (x_7)
Affirmative		
Impoverished family	Distance between students' home and admitted schools	*Dist_impover_fam* (x_8)
Inclusive Child	Distance between students' home and admitted schools	*Dist_inclu_child* (x_9)
Transfer		
Relocating	Distance between students' home and admitted schools	*Dist_relocate* (x_{10})
Teachers' Children	Distance between students' home and admitted schools	*Dist_teach_child* (x_{11})

(c) Exploratory Study

An exploratory study is often utilized to reveal patterns in datasets and generating hypothesis. This approach is operationalized through Exploratory Data Analysis (EDA). EDA may be viewed as the technique of study one or more datasets in an effort to get the insight of the underlying structured contained in datasets (Pearson, 2018). EDA involves data visualizations to identify patterns, trends, and relationships, identifying outliers and anomalies, examining relationships between variables, and detecting the presence of two or more groups of observations in datasets. For these purposes, EDA utilizes data visualization as well as data summarization through tables. In this study, we carried out exploratory study on datasets obtained from PPDB SMP 2025 in Padang. We make use of tables, bar plots, and boxplots.

(d) Clustering Analysis with HDBSCAN

HDBSCAN clustering is a density-based clustering algorithm. Density-based clustering is an approach where the clusters are contiguous dense regions in the data space separated by low density. The data points in the separating regions of the low point density are typically considered noise (Sander, 2017).

There are several algorithms for density-based clustering, namely DBSCAN (Ester et al, 1996) and DENCLUE (Hinneburg & Keim, 2003). However, these algorithms only being able to provide a flat labeling of the observations. Hence, Campello et al, (2013) proposed HDBSCAN as a hierarchical based-clustering to address this limitation.

For a proper formulation of HDBSCAN algorithm, we define the notations of core distance and a symmetric reachability distance, a new notion of *ε-core objects*, as well as the notion of a *conceptual*, transformed proximity graph, which will help us to explain a density-based clustering hierarchy.

Let $X = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n\}$ be a dataset of n objects and let D be an $n \times n$ matrix containing the pairwise distances $d(\boldsymbol{x}_p, \boldsymbol{x}_q) \in X$ for a metric distance $d(.,.)$. In the following, we define clustering algorithm HDBSCAN as can be found in Campello et al. (2013, 2015).

Definition 2.1. (Core Distance). *The core distance of an object $\boldsymbol{x}_p \in X$ with respect to m_{pts}, $d_{core}(\boldsymbol{x}_p)$ is the distance from $\boldsymbol{x}_p$ to its m_{pts}-nearest neighbor (including $\boldsymbol{x}_p$)*

Definition 2.2. (Mutual Reachability Distance). *The mutual reachability distance between two objects $\boldsymbol{x}_p, \boldsymbol{x}_q \in X$ with respect to m_{pts} is defined as*

$$d_{mreach}(\boldsymbol{x}_p, \boldsymbol{x}_q) = max\{d_{core}(\boldsymbol{x}_p), d_{core}(\boldsymbol{x}_q), d(\boldsymbol{x}_p, \boldsymbol{x}_q)\}$$

Definition 2.3. (Mutual Reachability Graph). *The mutual reachability graph is a complete graph, $G_{m_{pts}}$, in which the objects of X are vertices and the weight of each edge is the mutual reachability distance with respect to m_{pts} between the respective pair of objects.*

Preposition 2.1. *Let X be a set n objects described in a metric space by $n \times n$ pairwise distances. The clustering of this data obtained by DBSCAN* with respect to m_{pts} and some value ε is identical to the one obtained by first running Single-Linkage over the transformed space of mutual (with respect to m_{pts}), then, cutting the resulting dendogram at level ε of its scale, and treating all resulting singleton with $d_{core}(\boldsymbol{x}_q) > \varepsilon$ as a single class representing "Noise".*

The followings are the main steps of HDBSCAN.

HDBSCAN Algorithm.

1. *Compute the core distance with respect to m_{pts} for all data objects in X.*
2. *Compute an MST of $G_{m_{pts}}$, the Mutual Reachability Graph.*
3. *Extend the MST to obtain MST_{ext} by adding for each vertex a "self edge" with the core distance of the corresponding object as weight.*
4. *Extract the HDBSCAN* hierarchy as a dendogram from MST_{ext}:*
   a. *For the root of the tree assign all objects the same label (single "cluster").*
   b. *Iteratively remove all edges from MST_{ext} in decreasing order of weights (in case of ties, edges must be removed simultaneously):*
      i. *Before each removal, set the dendogram scale value of the current hierarchical level as the weight of the edges to be removed.*
      ii. *After each removal, assign labels to the connected components that contain the end vertices of the removed edges, to obtain the next hierarchical level: assign a new cluster label to a component if it still has at least one edge, else assign it a null label ("noise")*

For computational purpose, we made use of hdbscan function from dbscan R package of Hahsler et al (2019).

