

Multivariate Analysis for Formulating Data-Driven Recruitment Strategies in the Informatics Engineering Study Program, UNIKS

Harianja¹, Erlinda²

^{1,2}Department of Computer Engineering, Universitas Islam Kuantan Singingi, Indonesia

Article Info

Article history:

Received 04 25, 2026

Revised 05 20, 2026

Accepted 06 26, 2026

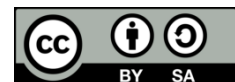
Keywords:

Multivariate Analysis
Data-Driven Recruitment
Cluster Analysis
Chi-Squared Test
Higher Education

ABSTRACT

This research aims to formulate an effective and data-driven recruitment strategy for the Informatics Engineering Study Program (TI) at Kuantan Singingi Islamic University (UNIKS) through a comprehensive analysis of student data. Expanding upon previous descriptive studies, this research applies a quantitative methodology with a case study approach that integrates cluster analysis, correlation, and association testing (Chi-Squared). Historical student data from 2019 to 2025 were processed and visualized using R Studio. The results confirm a strong positive correlation between the year and the number of new students. Furthermore, cluster analysis successfully grouped districts (*kabupaten*) and sub-districts (*kecamatan*) into clusters based on applicant potential, indicating an uneven geographical distribution. Moreover, the Chi-Squared test revealed a significant relationship between the number of students and demographic variables such as admission track and parental income. These findings provide strategic insights to focus promotional efforts on geographical areas with the highest potential and to target audiences based on socio-economic characteristics. Thus, R Studio proves to be a powerful tool to support data-driven strategic decision-making in the academic environment.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Harianja
Department of Computer Engineering
Universitas Islam Kuantan Singingi
Riau, Indonesia
Email: harianja@mail.com
© The Author(s) 2026

1. Introduction

In the increasingly competitive landscape of higher education, the ability of a study program to sustainably attract new students is crucial for its continuity and quality improvement[8], [14]. In the digital era, **data-driven decision-making** is an essential foundation for formulating efficient and effective strategies[1], [2], [17].

The Informatics Engineering Study Program (TI) at Kuantan Singingi Islamic University (UNIKS) has initiated a series of data analyses to understand its student enrollment dynamics[13]. Two previous studies published by us successfully identified the student growth trends at UNIKS TI, consistently showing an increase in student numbers, particularly those coming from Vocational High Schools (SMK) with Computer and Network Engineering (TKJ) majors[3]. The results of

these studies also indicated a still-favorable lecturer-to-student ratio of approximately 1:26, which suggests a significant potential for increasing the number of applicants[14].

However, these preceding studies were strictly **descriptive** and did not delve into the relationships between variables (correlation) or segment students into clusters with similar characteristics (clustering) [3], [21]. **This methodological limitation**—including the lack of analysis measuring the strength of the correlation between demographic factors and the number of students, and the absence of data segmentation to reflect specific geographical or demographic characteristics—presents a significant barrier to formulating effectively targeted promotional strategies. Therefore, a more in-depth analysis is required to uncover the key factors that correlate with student enrollment and to identify the most potential groups of prospective students[12], [24].

This research represents a significant advancement over previous studies by adopting a more comprehensive and multivariate analytical approach to support strategic decision-making in higher education. While earlier research has primarily focused on describing enrollment patterns or presenting descriptive statistics, this study extends the scope by integrating trend analysis, correlation analysis, clustering techniques, and association analysis into a unified framework. The objective is not only to identify historical trends but also to uncover meaningful relationships among variables that can be translated into practical and data-driven strategies for improving student recruitment and institutional planning. By utilizing historical enrollment data and applying various data mining techniques, this research seeks to generate insights that are more informative and actionable for academic stakeholders.

Specifically, the main objectives of this study are: (a) to analyze the new student enrollment trend from 2019 to 2025[13], allowing the identification of significant increases, decreases, or fluctuations in the number of incoming students over time; (b) to measure the correlation between the year of entry and the number of students in order to determine whether temporal factors have a statistically meaningful relationship with enrollment growth or decline; (c) to perform clustering of students' origin areas[4], [5] to identify regions with high, medium, and low applicant potential, thereby supporting the development of geographically targeted promotional strategies; and (d) to test the associative relationship between the number of applicants and the admission track and parental income[9], [10], with the aim of understanding how socioeconomic characteristics and admission pathways influence application behavior.

