Journal of Mathematics and Computing Science, 11 (2): 50-60, 2025

eISSN: 0128-0767

Clustering Student Internship Placement Using Wards Hierarchical Method

Wan Faizah Wan Yaacob[1], Norafefah Mohamad Sobri[2], Wan Fairos Wan Yaacob[3]*, Nik Nur Fatin Fatihah Sapri[4] & W. Khairiyah Hulaini Wan Ramli[5]
[1,2,5]Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Kelantan, Bukit Ilmu, Machang, Kelantan, Malaysia
[3,4]Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia

wfaizah98@uitm.edu.my, noraf378@uitm.edu.my, wnfairos@uitm.edu.my*, nikfatinfatihah@uitm.edu.my & wkriyah@uitm.edu.my

*Corresponding author

Received **; Received in revised **; Accepted ***
Available online ***
DOI: https://doi.org/10.24191/jmcs.*******

Abstract: Internship programs are widely recognized as an essential graduation requirement in higher education, as they provide students with practical experience that extends beyond classroom learning. Selecting an appropriate internship placement is therefore a critical process in ensuring that students gain meaningful exposure and achieve the intended learning outcomes. Although students' internship locations are dispersed nationwide, the factors influencing their choice particularly whether they prefer placements near their hometown or areas with higher internship availability have remained unclear. To address this gap, this study applies Ward's Hierarchical Clustering method to identify patterns in internship placement preferences among Mathematics students from a university in Kelantan. The analysis reveals three distinct clusters: a high-preference area (Kelantan), medium-preference areas (Selangor, Terengganu, Johor, and Perak), and low-preference areas (Pulau Pinang, Pahang, Kedah, Kuala Lumpur, Perlis, Putrajaya, Melaka, Negeri Sembilan, and Sabah). The main contribution of this study lies in providing an evidence-based classification of internship placement hotspots using a systematic clustering approach. These findings offer valuable insights for university internship managers in strengthening industry networks and strategically aligning placement opportunities with student preferences, ultimately enhancing the effectiveness of the internship program.

Keywords: Internship, Mathematics student, Placement, Ward Hierarchical Clustering

1 Introduction

Internship program has been part of the curriculum and one of the graduation requirements in most higher education institution. The internship can be defined as the process that involves the students being employed in any organization for a period of time, while receiving academic credit for their contribution to the organization with the optionality of monetary compensation for the students [1]. This program has become increasingly essential and plays an important role in giving exposure on real working life and professional career to the students. In addition, the internship experience exposes students to practical skills, improves their social relationships, motivates future learning and enhances their social personality [2]. Apart from that, the internship program also provides a platform for the students to engage with the workplace by applying theories and knowledge learnt from universities and gaining more understanding in their area of study.

Acquiring and selecting an internship placement is an important process for each student before starting any internship program. In Malaysia, students are usually given the choice of selecting their place of internship from all over the country, with the guidance from universities' internship management system. A good and proper selection of internship placement will lead to a major positive impact on internship experience and also positive effect on students' future careers. On the other hand, poor internship experience can lead to difficulties in job hunting, which will cause a prolonged period of early unemployment [3].

Despite the importance of placement selection, little is known about the underlying patterns that shape students' geographical choices of internship locations. Two theoretical perspectives suggest potential determinants. First, Place-Attachment Theory posits that individuals develop emotional and social ties to familiar environments, implying that students may prefer internship placements closer to their hometowns or within regions they are comfortable navigating [4]. Second, Mobility Constraint Theory suggests that financial, social, and logistical limitations can restrict a student's ability to travel far from home, thereby influencing their choice of internship location. Both perspectives indicate that placement decisions are not random but shaped by meaningful patterns related to familiarity and mobility [5].

However, existing research has not empirically examined whether these theoretical tendencies actually appear in student placement behaviour. Specifically, there is a lack of studies that identify whether students' internship choices naturally cluster around familiar regions or states offering more internship opportunities. Traditional descriptive statistics can list popular states but cannot reveal whether these states form distinct preference groups, nor whether locality and opportunity patterns emerge across regions.

Therefore, the objective of this study is to uncover spatial patterns in internship preferences among Mathematics students using Ward's hierarchical clustering. The study aims to determine whether internships cluster around home states, industrial hubs, or university-local regions, and to provide insights for internship using 189 students of Bachelor of Science (Mathematics) from UiTM Cawangan Kelantan. The result of the analysis is able to give a better understanding on the concentration of the internship placements from all over the country as chosen by these students.