(e) Density-Based Clustering Validity

For clustering validity, we employ density-based clustering validity (DBCV) of Moulavi et al (2014) as we are dealing with density-based clustering. Moulavi et al (2014) proposed that the globular clustering validations might fail for density based clustering. Chicco et al (2025) confirmed this result. They compared the performance of DBCV to Silhouette index (Rousseeuw, 1987), Davies-Bouldin index (DBI) (Davies & Bouldin, 1979), Calinski-Harabasz index (CHI) (Caliliski & Harabasz, 1974), Dunn index

(Dunn, 1974), and Gap statistic (Tibshirani et al., 2001), and found that DBCV is more effective. There are four quantities considered for density-based clustering validity, i.e., density sparseness of a cluster (DSPC), density separation (DS), validity index of a cluster (VC) and validity index of a clustering (DBCV). To formulate these four quantities we define the notion of all-points-core-distance (a_pts coredist) as in Definition 2.4.

Definition 2.4. (Core Distance of An Object). *The all-points-core-distance of an object* $\boldsymbol{o}$ *belonging to cluster* C_i *with respect to all other* $n_i - 1$ *objects in* C_i *is defined as*

$$a_{pts}coredist(o) = \left(\frac{\sum_{i=2}^{n_i} \left(\frac{1}{KNN(\boldsymbol{o}, i)} \right)^D}{n_i - 1} \right)^{-\frac{1}{d}}$$

where $KNN(\boldsymbol{o}, i)$ be the distance between objects o and its i^{th}nearest neighbors.

What follows are definitions of DSPC, DS, VC and DBCV.

Definition 2.5. (Density Sparseness of A Cluster). *The Density Sparseness of Clusters (DSC)* C_i *is defined as the maximum edge weight of the internal edges in* MST_{MRD} *of the cluster* C_i*, where* MST_{MRD}*is the minimum spanning tree constructed using* $a_{pts}coredist$ *considering the objects in* C_i*.*

Definition 2.6. (Density Separation). *The Density Separation of a Pair of Clusters (DSPC)* C_i *and* C_j*,* $1 \leq i, j \leq l, i \neq j$*, is defined as the minimum reachability distance between the internal nodes of the* MST_{MRD} *of clusters* C_i *and* C_j*.*

Definition 2.7. (Validity Index of A Cluster). *Validity index of a cluster* C_i *is defined as:*

$$V_C(C_i) = \frac{\min\limits_{1 \leq j \leq l, j \neq i} \left(DSPC(C_i, C_j) \right) - DSC(C_i)}{max\left(\min\limits_{1 \leq j \leq l, j \neq i} \left(DSPC(C_i, C_j) \right), DSC(C_i) \right)}$$

Definition 2.8. (Validity Index of A Clustering). *The validity index of a clustering solution* $C = \{C_i\}$*,* $1 \leq i \leq l$ *is defined as the weighted average of the Validity Index of all clusters in* C

$$DBCV(C) = \sum_{i=1}^{l} \frac{|C_i|}{|O|} V_C(C_i)$$

DSC evaluates how dense and connected points are inside a cluster. Meanwhile, DSPC quantifies how well a cluster is separated from others using density-based distances. Moreover, VC assesses the quality of each cluster, while DBCV evaluates the quality of clustering based on density separation and density connectedness**.** According to Moulavi et al (2014), if a cluster has better density compactness than density separation, we obtain positive values of the validity index. If the density inside a cluster is lower than the density that separates it from other clusters, the index is negative. For computational purpose, we made use of dbcv function from dbscan R package of Hahsler et al (2019).

3. Results and Discussion

(a) Exploratory Study

In this section, we present the results from performing exploratory study on PPDB 2025 datasets. We begin by displaying the number of students admitted in all SMP in Padang in 2025 as in Table 1. This table indicates that domicile, school report, and impoverished family paths are the ones with higher number of admitted students than the other paths. On the other hand, academic achievement path is the one with least number of admitted student. There was only one student who enrolled the school through this admission path. Hence, we exclude dataset from this admission path from our further discussion. We also omit datasets from non-academic achievement, inclusive child and teachers' children admission path for the same reason. There were only small number of students admitted through these admission

pathways.

Table 2. Number of students who passed the admission

Domicile	Admission Path						
	Merit-based			Affirmative		Transfer	
	School report	Academic Achievement	Non Academic Achievement	Impoverished family	Inclusive Child	Relocating	Teachers' Children
4460	1915	1	183	1082	66	143	38

Moreover, Table 2 exhibits the number of schools admitting students through a specific admission pathway. This table also shows that domicile, school report, impoverished family and relocating paths are the ones with higher number of schools admitting students than the other paths. Therefore, we only consider datasets from these pathways for subsequent analysis as previously mentioned. In terms of variables, we only take into account variables x_1, x_2, x_3, x_8 and x_{10}. It should be noted that we alternately use values of these variables on each individual student within a school and median values of these variables obtained from each school. Boxplots in Figures 3, Figure 4, and Figure 5 employed the former; meanwhile cluster analysis carried out in Section b utilized the latter.