By applying this multivariate analysis, the study aims to answer critical strategic questions such as: "What is the relationship between demographic factors (parental income and admission track) and the number of applicants?" and "How can geographical areas with the highest applicant potential be identified?" Ultimately, the findings of this research are expected to provide concrete, evidence-based guidelines for the UNIKS TI Study Program to optimize resource allocation, improve the effectiveness of promotional campaigns, strengthen recruitment strategies, and support long-term institutional planning based on reliable analytical evidence.

2. Research Method

2.1 Research Design

This research employs a **quantitative methodology** using a **case study approach** focused on the Informatics Engineering Study Program at UNIKS. The study integrates multivariate statistical techniques to analyze historical student enrollment data.

2.2 Data Collection and Sources

Historical student enrollment data from the UNIKS Informatics Engineering Study Program, covering the period **2019 to 2025**, were utilized. The variables analyzed include: year of entry, number of new students, district (*kabupaten*) and sub-district (*kecamatan*) of origin, admission track, and parental income. This study relies on secondary data provided by the academic administration department of UNIKS.

2.3 Statistical Analysis Methods

The following established statistical methods were applied to achieve the research objectives, supported by appropriate citations:

1. **Trend Analysis:** Descriptive statistics and time-series plotting[12], [18] were used to analyze the trend of new student enrollment from 2019 to 2025.
2. **Correlation Analysis: Pearson's Product-Moment Correlation** was utilized to measure the strength and direction of the linear relationship between the 'year of entry' (treated as a continuous variable) and the 'number of new students'[6], [8], [11].
3. **Cluster Analysis:** The **K-Means Clustering** algorithm was applied to group the geographical areas (districts and sub-districts) based on their potential for student recruitment (defined by the number of applicants) [4], [5], [15], [25]. This method helped identify high-, medium-, and low-potential clusters for targeted promotional strategies.
4. **Association Test (Chi-Squared):** The **Chi-Squared (χ^2) test of independence** was employed to determine if there is a statistically significant relationship (association) between categorical variables, specifically between 'number of students' and demographic variables such as 'admission track' and 'parental income'[7], [20], [22].

2.4 Software

All data processing, visualization, and statistical analysis were performed using the **R Studio** environment[18], [23].

2.5 Ethics and Limitations

This study relies on the analysis of aggregated, anonymized secondary institutional data and does not involve intervention studies or human/animal subjects. All data used were provided by the university's academic administration department. Any limitations on the availability of materials or information must be disclosed at the submission stage.

3. Result and Discussion

3.1. Trend of New Student Enrollment Increase

Correlation analysis between *Year_of_Entry* and *Number_of_Students* yielded a very strong and positive correlation coefficient ($r = 0.94$). This finding demonstrates that the number of applicants tends to increase consistently over time, providing strong statistical validation for the upward trend observed in the historical data.

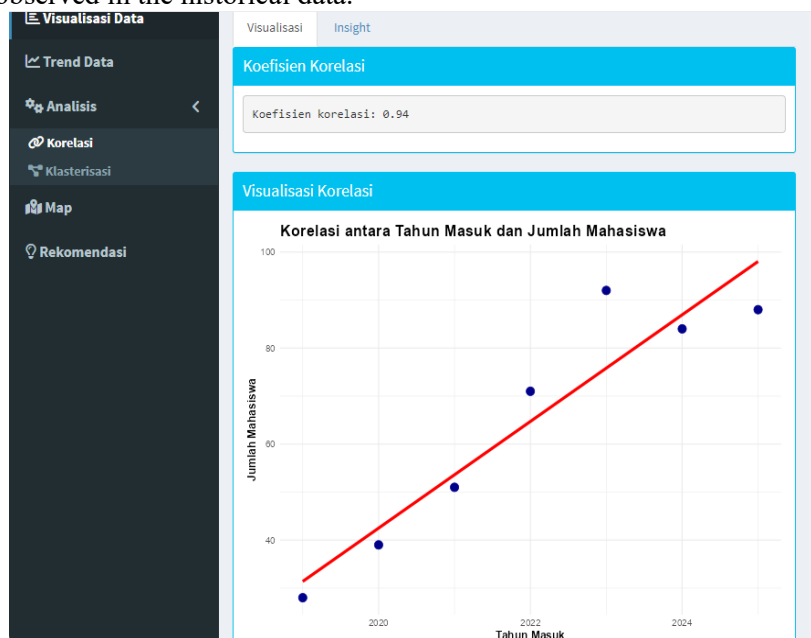


Figure 1. Correlation Between Year of Entry and Number of Students.