2 Literature Review

A Impact, Effectiveness and Influencing Factors of Internship Programs

Most of the previous research on student internship has adopted a predominately quantitative approach [6]. A study by Anjum [7] evaluated the impact of internship programs on professional and personal development of business students in Pakistan by using descriptive analysis and scale measurement analysis. The results of the study described that the internship programs have an impact on the professional growth and skills of the students, which affects their personal development, skills and capabilities. The findings are consistent with previous study from Gault et al. [8] which indicates significant early career advantages for undergraduates with internship experience, as well as positive impacts for marketing educators, university administrators and intern employers. Meanwhile, Isa et al. [9] conducted a study on a sample of 387 students to examine the effectiveness of internship program and the interns' performance from the employers' or industry's perspective. Their analysis was based on the descriptive statistics, reliability and correlation tests, where their findings stated that the overall interns' performance evaluated by the industries was categorized as 'Good', and the interns possessed many good attributes. Besides that, a few studies have also been done in determining the factors that influence the satisfaction level of students with respect to their internship program. By using correlation test and multivariate regression model, Hergert [10] found that there was a strong statistical correlation between perceived value of the internship with the students' demographic profile, the structure of the internship, and the connection to the students' career plans. On the other hand, Jawabri [11] analysed the data of 70 students from 5 different universities across UAE using descriptive analysis, factor analysis and inferential analysis procedures which resulted in eight factors of effectiveness in internship.

Overall, existing literature has been strong in evaluating internship impacts, effectiveness, and satisfaction, but it lacks an examination of how students select internship locations and whether systematic patterns underlie these choices. Most studies treat internship experience as uniform across settings, overlooking the influence of geographical preference, proximity, or opportunity distribution. Thus, a gap remains in understanding the clustering of internship placement choices, which is essential for institutions seeking to plan and allocate resources effectively.

The present study extends the literature by focusing not on internship outcomes, but on the geographical patterns of students' placement decisions, using Ward's hierarchical clustering to identify natural groupings of preferred locations. This approach provides a structural understanding of internship selection behaviour that has yet to be explored in earlier studies.

B Cluster Analysis

Cluster analysis has been widely used in many fields of study, such as in market research fields, epidemic studies, engineering, anthropology, environment, food science and many more. The concept of clustering is different from classification method, where cluster analysis is a multivariate technique used to sort a huge data set and place similar observations (objects) into the same group, which is called the cluster. The observations within each group are close to each other (similar observations), but the clusters themselves are dissimilar [12]. For example, Sya'iyah et al. [13] used the K-Means algorithms to cluster 724 student data into three groups (clusters) of students with characteristics of excellent performance, standard performance and underperformance. While Canlas [14] used Modified Case Based Reasoning method of data mining to develop a model for student's internship placement. Not much work has been done to cluster the internship placement using method of clustering.

Although various clustering techniques, including hierarchical and non-hierarchical methods, have been applied in different fields, the use of clustering to analyse internship placement patterns remains limited. Ward's hierarchical clustering is one of the most widely used methods due to its ability to form clusters by minimizing within-group variance, producing interpretable and stable groupings. Several studies highlight its versatility: Sapri et al. [15] applied Ward's method to classify dengue-endemic regions in Selangor, demonstrating its effectiveness in spatial pattern identification. Beckstead [16] used it to group nurses based on attitudinal similarities, showing its applicability to behavioural data. Marinova-Boncheva [17] applied it to stock market data to detect corporate groupings, illustrating their strength in financial domains. While each of these studies successfully showcased the technique's utility, none addressed educational or internship-related decision-making contexts.

A review of the literature reveals that although clustering has been extensively applied in many disciplines, very few studies have used clustering to examine students' internship placement preferences. Existing educational clustering studies typically focus on academic performance or satisfaction, whereas internship-related research tends to rely on descriptive or regression-based analyses. This leaves a gap in understanding whether students' internship choices form identifiable geographical clusters. The present study builds on prior work by applying Ward's hierarchical clustering specifically to the spatial distribution of internship placements. By doing so, it extends the use of clustering into a new domain internship preference mapping providing a methodological contribution and offering insights that prior studies have not explored.

3 Methodology

A Descriptive Statistics

The data used in this study were first explored by using descriptive analysis. The descriptive analysis was used to summarize the number of internship students by state, number of internship students by district of internship, and number of student's origin by state.