Table 3. Number of Schools Admitting Students Through a Specific Admission Path

	Admission Path							
		Merit-based			Affirmative		Transfer	
	Domicile	School report	Academic Achievement	Non Academic Achievement	Impoverished family	Inclusive Child	Relocating	Teachers' Children
Present	45	45	1	23	43	27	42	22
None	0	0	44	22	2	18	3	23

Let us now consider the results obtained through exploratory study on distance datasets from the four admission pathways above. We begin by taking the number of students admitted for each admission path into account as in Figure 1 and Figure 2. Each panel within Figure 1 exhibits these numbers from domicile and impoverished family. Meanwhile, each panel within Figure 2 display these numbers for relocating and report grade paths. These panels denote that the number of students admitted through each admission paths varies across school. Closer inspection on all panels within Figure 1 and Figure 2 show that SMPN 18 and SMPN 20 are two schools admitted larger number of students through all admission pathways than the other schools. On contrast, SMPN 37 and SMPN 44 are two school admitted the smaller number of students.

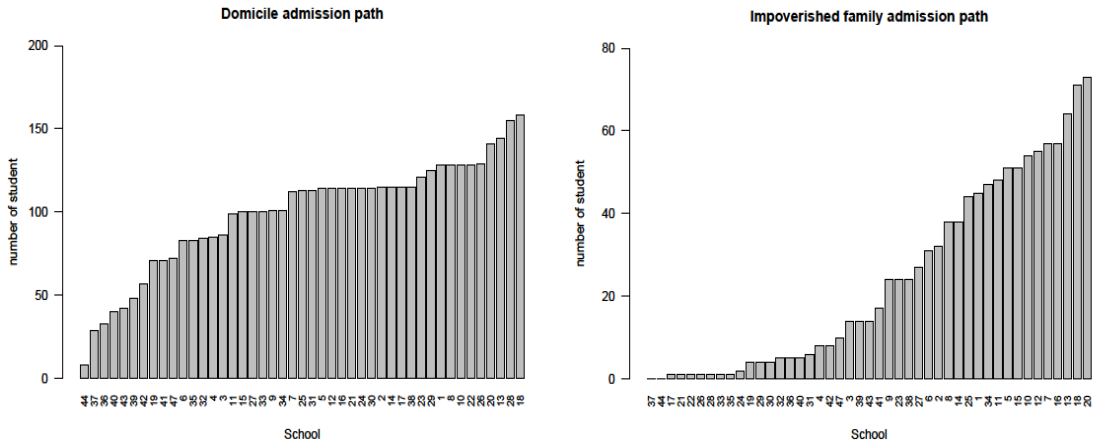

Figure 1. Barplots for Number of Students Admitted for Domicile and Impoverished Family Admission Path.

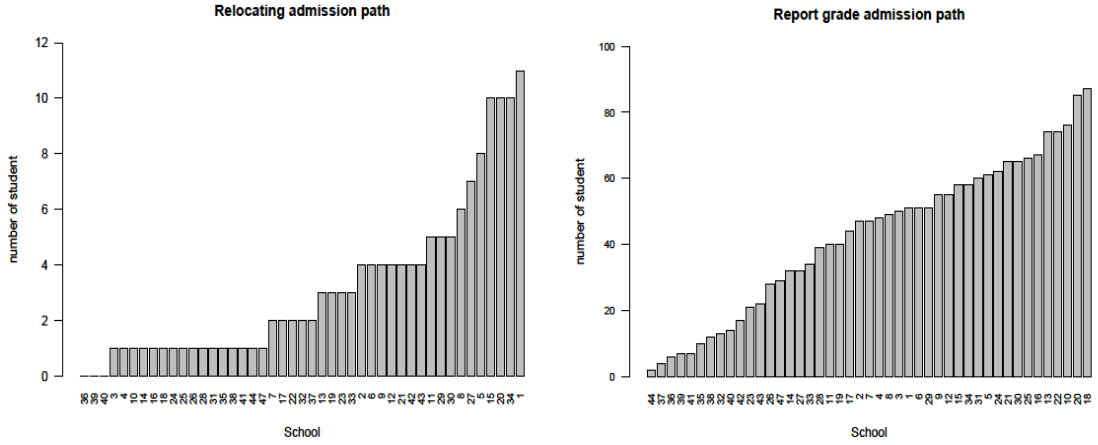

Figure 2. Barplots for Number of Students Admitted for Relocating and Report Grade Admission Path.