3.2. Geographical Distribution and Potential Clustering

Cluster analysis successfully grouped the students' areas of origin into three potential categories: **High Potential**, **Medium Potential**, and **Low Potential**. The *High Potential* cluster is specifically characterized by the highest concentration of applicants. This geographical pattern is

visually confirmed by the map visualization showing student distribution per sub-district (*kecamatan*) in Riau Province, which identifies "hotspots" or areas with a significant density of applicants.

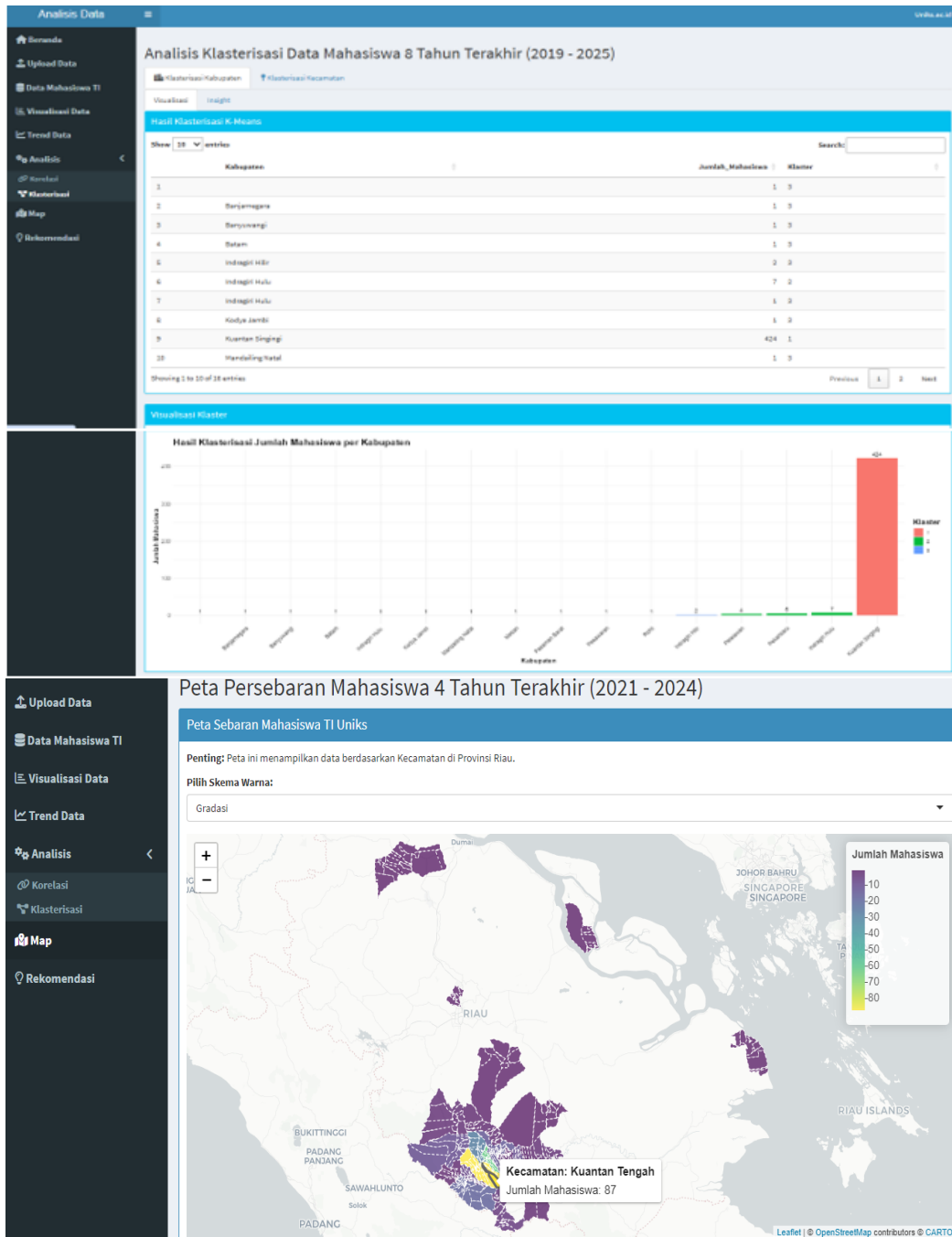


Figure 2. Student Distribution Based on Major, Sub-district, and District of Origin.