B Spatial Analysis

In this study, a spatial map was used to visualize the student's internship placement concentration. For this analysis, a spatial map was plotted to observe the distribution of student's internship placement and student's origin of each district. From the spatial map, we can observe the hotspot area of student's internship placement. The spatial map produced in the study was analysed using spatial, map and tmap packages in R software version 3.5.2. The spatial maps are produced with a variation of yellow and red colour in which the lightest yellow represents the lowest count of placement while the darkest red represents the highest count of placement in a specific district. The lightest yellow indicates a range of placement count between 0 to 5 placements while the darkest red indicates a range of more than 30 placements during the period of study. These visualisations allowed the identification of clear spatial hotspots that indicate districts with higher student internship concentration. In addition, the mapping component of this study was intended as a descriptive spatial visualisation rather than a formal spatial statistical analysis. Thus, no spatial autocorrelation or hotspot statistical tests (e.g., Moran's I or Getis Ord Gi*) were applied. Future work may extend this by incorporating spatial statistical testing to validate clustering patterns more rigorously.

C Euclidean Distance

Euclidean distance is widely adopted in spatial analysis because it provides the most direct and intuitive measure of geographic separation between two locations. When geographic coordinates are projected into a planar coordinate system (such as UiTM), Euclidean distance accurately represents real-world straight-line distance, making it theoretically coherent with the geometric principles underlying spatial data. Its mathematical simplicity allows for efficient computation, which is essential for methods such as hierarchical clustering, k-means clustering, and spatial proximity analysis that rely on repeated distance calculations. Furthermore, many geographic phenomena exhibit spatial autocorrelation, whereby nearby locations tend to share similar characteristics. There are many ways to measure statistical distance such as Canberra metric, Czekanowski coefficient, Minkowsi metric and Euclidean distance. However, Euclidean distance or also known as geometric matrix is the most frequently used in research. Euclidean distance represents the length of the hypotenuse of the right-angled triangle [18]. A straight-line distance between points as in Cartesian method is also a Euclidean distance. Euclidean distance calculation is based on the square root of the sum of the square differences between the origin and destination values for all variables [19]. In two-dimension plane, the observations are plotted by using scatter plot, and the distance between the pair of points can simply be measured. Generally, the Euclidean distance is calculated using the following equation:

$$D(\boldsymbol{a},\boldsymbol{b}) = \sqrt{(a_1-b_1)^2+(a_2-b_2)^2} \tag{1}$$

where;

a and ***b*** represent vectors

a_1, a_2, b_1 and b_2 represent the element of the observation vector ***a*** and ***b*** respectively

Consequently, it is clear that Eq. (1) is the same as the Pythagoras theorem. Hence, the Euclidean distance can be calculated as follows if the vectors contain *n* dimensions:

$$D\,(\boldsymbol{a},\boldsymbol{b}) = \sqrt{(a_1-b_1)^2+(a_2-b_2)^2+.......+(a_i-b_i)^2+(a_n-b_n)^2} = \sqrt{\sum_{i=1}^{n}(a_i-b_i)^2} \tag{2}$$

where i denotes points of coordinates.

Therefore, from Eq. (2), the squared Euclidean distance can be written as follows:

$$D^2\,(\boldsymbol{a},\boldsymbol{b}) = \sum_{i=1}^{n}(a_i-b_i)^2 \tag{3}$$

However, the distance does not behave simply as the formula stated above due to the geomorphological and historical characteristics. Using correction factors, the equation can be adjusted, such as by using regression analysis. The correction factor for the different distance and regions form brings the Euclidean and the actual distances closer with minimized error [20].

D Wards Hierarchical Linkage Clustering

Wards Hierarchical Linkage Clustering or also known as Wards Minimum Variance Method (MVM) is a method that suggests the decision on which pair of clusters to be joint should be based on the optimal value of an objective function [21 – 24]. The objective of Wards procedure is to unify groups where variance within them does not increase drastically and thus producing the most possible homogeneous clusters [21]. Wards' technique also has a characteristic of dividing the various stations into regions with approximately equal size. This characteristic is useful to ensure each group has the minimum number of stations required for the application of the appropriate regional estimation technique. The similarity metric, which is one of the components in the agglomerative algorithms of Wards' technique, is used to determine which pair of subsets will be merged. In this study, the Euclidean distance was used. It considers two clusters which are Cluster A and Cluster B. The sum of squares errors (SSE) due to the merge of two clusters are defined as follows:

$$SSE_A = \sum_{i=1}^{n_A}(\boldsymbol{a}_i - \bar{\boldsymbol{a}})'(\boldsymbol{a}_i - \bar{\boldsymbol{a}}) \tag{4}$$

$$SSE_B = \sum_{i=1}^{n_B}(\boldsymbol{b}_i - \bar{\boldsymbol{b}})'(\boldsymbol{b}_i - \bar{\boldsymbol{b}})$$