Furthermore, we present boxplots for the distance between individual students' homes and the admitting schools within each school in Figures 3 and Figure 4. In general, these figures show that students who admitted through domicile admission path live in closer proximity to the school. On contrast, most of the students who admitted through school report admission path live at a distance from the school. Some of the admitted applicants through this admission pathway even continued to use their out-of-district addresses outside Padang.

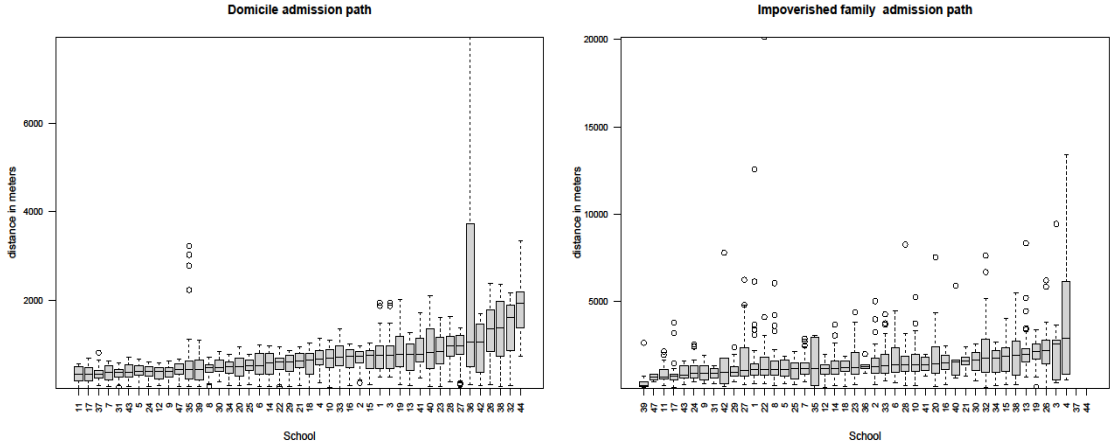

Figure 3. Boxplots for distance between the admitting school and student domicile for domicile and impoverished family admission path.

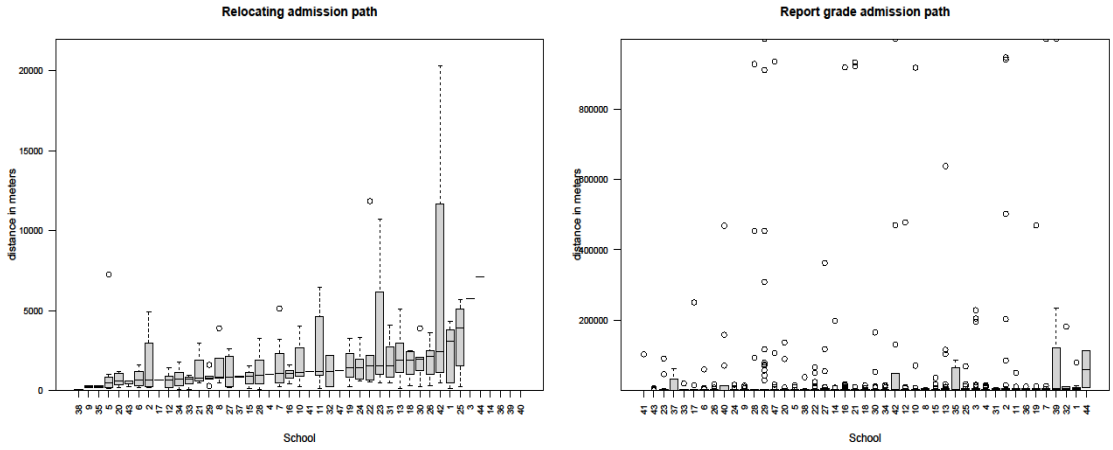

Figure 4. Boxplots for distance between the admitting school and student domicile for relocating and report grade

admission path.

Moreover, we present boxplots for the scores of individual student report cards from each school in Figure 5. Figure 5 indicates the presence of top five schools, namely, SMPN 1, SMPN 7, SMPN 8, SMPN 11, and SMPN 2. Students enrolled these schools through score of school report card have higher marks than the other public schools in Padang. On the other hand, those who admitted to SMPN 37 and SMPN 32 have lower marks.

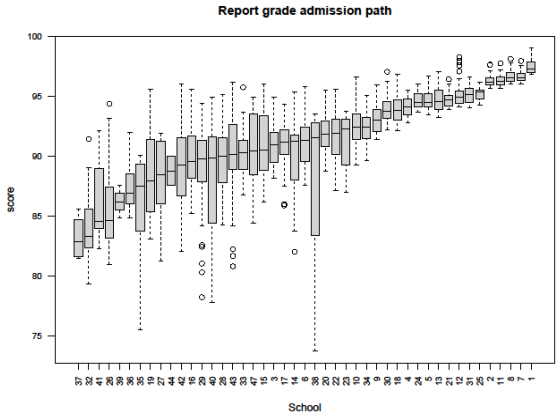

Figure 5. Boxplot for school report card scores of students admitted to each junior high school.

