

Identifying key patterns of college student's background through exploratory data analysis

Sitti Rahmah Jabir¹, Herdianti²

^{1,2} Faculty of Computer Science, Universitas Muslim Indonesia, Makassar, Indonesia

Article Info

Article history:

Received Feb 9, 2026

Revised Mar 5, 2026

Accepted Mar 10, 2026

Keywords:

Big Data;
Business Analytics;
Exploratory Data Analysis;
Missing Values;
Student's College Background.

ABSTRACT

The declining of student interest had forced universities to examine the characteristics of each student. According to higher education statistics on the number of new students, fluctuating values have been found in recent years. Several research used exploratory data analysis (EDA) approach to analyze new student admissions data. EDA is offered a summary of the dataset analysis and preliminary findings. There are variables decided to be dropped because consisted high number of missing values. On the other hand, some data filled with mean and mode because the number of missing not more than 20%. The missing values in each of attribute might be cleaned using another way. The admission team in university might encourage the registrants to complete and input correct data to the system. Based on the visualization, we found that some college students applied to university from several background of area, demographic and etc. The marketing division might apply another strategy is area had small number of college which is Kalimantan. Public health, computer science and insutry technology are major that have potential to be promoted due to the job prospects.

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



Corresponding Author:

Sitti Rahmah Jabir,
Faculty of Computer Science,
Universias Muslim Indonesia,
Jl. Urip Sumoharjo Km. 5, Makassar, Sulawesi Selatan, 90231, Indonesia.
Email: rahmahjabir@umi.ac.id

Introduction

The declining of student interest had forced universities to examine the characteristics of each student. According to higher education statistics on the number of new students, fluctuating values have been found in recent years. According from data statistics, number of student has been declined from 2022 to 2023. The data showed 58.290 students joined the university and followed by next year was 51.525. Furthermore, several factors contributing to the decline in student interest in continuing their studies included the COVID-19 pandemic, the prolonged implementation of the independent student admissions pathway, the large quota requirement, and economic factors (Universitas Ahmad Dahlan, 2023). Several studies have conducted demographic analyses to analyze prospective students' interest in higher education. One researcher conducted a study analyzing student enrollment using a big data approach. The results of this study assisted schools to increase the intensity of digital promotion and identify the majority of applicants through demographic analysis. Big data analytics capabilities (BDAC) is a constitutes a key determinant of decision-making quality, encompassing an organization's proficiency in leveraging technological infrastructure, human competencies, and data governance

mechanisms to gain competitive advantage. Through the application of advanced statistical methods to large-scale datasets, BDA facilitates performance enhancement and supports the generation of economic value (Dang et al., 2025). Furthermore, another study used an exploratory data analysis (EDA) approach to analyze new student admissions data. This study focused on analyzing student data features such as gender, date of birth, region of origin, study program, and others. The goal was to identify features that could be used for modeling in subsequent research (Septiani & Iqbal, 2024).

Conceptually, exploratory data analysis is the initial stage in data analysis, using software or automation to display data and identify relationships between different variables or features (Jabir et al., 2022). Related research analyzed the decision-making approach to college using a random forest modeling approach (Jabir et al., 2024). Other researchers used visualization data to understand prospective new student data, where the results of the research were in dashboard to monitor new student admissions in the form of visual information. The visualization results helped to improve administrative and promotional activities (MZ et al., 2022). In this study, the limitation is only applied for Student in Muslim University of Indonesia (UMI). The college student's will be analysed the student enthusiasm. UMI is a university that was founded in 1953. Where it is known that UMI is the largest university in Eastern Indonesia (UMI, n.d.). Based on references to several previous studies, researchers will apply Big Data and Exploratory Data Analysis (EDA) to determine the characteristics of students entering PTS, especially UMI. Exploration of this data is expected to optimize admission strategies and increase the number of new students (MABA). This study aims to analyze student admissions data using a big data approach and exploratory data analysis (EDA) to provide strategies for increasing student enthusiasm. The final result of the study is to see features that can be used for modeling in further research.