(5)

$$SSE_{AB} = \sum_{i=1}^{n_{AB}}(\boldsymbol{y}_i - \bar{\boldsymbol{y}}_{AB})'(\boldsymbol{y}_i - \bar{\boldsymbol{y}}_{AB}) \tag{6}$$

where;

$\boldsymbol{a}_i$ is the i^{th} observation vector and $\bar{\boldsymbol{a}}$ is the centroid of Cluster A
$\boldsymbol{b}_i$ is the i^{th} observation vector and $\bar{\boldsymbol{b}}$ is the centroid of Cluster B
$\boldsymbol{y}_i$ is the i^{th} observation vector and $\bar{\boldsymbol{y}}_{AB}$ is the centroid of Cluster AB

Hence, Wards' method calculates the distance between cluster members and the centroid by minimizing the sum of squared Euclidean distance between centroid point and each of the point of cluster members presented in Eq. (4) – Eq. (6). The sum of squared for all points inside a cluster divided by the number of points in the cluster is defined as the centroid. The first cluster is formed when the smallest sum of square from the pair of sample unit is computed. The cluster member is grouped such that the sum of squared error is minimized in each step of the procedure. When all objects are combined into a single large cluster, the procedure is stopped. Wards' method can be simply written as follows:

$$I_{AB} = \frac{n_A n_B}{n_A + n_B}(\bar{\boldsymbol{a}} - \bar{\boldsymbol{b}})'(\bar{\boldsymbol{a}} - \bar{\boldsymbol{b}}) \tag{7}$$

where;

$\bar{\boldsymbol{a}}$ and $\bar{\boldsymbol{b}}$ is defined as the centroids of Cluster A and Cluster B respectively.
n_A and n_B is defined as the size of Cluster A and Cluster B respectively.

The Wards hierarchical clustering is known to be the most effective and frequently used of hierarchical method [20]. In this study the method was performed using hclust() function with method = "ward.D2" in the stat package using R software. Mean of dendrogram is produced to visualize the results of clustering procedure. Dendrogram is a two-dimensional diagram that shows the hierarchical relationship between object, and it is created as an output from the hierarchical clustering. Dendrogram works as the best way to allocate objects to clusters. It illustrates the information in the amalgamation table in the form of a tree diagram. It is a network structure which constitutes a root node that produces several nodes connected by edges or branches. The last nodes of the hierarchy are called leaves. The distance or the dissimilarity between clusters is represented by the horizontal axis of the dendrogram. The objects and clusters are represented by vertical axis. Each fusion (joining) of two clusters is represented on the graph by the splitting of a horizontal line into two horizontal lines. The distance (dissimilarity) between the two clusters is shown by the short vertical bar. Dendrogram is the most useful and fairly simple to interpret as there are smaller number of cases [24].

E The Data

The study used secondary data of student internship placement of Bachelor of Science (Mathematics) from 2020 to 2021. The data were collected from Universiti Teknologi MARA Cawangan Kelantan, one of the universities located in the east coast Malaysia, with approval for the purpose of educational data mining. The data described the information about the placement of the student's internship which included the location of placement by district and state and the student's origin by state. The data consisted of 189 internship students that were placed at various states in Malaysia such as Kelantan, Pahang, Selangor and others. Longitudinal and latitudinal were also obtained to analyse the internship data for spatial mapping. Table 1 summarizes the description of the data.

Table 1**:** Description of the Data

Variable	Description	Unit
District	District of placement	District
State	State of placement	State
Origin	Student state of origin	State
Long	Longitudinal	°East
Lat	Latitudinal	°North

4 Result and discussion

A Descriptive Statistics

Table 2 reports the total number of student internships and their origin by state which is detailed out by specific district with the highest number of student internship. Based on Table 2, the highest total number of student internship placements was recorded in the state of Kelantan and majority of the student's origin appears to be from Kelantan followed by Selangor. Kota Bharu district appeared to be the district of Kelantan with the highest number of students' placement. It was followed by the states of Selangor and Terengganu with the highest figures of placement concentration situated in the districts of Petaling Jaya and Kuala Terengganu, respectively. This is consistent with the distribution of concentration by district as displayed in Figure 1. The map shows that student internship placements are unevenly distributed across Malaysia, with a noticeably higher concentration in Peninsular Malaysia compared to East Malaysia. Several districts particularly in the west coast states such as Penang, Selangor, Perak, and Johor display darker shades, indicating higher counts of student internships. In contrast, most areas in Sabah and Sarawak show very low participation, with nearly all districts appearing in light shades, suggesting minimal internship placements. Overall, the distribution pattern highlights that students tend to secure internships primarily in more urbanised and economically active regions of Peninsular Malaysia, while placements in East Malaysia remain limited.