(b) Clustering The Schools

In this section, we present the results obtained from clustering the schools using variables x_1, x_2, x_3, x_8 and x_{10} by implementing HDBSCAN algoritm (Campello et al, 2013). We carry out data standardization prior to fitting the clusters. We performed cluster analysis by using HDBSCAN algorithm with hyper parameter min Pts = 3 for all combinations of the above variables and examined the resulting DBCV values. We obtained seven combinations with DBCV values greater than 0,5 as presented in Table 4. These DBCV values indicate that the resulting clusters are moderately good.

However, among the clusters obtained by seven combinations of variables above, we only consider the ones involving variables x_1, x_2 and x_3, namely combinations 1, 4, 5, 6 and 7. We only consider these variables along with their combinations owing to the fact that there are large enough number of measurements on these variables as shown in Table 2.

Table 4. DBCV Values of Clustering Which are Greater than 0,5

Combination	Variables	DBCV
1	x_1, x_2	0,590
2	x_{10}	0,587
3	x_{10}, x_2	0,567
4	x_3, x_2	0,551
5	x_3	0,538
6	x_2	0,530
7	x_1	0,506

Furthermore, fitting the cluster using variables x_1 and x_2, we obtain the same cluster as of using variable x_1. Likewise, fitting the cluster using variables x_2 and x_3, we obtain the same cluster as of using variable x_3. Therefore, we only discuss the clusters yielded using variables x_1, x_2 and x_3, separately. Figure 6, Figure 7, and Figure 8 present the clusters obtained using variables x_1, x_2 and x_3, consecutively.

Moving on now to consider the results from performing Cluster Analysis by employing HBDSCAN algorithm, one found that there are 8 clusters in dataset along with a group of observations considered as noises as presented in Table 5 by fitting clustering using variable x_1. As shown in boxplots of Figure 6, the schools belong to either Cluster 3, 4, 5, 6, 7, or 8 are the ones with have low variability in terms of

median distance values. Meanwhile, the schools within Clusters 3, 4 and 5 are the ones with low median distance values. This indicates that, in general, the schools belong to these clusters that enrolled students live nearby. The list of these schools can be seen in Table 5. Moreover, group of schools which viewed as noise as can be seen as a cluster with 0 label is the group of schools which may not belong to any cluster. In terms of the values of median distance, the schools considered as noise are the ones with median distance values vary.

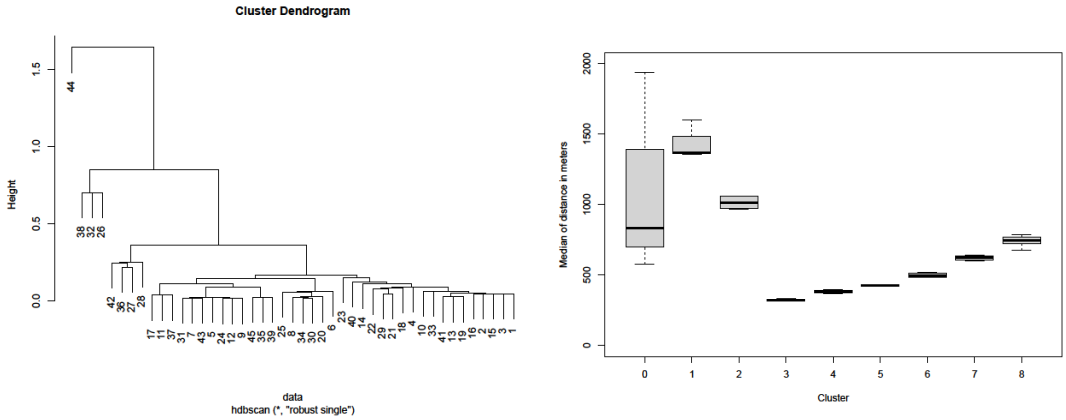

Figure 6. Dendogram and Boxplots for Clusters Obtained Using Variable x_1.

Turning now to cluster validity, Table 5 shows that Cluster 6 which is considered as a most compact cluster. Also, this cluster is found to be well separated from the other clusters since it has high DCSP values as shown in Table 6. However, this cluster has the lowest cluster validity index. On the other hand, Cluster 2 along with Cluster 8 are considered to be the least compact one, but at the same time they have high cluster validity.