Method

The data was processed using R-Studio and Tableau software. The data collected spanned several years, and therefore, an integration process was performed using R Studio and missing data were cleaned. The data exploration process can be used to examine the characteristics of each new student who has enrolled at UMI since 2023 until 2024. The data obtained from university's department which contained 16.167 rows and 52 attributes. The data will explored used R Studio and Tableau tools specifically Tableau Desktop. Data integration and data pre-processing will be applied using R-studio and Tableau will be utilized for producing visualizations and viewing insight of the data. The interactive visualizations provided make it easy for users to view information from the data and help in identifying trends, identifying opportunities, and making data-driven decisions.

This research applied some research steps, which were:

1. Problem Analysis: Observing the problem. Observing the declining interest in new student admissions.
2. Data Collection: Data was obtained from UMI's internal data, taking new student admissions data from 2023-2024.
3. Data Integration: Integrating several collected data sets into one to facilitate analysis. Some missing values will be dropped from the data.
4. Data Pre-processing: the missing data will be dropped from the data. Complete case analysis (CCA) is a method to delete column which contain all missing values in each row (Popoola et al., 2025). Complete case analysis (CCA) is a method to delete column which contain all missing values in each row (Popoola et al., 2025). The CCA is chosen because some attributes had 100% missing values.
5. Exploratory Data Analysis: Exploring the combined dataset by examining several available features and presenting them in visualizations.
6. Evaluation: The visualization results will be analyzed to determine student characteristics and assist in developing strategies to increase student engagement. Result will combine with previous studies.

According to the implementation method, it attached with some previous work such as the study using title Student Enrollment Trend Analysis Using Big Data, the result showed that the

vocational high school program had seen a 15% increase in enrollment annually, while the high school program has seen a 20% decrease in the last year. The majority of enrollees were from the Medan area (Septiani & Iqbal, 2024). Other research applied exploratory data analysis (EDA) to analyze student data, the final viewed 14 features were recommended for further analysis. Other features with missing values, imbalances, and outliers were not recommended for use (Griha Tofik Isa et al., 2023). According to that research, it also will be applied to this research whether understood the important variables to include. Similar research using specific object, EDA was used to understand pattern of gender and team-based which focused on descriptive and comparative analyses. Insights included the relationships between age, physical characteristics, and activity status, highlighting trends such as greater physical activity among lighter individuals and team-specific differences in gender composition. The study also developed knowledge of R programming and critical thinking (Katyul et al., 2025). On the other hand, other research used another method to understand student data. Machine learning was applied. The demographic variables were analysed to understand the effect of the variables to the student's performance in positive and negative both the aspects (Srivastava et al., 2023). In addition, Business Intelligence also applied by another research. They utilized the method to understand new college student data and found that women were the majority of gender which contributed more in that year admission. Other than that, area and age were other variables which can be considered to next strategy for promotion (Paramita & Ramadhan, 2024). According to the previous research, this study will apply exploratory data analysis approach to understand the college student admission data in UMI for better promotion in the future enrollment.

In this study, the data will be analyzed. The initial step in the data analysis process is to select the appropriate data visualization tools that will assist in addressing the business questions identified earlier. The next step involves analyzing the data visualization outcomes and assessing their effectiveness. The visualization results will be reviewed to provide valuable information and insights into the business questions established during the project planning phase. The final step is to confirm that the data visualization aligns with and fulfills the business objectives (Tri et al., n.d.). By analyzing the data, organization might create more an effective and informed business strategy. The approach is called business analytics. Business analytics (BA) is an approach to make decision based on data. For improving business performance, organization has to interpret and extract knowledge of the data, BA is one of a proper approach (Fattah et al., 2025). Data quality, consistency, and coverage play a critical role in achieving expected results. Although business intelligence frameworks are capable of processing diverse data types, success largely depends on the availability of clean and actionable data. It is also important to recognize that ensuring data quality is a continuous effort that requires ongoing attention (Bany Mohammed et al., 2024).