Table 2: Summary on number of student internship and origin by state and district

State	Total student Origin	Total Student Internship	District with the highest number of student internship
Johor	20	19	Johor Bharu (6)
Kedah	9	11	Kuala Muda (4)
Kelantan	35	36	Kota Bharu (18)
Kuala Lumpur	9	11	Kuala Lumpur (11)
Melaka	3	4	Jasin (2)
Negeri Sembilan	4	5	Seremban (3)
Pahang	14	12	Kuantan (8)
Perak	21	21	Hilir Perak (5)
Perlis	2	2	Perlis (2)
Pulau Pinang	10	8	Seberang Perai Utara & Timur Laut (3)
Putrajaya	0	2	Putrajaya (2)
Sabah	5	5	Kota Kinabalu (2)
Selangor	28	28	Petaling Jaya (14))
Terengganu	27	25	Kuala Terengganu (9)
Sarawak	2	0	0
Labuan	0	0	0

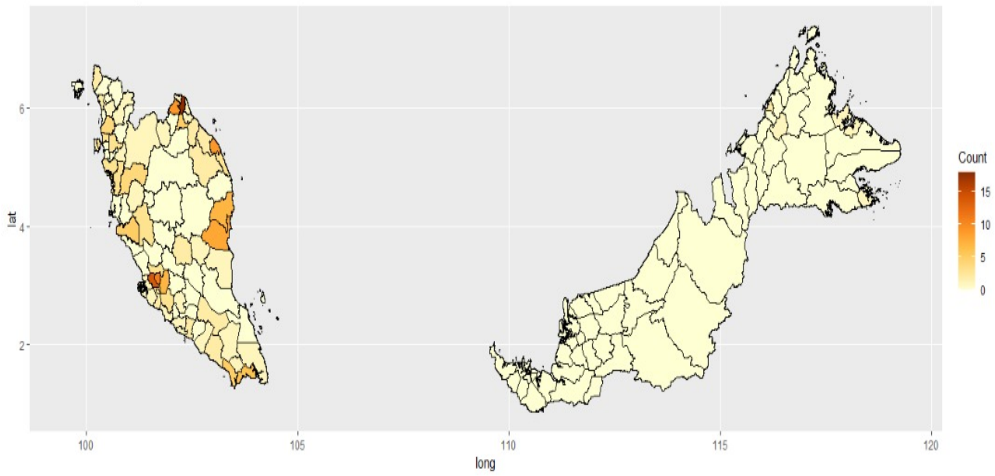

Figure 1: Distribution of student internship by district

B Spatial Analysis

This section reports the spatial map analysis of student internship placement concentration and the origin of the student by state. The map demonstrated the result of analysis that visualized the concentration of preference in the number of student internship across Peninsular Malaysia and Sabah and Sarawak across 2 years (2020 – 2021) by states. Figure 2(a) displays the spatial distribution of students' hometowns, while Figure 2(b) illustrates the distribution of their internship placements across Malaysian states. Both maps show a clear concentration in the east-coast and northern regions of Peninsular Malaysia. Kelantan records the highest number of students and also the highest number of internship placements, followed by neighbouring states such as Terengganu, Pahang, and Perak. In contrast, Sabah, Sarawak, Labuan, and Perlis show very low counts in both origin and placement, indicating limited student presence and minimal movement toward these regions.

The similarity in spatial patterns between students' origin and internship placement suggests that many students prefer to undertake their internships near their hometowns. This aligns with PlacenAttachment Theory, which proposes that individuals tend to remain in familiar environments, and with Mobility Constraint Theory, where financial, logistical, or familial factors may limit students' willingness to travel far from home. The higher placement counts in Selangor and Johor, despite lower origin counts, may reflect the attraction of states with greater internship availability.

These spatial patterns directly support the clustering results, which classify Kelantan as a high-preference area, followed by a medium-preference cluster comprising Terengganu, Pahang, Perak, Johor, and Selangor, while the remaining states fall into a low-preference cluster. The spatial maps therefore provide clear visual evidence that internship choices are not randomly distributed but form identifiable geographical groupings shaped by locality and opportunity factors.