Table 5. Cluster Membership Attained from Using Median of Distance Data for Domicile Admission Pathway

Cluster	School	Median of Unstandardized Distance			DSC	VC
		Min	Median	Max		
0	14, 23, 40, 44	575.55	833.84	1939.32		
1	26, 32, 38	1359.12	1366.57	1603.52	0.104	0.55
2	27, 28, 36, 42	964.93	1014.63	1062.72	0.025	0.75
3	11, 17, 37	319.85	322.72	332.77	0.056	0.67
4	5, 7, 9, 12, 24, 31, 43	370.93	383.07	395.45	0.032	0.79
5	35, 39, 47	426.23	426.65	434.60	0.049	0.79
6	6, 8, 20, 25, 30, 34	484.93	493.97	516.90	0.691	0.23
7	18, 21, 22, 29	602.07	622.69	643.30	0.216	0.62
8	1, 2, 3, 4, 10, 13, 15, 16, 19, 33, 41	675.11	745.11	786.57	0.023	0.77

Table 6. DSPC Values for Clusters Fitted Using Median of Distance Data for Domicile Admission Pathway

		Cluster						
		1	2	3	4	5	6	7
Cluster	2	0.917						
	3	0.563	0.273					
	4	1.115	0.155	0.471				
	5	0.231	0.641	0.287	0.839			
	6	1.692	2.801	2.447	Inf	2.115		
	7	0.572	1.682	1.328	1.880	0.995	0.903	
	8	0.817	0.101	0.172	Inf	0.540	Inf	1.581

The following is the results obtained by fitting clustering using variable x_2. As presented in Figure 7 and Table 7, one obtained 6 clusters schools along with a group of observations considered as noises. What is

interesting about this resulting clustering is the presence of schools belong to either Cluster 1 or Cluster 6. While the ones belong the former are the schools with lowest values for median score of school report, the ones belong to the latter are the schools with highest values. This result indicates the presence of the schools with either performing students or non-performing ones. This eventually resulted the emergence of prestigious schools that are more preferable than the other schools oncoming years.

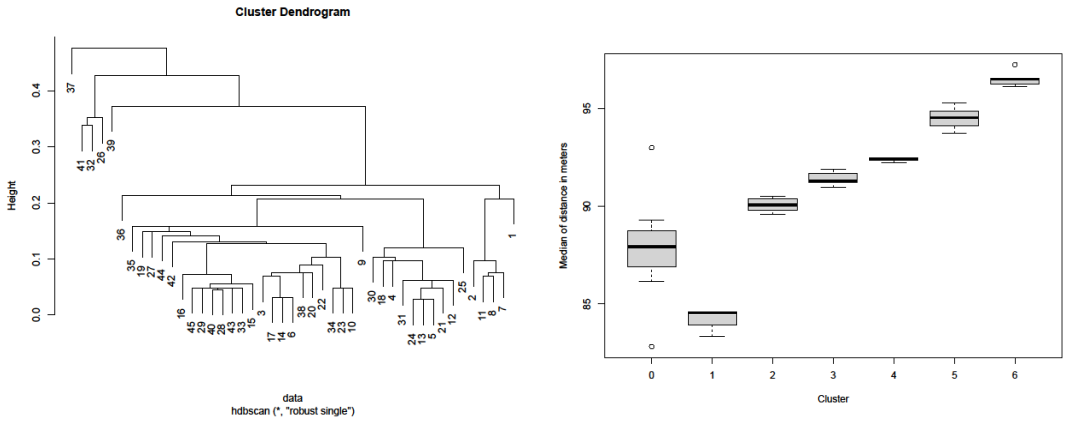

Figure 7. Dendogram and Boxplots for Clusters Attained Using Variable x_2.

Furthermore, with regard to the cluster validity as presented in Table 8, this cluster also considered to be the most compact cluster as well as the most distinctive one relative to the other. The schools belong this cluster are close together and strongly separated from the rest of schools in Padang simultaneously.

Table 7. Cluster Membership Yielded from Using Median of Score Data for School Report Card Admission Pathway

Cluster	School	Median of Unstandardized Score			DSC	VC
		Min	Median	Max		
0	37, 39, 36, 35, 19, 27, 44, 42, 9	82.83	87.94	93.00		
1	32, 41, 26	83.33	84.55	84.60	0.069	0.82
2	16, 29, 40, 28, 43, 33, 47, 15	89.58	90.07	90.51	0.090	0.45
3	3, 17, 14, 6, 38, 20, 22	90.98	91.30	91.88	0.134	0.65
4	23, 10, 34	92.25	92.43	92.43	0.048	0.71
5	30, 18, 4, 24, 5, 13, 21, 12, 31, 25	93.75	94.54	95.31	0.072	0.71
6	2, 11, 8, 7, 1	96.15	96.50	97.25	0.344	0.76

Table 8. DSPC Values for Clusters Fitted Using Median of Score Data for School Report Card Admission Pathway

		Cluster				
		1	2	3	4	5
Cluster	2	1.219				
	3	0.379	0.544			
	4	1.054	0.165	0.379		
	5	1.636	0.251	0.961	0.582	
	6	3.224	1.839	2.549	Inf	1.433

Having discussed the results obtained from fitting clustering using variables x_1 and x_2, let us now to examine the ones from utilizing variable x_3. Figure 8 and Figure 9 display dendogram and boxplots yielded for these results. What is striking about these results is that the emergence of large number of schools considered as noise as shown Table 9. The schools belong to this group marked by excessive variability. This reveals a tendency among parents who enrolled their children through school report admission pathway to neglect the relevance of home-school distance. Some of these schools are belong to the Cluster 6 from the previous result such as SMPN 1, SMPN 2, SMPN 7, SMPN 8 and SMPN 11.