However, not all the data consisted filled complete. Some data contained missing values. Missing value is known as the value which has not existed on the observation. It might be caused by several factors, such as uncomplete survey, irregular events, bad connection, and scheduled maintenance or any other circumstances (Li et al., 2024). The issue is faced in a common numerous scientific disciplines (Xiao et al., 2025). Missing data has to be handled carefully. Several statistical methods have been used for the case (Xiao et al., 2025). Handling missing data is a basic requirement for classifying patterns in data mining, particularly classification. Improper handling of missing data can result in misclassification. Even without classification, unclean data can also provide incorrect information in the data exploration process for strategic purposes (Desiani et al., 2021). The data will be analysed to see the number of missing values in each of variables contained. One of researcher viewed the number of data and compared total of missing value in values and percentage (Downing, 2025). According to appearance of the missing value, some problems might be occurred such as decreased performance, data analysis problems, and biased results caused by the difference between missing and complete values. To handle it, some statistical techniques are used for example mean, regression, k-nearest neighbor, ensemble based which consider proper to implement (Emmanuel et al., 2021). On the other hand, one of researcher applied conditional deletion of missing value around 50% of total in one attribute. It is assumed that features that have more than 50% missing values cannot provide adequate information or reliable insights for future analysis (Nguyen & Byeon, 2024). Moreover, other research is found that only 20% of the tolerance of missing values can be accepted. After checking the missing data

from row to row, then the attribute will be removed if more than acceptable value. Missing data is imputed using the mean (for numeric features) or mode (for categorical features) if there are a small number of missing values (Irwanto, 2024). In this research, the missing values will be imputed with tolerance of missing not more than 20%, other than that the attribute will be dropped.

Exploratory data analysis (EDA) is one of step that helped data analyst to understand the data. That technic provided data knowledge using some charts, graph or statistic information. Using the output, it night. EDA is offered a summary of the dataset analysis and preliminary findings. One of researcher applied for analysed which public high schools are more likely to have underrepresented students (Paramita & Ramadhan, 2024). EDA also is an approach to data visualization and analysis that has evolved alongside the development of scientific methods for data visualization, emphasizing analytical reasoning through interactive visual interfaces (Orji et al., 2022). One of study used EDA to maximize data's value and guided also developed research program. There are two types of EDA which univariate and multivariate. Univariate helped to find pattern of single variable using box plot, histogram, and kernel density. For multivariate, EDA helped to analyze large dataset to find trends and clusters (Ameer & Fatahi Valilai, 2025). On the other hand, spatial analysis also used to understand the number of students in each area. Spatial analysis is a clustering analysis to identify spatial patterns and dependencies (Zia et al., 2024). It showed the map of the distribution of students' areas of origin.

Results and Discussions

```
> sapply(maba, function(x) sum(is.na(x)))
      camab_no_reg      code_thn_akademik      code_gel_penerimaan      camabNamaLengkap
      0              0              0              0
camabProdiPilihan1      nm_prodi_pil_1      kd_fakultas_pil_1      nm_fakultas_pil_1
      0              0              0              0
      jenjang_pil_1      camabProdiPilihan2      nm_prodi_pil2      kd_fakultas_pil2
      0              4804              4804              4804
      nm_fakultas_pil2      jenjang_pil2      camabProdiPilihan3      nm_prodi_pil_3
      4804              4804              4804              4804
      kd_fakultas_pil_3      nm_fakultas_pil_3      jenjang_pil_3      camabBebasTes
      4804              4804              4804              2269
      camabTempatLahir      camabTanggallahir      camabJenisKelamin      camabAgama
      0              0              0              0
camabKewarganegaraan      camabPekerjaan      pekerjaan_maba      camabAlamatTinggal
      0              16167              16167              0
      camabPropinsi      nm_provinsi      camabKab      nm_kab
      0              0              0              0
      camabTelpHP      camabEmail      camabNamaBapak      camabPendBapak
      0              16167              0              16167
camabPekerjaanBapak      pekerjaan_bapak      camabPenghasilanBapak      camabTelpHP_ayah
      337              1151              725              405
      camabPendIbu      camabPekerjaanIbu      pekerjaan_ibu      camabPenghasilanIbu
      16167              337              6696              725
      camabTelpHP_ibu      camabAlamatOrtu      kd_prodi_lulus      nm_prodi_lulus
      399              337              0              0
      kd_fakultas_lulus      nm_fakultas_lulus      jenjang_prodi_lulus      regmbNoInduk
      0              0              0              0
```