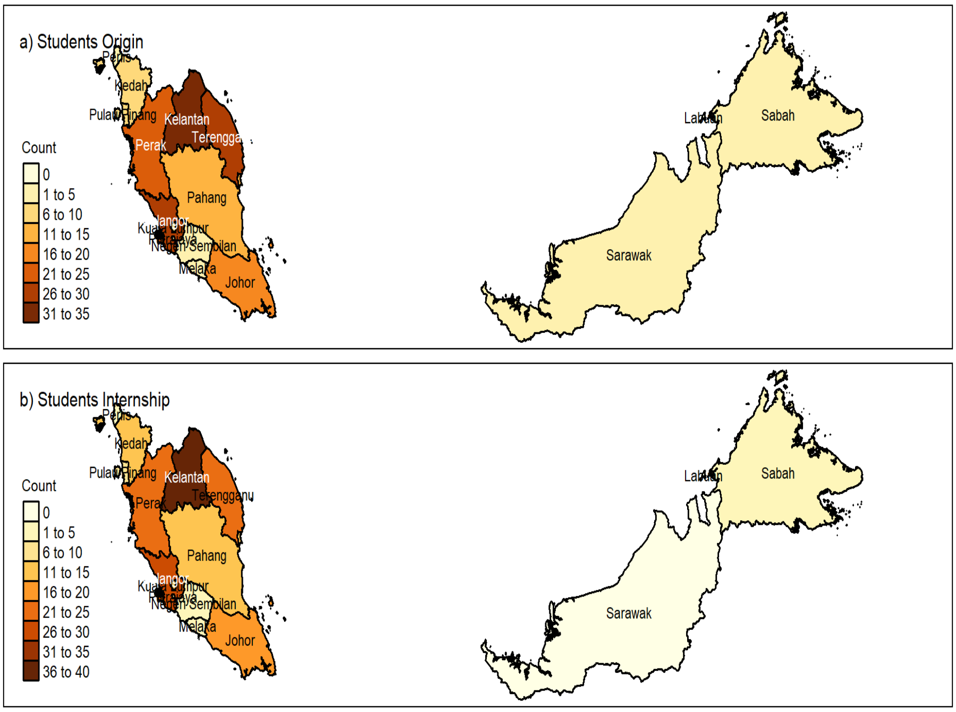

Figure 2: Spatial Map of (a) student origin and (b) student internship by State

C Clustering Analysis of Student Internship by Districts

This study clustered the student's internship placement using Wards Hierarchical Clustering method. The technique clustered the districts of internship placement based on minimum variance within the clusters with minimum increment of sum of square. Figure 3 displays the dendrogram of 14 states in Malaysia. The Ward's hierarchical clustering produced three distinct clusters of Malaysian states based on the similarity of internship placement patterns. Cluster 1 consists of the majority of Peninsular Malaysian states Pulau Pinang, Pahang, Kedah, Kuala Lumpur, Perlis, Putrajaya, Melaka, Negeri Sembilan and Sabah. This large grouping suggests that these states share comparable characteristics in terms of internship attractiveness, possibly due to moderate-to-high urban development, accessibility, and consistent availability of internship opportunities across multiple sectors.

Cluster 2, which contains only Kelantan, indicates that Kelantan's internship placement pattern is uniquely different from all other states. This may be influenced by strong hometown preferences among students from UiTM Kelantan, limited industrial sectors, and high familiarity with the local environment, aligning with Place-Attachment Theory.

Cluster 3 groups Johor, Perak, Selangor, and Terengganu, states known for stronger industrial presence and higher availability of internship placements. This cluster reflects opportunity-driven preference patterns, where students tend to choose locations with better industry exposure, supporting Mobility Constraint Theory, as students may travel to these states because the professional advantages outweigh travel constraints.

Overall, the clustering results demonstrate clear, non-random geographical patterns in internship selection, driven by familiarity, accessibility, and job opportunity structures.