These schools are the ones preferred by high performing students.

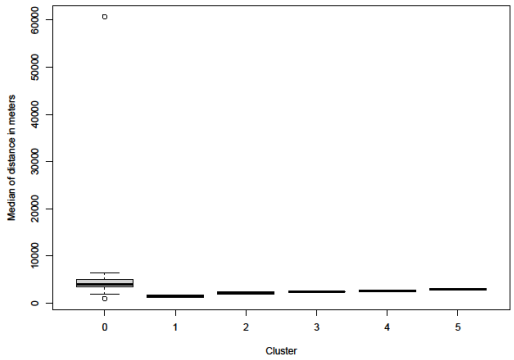

Figure 8. Dendogram and Boxplots for Clusters Produced Using Variable x_3.

Moreover, compared to the results obtained by employing variable x_1, which is the distance between students' home and admitted school for domicile admission pathway, the ones yielded using variable x_3 reveal that the clusters of schools with much higher values for median of distance. Apart from Cluster 1, all schools in the remaining clusters have values of median for distance between students' home and admitted school exceeding 2000 meters.

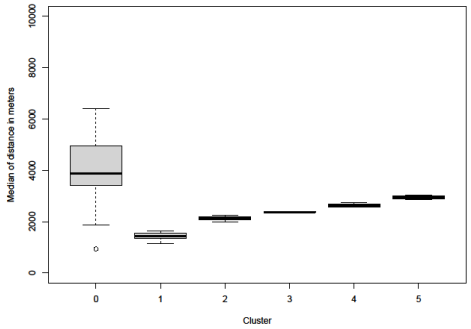

Figure 9. Revised Version of Boxplots for Clusters Yielded Using Variable x_3.

In terms of cluster validity, both Table 9 and Table 10 reveal that the resulting clusters have both low compactness and separateness even though they have high cluster validity indices. This indicates that using information on distance between students' home and the admitted school for school report admission pathway to fitting clustering one ends up with the results are not entirely correct.

Table 9. Cluster Membership Yielded from Using Median of Distance Data for School Report Card Admission Pathway

Cluster	School	Median of Unstandardized Distance			DSC	VC
		Min	Median	Max		
0	1, 2, 4, 5, 7, 11, 19, 31, 32, 36, 39, 41, 44	937.15	3879.94	60727.81		
1	6, 9, 17, 20, 23, 24, 26, 28, 29, 33, 37, 40, 43, 47	1157.12	1443.02	1628.86	0.0079	0.79
2	14, 16, 18, 21, 22, 27, 38	1975.06	2133.75	2237.68	0.0089	0.84
3	30, 34, 42	2348.95	2373.60	2403.94	0.0022	0.91
4	8, 10, 12, 15	2596.18	2607.19	2740.46	0.0116	0.47
5	3, 13, 25, 35	2861.82	2961.80	3032.61	0.0045	0.80

Table 10. DSPC Values for Clusters Fitted Using Median of Distance Data for School Report Card Admission Pathway

		Cluster			
		1	2	3	4
Cluster	2	0.160			
	3	0.037	0.122		
	4	0.086	0.055	0.048	

5	0.064	0.096	0.026	0.022

4. Conclusion

In this study we carried out exploratory study on PPDB 2025 for public Junior High School in Padang by employing Exploratory Data Analysis approach along with Clustering Analysis using HBDSCAN. We analyzed datasets on distance between students' home and admitted school for all admission pathways as well as score of student school report. Our study reveals the emergence of group of prestigious schools along with group of schools that admitted students live nearby the schools.