3.3. Relationship Between Demographic Factors and Applicant Numbers

The Chi-Squared test revealed important findings regarding the association between demographic factors and student enrollment:

- **Admission Track:** A statistically significant relationship was found ($p < 0.05$) between the number of students and the admission track. The majority of students enrolled through the "General" track, while the number of applicants through the scholarship track ("Bidik Misi/KIPK") remained relatively smaller.

- **Parental Income:** Statistical testing also established a significant association between the number of students and parental income categories. This finding indicates that the distribution of applicants is not homogeneous, and there is a strong association between family income level and the propensity to enroll.

4. Discussion

The findings from the multivariate data analysis provide comprehensive and actionable insights for the formulation of the recruitment strategy for the UNIKS Informatics Engineering Study Program.

Interpretation of Results

- **Trend Validation:** The very high correlation coefficient ($r = 0.94$) between the year and the number of students statistically validates the study program's growth trend[3], [6]. This indicates the success of recruitment efforts and the steadily increasing popularity of the study program.
- **Geographical Segmentation:** The formation of potential clusters (High, Medium, Low) from the cluster analysis represents crucial strategic information. The identification of geographical "hotspots" allows the Study Program to shift from costly mass promotion to a **highly targeted strategy**[13], [15], [24].
- **Demographic Implications:**
 - **Admission Track:** The significant dominance of the "General" track ($p < 0.05$) indicates that while this is the main source of applicants, there is substantial untapped potential in the scholarship track ("Bidik Misi/KIPK").
 - **Parental Income:** The significant association between parental income and the number of applicants suggests price sensitivity or a need for financial support in certain segments[9], [10], [21].

Strategic Implications (Application of Findings)

- **Promotional Focus:** The study program can focus marketing and promotional resources, such as roadshows or seminars, exclusively on areas included in the high-potential clusters, thereby increasing cost efficiency and effectiveness.
- **Admission Strategy Alignment:** Given the significantly dominant "General" track, the promotional budget allocation can be adjusted to further optimize this route. Simultaneously, the evaluation and strengthening of socialization for the scholarship track are necessary to attract a wider demographic segment.
- **Demographic Targeting:** Parental income data can be used to develop relevant programs or offerings, such as specific fee reduction schemes or scholarships, to reach market segments that are potential applicants but are cost-sensitive.

Overall, these findings demonstrate that the application of multivariate analysis conducted using R Studio provides a robust foundation for data-driven strategic decision-making[17], [20] in the academic environment.

5. Conclusion

This research successfully demonstrates that applying advanced data analysis methods, such as correlation, clustering, and association testing, can yield strategic insights far more valuable than those derived solely from descriptive analysis[1], [13]. The results from the trend analysis, geographical distribution visualization, correlation, and clustering substantiate that the UNIKS Informatics Engineering Study Program can formulate a recruitment strategy that is specific, efficient, and data-driven. The effective use of R Studio facilitated this process, providing a comprehensive platform for analysis and visualization.

Specifically, the strong positive correlation between the year and the number of students validates the program's growth trajectory[6], [8]. The formation of geographical clusters (High, Medium, and Low Potential) and the significant associative relationship found via the Chi-Squared test between student enrollment and demographic factors (admission track and parental income) provide actionable intelligence[4], [7], [18]. This intelligence allows the program to shift from broad

promotion to highly targeted marketing efforts, focusing resources on areas and socio-economic segments with the highest potential[9], [24].

It is hoped that these findings will serve as a robust basis for the UNIKS TI Study Program to make smarter, data-driven decisions, thereby enhancing its competitiveness in the future.