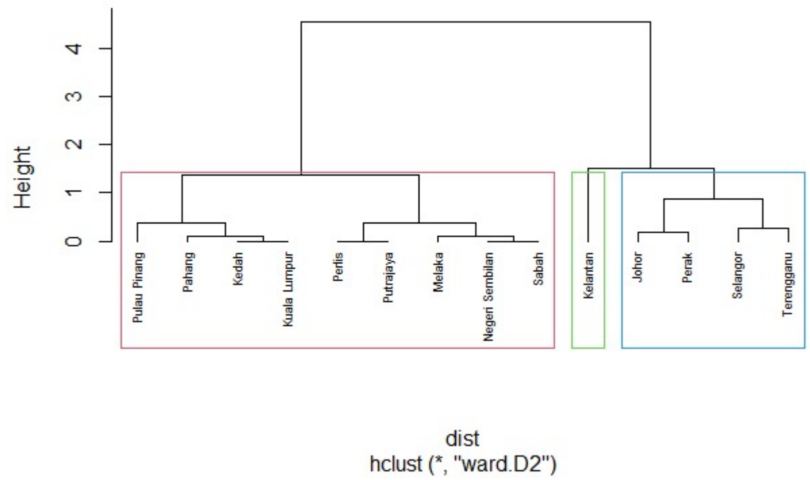

Figure 3: Dendrogram of Wards Hierarchical Clustering for Student's Internship Placement

Based on the findings of dendrogram, it can be clearly seen the number of clusters that made up the choice of Mathematics student's internship placement consisted of three clusters which were Cluster 1 (High internship placement area), Cluster 2 (Medium internship placement area) and Cluster 3 (Low internship placement area). The list of states according to clusters obtained is shown in Table 3 below. This finding is consistent with what has been portrayed by spatial mapping in Table 3.

Table 3: Cluster of student internship by state.

High internship placement area	Medium internship placement area	Low internship placement area
Kelantan	Selangor Terengganu Johor Perak	Pulau Pinang Pahang Kedah Kuala Lumpur Perlis Putrajaya Melaka Negeri Sembilan Sabah

5 Conclusion

This study provides an empirical examination of internship placement preferences among Mathematics students, highlighting the spatial patterns that shape decision-making. Using Ward's Hierarchical Clustering, three distinct clusters were identified: Kelantan as a high-preference state, Selangor, Terengganu, Johor, and Perak as medium-preference states, and the remaining regions as low-preference areas. The findings underscore two key determinants of student choices: proximity to home or university, consistent with Place-Attachment Theory, and the availability of industry opportunities, in line with Mobility Constraint Theory. By integrating spatial mapping with hierarchical clustering, this research contributes to the limited body of literature on geographically clustered internship preferences. The results carry practical implications for university internship management. Strengthening partnerships in high-demand regions such as Kelantan and Selangor can optimize placement opportunities, while diversifying options in medium-preference states may reduce competition. Furthermore, identifying low-preference regions offers opportunities to enhance student awareness and logistical support, thereby broadening the scope of industrial collaboration. Overall, the study demonstrates that internship choices are structured and predictable, shaped by both personal attachments and regional economic landscapes.

6 Limitations and Future Recommendations

This study is limited by its relatively small sample size and focus on Mathematics students, which restricts the generalizability of the findings. The reliance on Euclidean distance and Ward's hierarchical clustering, while effective for identifying broad spatial patterns, may not fully capture the complexity of internship decision-making. Future research should expand the dataset to include students from multiple programs and employ more advanced analytical techniques such as machine learning clustering, spatial autocorrelation, or models incorporating socioeconomic variables. Such approaches would enhance the robustness of spatial analysis and provide deeper insights into the determinants of internship placement preferences.

Acknowledgements

We would like to express our appreciation to the faculty of Mathematical & Computer Sciences, Universiti Teknologi MARA Cawangan Kelantan for the support in providing the data. The authors are also grateful to the Editor, Associate Editor, and anonymous referees for the helpful comments and suggestion

Conflict of Interest Statement

The authors declare that this research was conducted without any personal, commercial, or financial conflicts of interest, and that no conflicts exist with the funder.

References

[1] Y. C. Fei, W. Waheeb and S. B. Sulaiman, "Student internships: A selective review", *Journal of Applied Technology and Innovation (e-ISSN: 2600-7304)*, vol. 4, no. 2, pp. 1 – 5 , 2020.

[2] D. Calloway and S. M. Beckstead, "Reconceptualizing internships in management education", *Journal of Management Education*, vol. 19, no. 3, pp. 326 – 41, 1995.

[3] A. Cerulli-Harms, "Generation internship: the impact of internships on early labour market performance", *I Z A Institute of Labour Economics*. IZA DP No. 11163, 2017.

[4] M. Lewicka (2011). Place attachment: How far have we come in the last 40 years?. *Journal of Environmental Psychology*, 31(3), 207-230.

[5] T. Schwanen, K. Lucas, N. Akyelken, D. Solsona, D.C. , J. A. Carrasco & T. Neutens, (2015). Rethinking the links between social exclusion and transport disadvantage through the lens of social capital. *Transportation Research Part A: Policy and Practice*, 74, 123-135.