References

1. Bjerre-Nielsen, A., Christensen, L. S., Gandil, M. H., & Sievertsen, H. H. (2023). Playing the system: address manipulation and access to schools. -: -.
2. Black, S. E., & Machin, S. (2011). Housing valuations of school performance. In E. A. Hanushek, S. Machin, & L. Woessmann, Handbook of the Economics of Education (pp. 485-519). Amsterdam: Elsevier.
3. Caliliski, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. Communications in Statistics, 3(1), 1-27.
4. Campello, R. J., Kröger, P., Sander, J., & Zimek, A. (2020). Density-based clustering. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10(2), e1343.
5. Campello, R. J., Moulavi, D., Zimek, A., & Sander, J. (2015). Hierarchical density estimates for data clustering, visualization, and outlier detection. ACM Transactions on Knowledge Discovery from Data (TKDD), 10(1), 1-51.
6. Campello, R. J., Moulavi, D., & Sander, J. (2013). Density-based clustering based on hierarchical density estimates. Pacific-Asia conference on knowledge discovery and data mining (pp. 160-172). Berlin: Springer Berlin Heidelberg.
7. Chicco, D., Sabino, G., Oneto, L., & Jurman, G. (2025). The DBCV index is more informative than DCSI, CDbw, and VIASCKDE indices for unsupervised clustering internal assessment of concave-shaped and density-based clusters. PeerJ Computer Science, 11, e3095.
8. Davies, D. L., & Bouldin, D. W. (1979). A Cluster Separation Measure. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1(2), 224-227.
9. Dunn, J. C. (1974). Well-separated clusters and optimal fuzzy partitions. Journal of cybernetics, 4(1), 95-104.
10. Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. The Second International Conference on Knowledge Discovery and Data Mining (pp. 226-231). Porland Oregon: AAAI Press.
11. Gibbons, S., & Machin, S. (2003). Valuing English primary schools. Journal of urban economics, 53(2), 197-219.
12. Gibbons, S., Machin, S., & Silva, O. (2005). Competition, choice and primary school performance. London School of Economics. London: Centre for Economic Performance, London School of Economics Working Paper.
13. Hahsler, M., Piekenbrock, M., & Doran, D. (2019). dbscan: Fast Density-Based Clustering with R. Journal of Statistical Software, 91(1), 1-30.
14. He, S. Y., & Giuliano, G. (2017). School choice: understanding the trade-off. Transportation, 45(5), 1475-1498.
15. Hinneburg, A., & Keim, D. A. (2003). A general approach to clustering in large databases with noise. Knowledge and information systems, 5(4), 387-415.
16. Iskandar, & Ernawati. (2023). Efektivitas Sistem Zonasi Penerimaan Peserta Didik Baru (PPDB) di SMA Negeri 1 Tanah Putih Tanjung Melawan Kabupaten Rokan Hilir Tahun Ajaran 2020/2021. Jurnal Buana, 7(2), 342 - 350 .
17. Kementerian Pendidikan Dasar dan Menengah. (2025). Peraturan Menteri Pendidikan Dasar dan Menengah Republik Indonesia Nomor 3 Tahun 2025 tentang Sistem Penerimaan Murid Baru (SPMB). Jakarta. Retrieved from https://peraturan.bpk.go.id/Details/315671/permendikdasmen-no-3-tahun-2025
18. Mahyani, E. R., Wahyunengseh, R. D., & Haryanti, R. H. (2019). Public Perception of Zoning School Policy in Surakarta Public Senior High Schools. 1st International Conference on Administration Science (ICAS 2019). 343. Bandung: Atlantis Press.
19. Moulavi, D., Jaskowiak, P. A., Campello, R. J., Zimek, A., & Sander, J. (2014). Density-Based

Clustering Validation. Proceedings of the 2014 SIAM international conference on data mining (pp. 839-847). Pennsylvania: Society for Industrial and Applied Mathematics.
20. Nurhidayat, H., Hariyanto, & Juliane, C. (2024). Data Mining Implementation in Admission of New Students Using Zone Systems. Indonesian Journal of Computer Science, 13(1), 229-240.
21. Pearson, R. K. (2018). Exploratory data analysis using R. Chapman and Hall/CRC.
22. Reardon, S. F. (2019). Educational opportunity in early and middle childhood: Using full population administrative data to study variation by place and age. RSF: The Russell Sage Foundation Journal of the Social Sciences, 5(2), 40-68.
23. Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of computational and applied mathematics, 20, 53-65.
24. Salim, F. P., & Nora, D. (2022). Dampak Penerimaan Peserta Didik Baru (PPDB) Sistem Zonasi (Studi Kasus: Penerimaan Peserta Didik Baru Sekolah Dasar di Kecamatan Matur). Naradidik: Journal of Education and Pedagogy, 1(1), 67 - 77.
25. Sander, J. (2017). Density-based clustering. In C. Sammut, & G. I. Webb, Encyclopedia of Machine Learning and Data Mining (2 ed., pp. 349-353). Sidney: Springer.
26. Sulistyosari, Y., Wardana, A., & Dwiningrum, S. I. (2023). School zoning and equal education access in Indonesia. International Journal of Evaluation and Research in Education, 12(2), 586 - 593.
27. Sutadi, Sumirah, D., Lisyani, & Agustina, D. (2025). Analysis of the School Zoning System and its Impact on the Education. PPSDP International Journal of Education, 4(1), 91-107.
28. Tibshirani,, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. Journal of the royal statistical society: series b (statistical methodology), 63(2), 411-423.