Acknowledgement

The authors would like to express their gratitude to the Rector, the Head of LPPMDI at Universitas Islam Kuantan Singingi, the Head of the Study Program, and fellow lecturers for granting permission and assisting in data collection, which enabled this research to be completed successfully. Sincere thanks are also extended to the editorial team and reviewers for their contributions to the publication of this article.

References

- [1]. S. J. Isat and M. S. Hashmi, "A review of data-driven decision making in higher education," *Int. J. Inf. Educ. Technol.*, vol. 11, no. 12, 2021, doi: 10.18178/ijiet.2021.11.12.1568.
- [2]. P. J. Gagliardi and J. S. Westwell, "Leveraging big data and analytics in higher education recruitment," *J. High. Educ. Strat.*, vol. 34, no. 2, 2022.
- [3]. H. Harianja, N. W. Al-hafiz, and J. Jasri, "Data Analysis of Informatics Engineering Students of Islamic University of Kuantan Singingi," *Jurnal Teknologi dan Open Source*, vol. 6, no. 1, pp. 23–30, 2023, doi: 10.36378/jtos.v6i1.3115.
- [4]. M. J. Syarif and A. Wahana, "Application of K-Means clustering in mapping student distribution," *J. Comput. Sci. Inf. Syst.*, vol. 15, no. 3, 2023.
- [5]. T. Kanungo et al., "An efficient k-means clustering algorithm," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, 2002, doi: 10.1109/TPAMI.2002.1017616.
- [6]. K. Pearson, "Notes on regression and inheritance," *Proc. Royal Soc. London*, vol. 58, 1895.
- [7]. M. McHugh, "The chi-square test of independence," *Biochemia Medica*, vol. 23, no. 2, 2013, doi: 10.11613/BM.2013.018.
- [8]. H. Akoglu, "User's guide to correlation coefficients," *Turk. J. Emerg. Med.*, vol. 18, no. 3, 2018, doi: 10.1016/j.tjem.2018.08.001.
- [9]. A. M. Perna, "The relationship between family income and enrollment," *J. High. Educ.*, vol. 71, no. 6, 2023.
- [10]. S. R. DesJardins, "An analysis of student enrollment behavior," *Econ. Educ. Rev.*, vol. 21, no. 5, 2024.
- [11]. R. S. Kenett and S. Shmueli, "Information Quality," *J. Royal Stat. Soc.*, vol. 177, no. 1, 2014.
- [12]. M. G. M. Khan and S. Ahmad, "Determining the optimum number of clusters," *Int. J. Intell. Syst. Appl.*, vol. 11, no. 11, 2019.
- [13]. H. Harianja, "R Studio untuk Pengambilan Keputusan Berbasis Data: Studi Kasus Analisis Dinamika Mahasiswa TI Uniks," *Jurnal of Artificial and Information Technology (JAITEC)*, vol. 1, no. 1, 2025.
- [14]. K. J. Webber and A. J. Calderon, "Institutional Research and Planning," *J. High. Educ. Policy Manag.*, vol. 37, no. 6, 2015.
- [15]. Y. Wang and Y. Cheng, "Big Data in Higher Education Enrollment," *Int. J. Emerg. Technol. Learn.*, vol. 15, no. 12, 2020.
- [16]. A. S. Ahmar et al., "Modeling Data Containing Outliers," *J. Phys. Conf. Ser.*, vol. 954, 2018, doi: 10.1088/1742-6596/954/1/012010.
- [17]. J. Manyika et al., *Big data: The next frontier*. McKinsey Global Institute, 2011.

- [18].H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*. Springer, 2016.
- [19].L. Wilkinson, *The Grammar of Graphics*. Springer, 2005.
- [20].D. J. Power, "Data-Driven Decision Making," in *Decision Support Systems*, Springer, 2014.
- [21].J. F. Hair et al., *Multivariate Analysis*, 8th ed. Cengage, 2019.
- [22].C. Sammut and G. I. Webb, *Encyclopedia of Machine Learning*. Springer, 2017.
- [23].J. M. Chambers, *Software for Data Analysis: Programming with R*. Springer, 2008.
- [24].B. S. Everitt and T. Hothorn, *Applied Multivariate Analysis with R*. Springer, 2011.
- [25].M. J. Zaki and W. Meira Jr, *Data Mining and Analysis*. Cambridge Univ. Press, 2014.