[6] A. L. H. Cheong, N. Yahya, Q. L. Shen and A. Y. Yen, "Internship experience: An in-depth interview among interns at a business school of a Malaysian private higher learning institution", *Procedia-Social and Behavioral Sciences*, vol. 123, pp. 333 – 43, 2014.

[7] S. Anjum, "Impact of internship programs on professional and personal development of business students: A case study from Pakistan", *Future Business Journal*, vol. 6, no. 1, pp. 1 – 13 , 2020.

[8] J. Gault, J. Redington and T. Schlager, “Undergraduate business internships and career success: are they related?”, *Journal of marketing education*, vol. 22, no. 1, pp. 45 – 53, 2000.
[9] M. A. M. Isa, Z. Zaharum, N. A. Ramdhan, M. M. Malom and F. Nasrul, “The internship students’ performance from the perspective of industry: Case of Islamic banking students”, *Journal of Islamic*, vol. 5, no. 28, pp. 111 – 121 , 2020.
[10] M. Hergert, “Student perceptions of the value of internships in business education”, *American Journal of Business Education (AJBE)*, vol. 2, no. 8, pp. 9 – 14, 2009.
[11] A. Jawabri, “Exploration of internship experience and satisfaction leading to better career prospects among business students in UAE”, *American Journal of Educational Research*, vol. 5, no. 10, pp. 1065 – 1079, 2017.
[12] A. F. Alkarkhi and W. A. Alqaraghuli, *Easy statistics for food science with R*, Academic Press, 2018.
[13] K. Sya’iyah, H. Yuliansyah and I. Arfiani, “Clustering student data based on K-means algorithms”, *Int. J. Sci. Technol. Res*, vol. 8, no. 8, pp. 1014 – 1018, 2019.
[14] F. Q. Canlas, Data Mining Model for Student Internship Placement Using Modified Case Based Reasoning. In: Masrour T., El Hassani I., Cherrafi A. (eds) Artificial Intelligence and Industrial Applications. A2IA 2020. *Lecture Notes in Networks and Systems,* vol. 144, pp. 260 – 269, Springer, Cham. https://doi.org/10.1007/978-3-030-53970-2_24, 2021
[15] N. N. F. F. Sapri, W. F. W. Yaacob, Y. B. Wah and S. S. S. A. Rahim, “Spatio-temporal clustering of dengue incidence”, *Universal Journal of Public Health,* vol. 9, no. 3, pp. 120 – 130, 2021.
[16] J. W. Beckstead, “Using hierarchical cluster analysis in nursing research”, *Western Journal of Nursing Research*, vol. 24, no. 3, pp. 307 – 319, 2002.
[17] V. Marinova–Boncheva, “Using the agglomerative method of hierarchical clustering as a data mining tool in capital market”, *International Journal Information Theories & Applications*, vol. 15, pp. 382 – 386 , 2008.
[18] T. Löster, “The evaluation of CHF coefficient in determining the number of clusters using Euclidean distance measure”, *The 8th International Days of Statistics and Economics*, 2014.
[19] D. N. S. Gonçalves, C. de Morais Gonçalves, T. F. de Assis and M. A. da Silva, “Analysis of the difference between the euclidean distance and the actual road distance in Brazil”, *Transportation Research Procedia*, vol. 3, pp. 876 – 885, 2014.
[20] P. Govender and V. Sivakumar, “Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019)”, *Atmospheric Pollution Research*, vol. 11, no. 1, pp. 40 – 56, 2020.
[21] W. A. Pansera, B. M. Gomes, A. V. Boas and E. L. Mello, “Clustering rainfall stations aiming regional frequency analysis”, *Journal of Food, Agriculture & Environment*, vol. 11, no. 2, pp. 877 – 885, 2013.
[22] S. Sreedhar Kumar, M. Madheswaran, B. A. Vinutha, H. Manjunatha Singh and K. V. Charan. “A brief survey of unsupervised agglomerative hierarchical clustering schemes”, *Int J Eng Technol*, vol. 8, no. 1, pp. 29 – 37, 2019.
[23] T. Strauss and M. J. von Maltitz, “Generalising Ward’s method for use with Manhattan distances”, *PloS one*, vol. 12(1), e0168288, 2017.
[24] NCSS,LLC. NCSS Statistical Software. NCSS.com [online] Chapter 445 Hierarchical Clustering/Dendrograms, http://ncss.wpengine.netdna-cdn.com/wpcontent/themes/ncss/pdf/Procedures/NCSS/Hierarchical_ClusteringDendrograms.pd