Figure 1. Missing Values in Attributes

According to the figure 1, some attributes consisted missing values. There are variables decided to be dropped because consisted high number of missing values. Referring to previous research, only 20% of missing values can be accepted. However, some attributes did not attach any data such as occupation of student, student father's occupation, student's mother education background. All in all, the variables is dropped from data. On the other hand, some data filled with mean and mode because the number of missing considered low. Mean imputation replaced the missing value using average of respective coloumn (Kazadi Mbamba et al., 2025). From 52 attributes, 11 attributes dropped from the data in pre-processing phase. In this data, the data will be replaced by mean or mode if proportion of missing value is lower than 20%. Some variables which had 100% missing values is dropped such as student's job, student's nationality, email of student, ID job of student's father and student's mother

education. The attributes which applied mean and mode imputation are student's father job, income of student's father, and others.

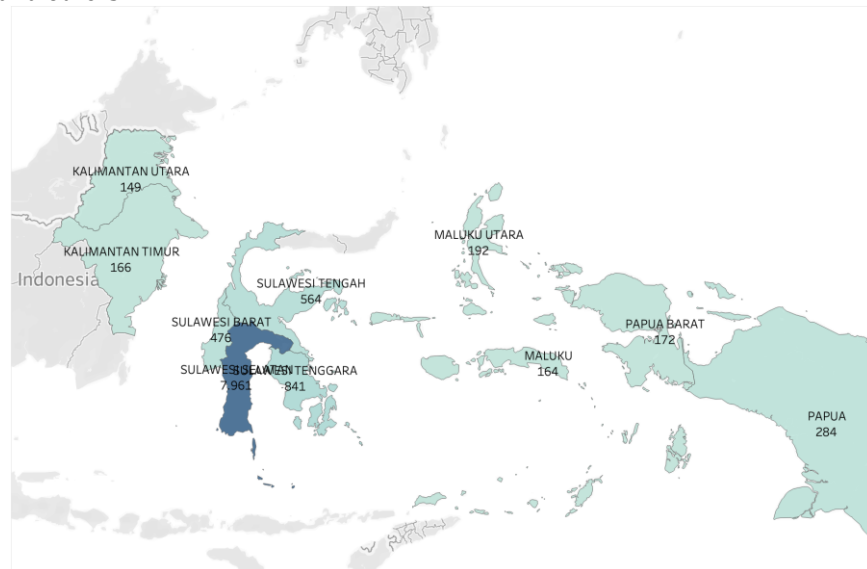


Figure 2. Spatial Analysis

The data above showed the number of new students aorund 10,405 out of a total of 11,925 from 2023-2024. Accoding to the demographic image above, South Sulawesi was the region with the highest number of new students approximately 8 thousands. Where the dark blue color indicated that South Sulawesi is the most dominant province as the origin of new students in two years. According to the BPS data, Sulawesi Selatan is known the highest number of students in east Indonesia (Badan Pusat Statistik, 2025). It will be oppportunity from university as known the university in sulawesi selatan area so it will be easier to conduct the promotion.

Nm Provinsi	F	Code Thn Akademik	
		20231	20241
SULAWESI SELATAN	4.038	3.923	
SULAWESI TENGGARA	410	431	
SULAWESI TENGAH	275	289	
SULAWESI BARAT	246	230	
PAPUA	141	143	
MALUKU UTARA	94	98	
PAPUA BARAT	86	86	
KALIMANTAN TIMUR	80	86	
MALUKU	85	79	
KALIMANTAN UTARA	84	65	
GORONTALO	58	74	
NUSA TENGGARA BARAT	55	69	
NUSA TENGGARA TIMUR	48	52	
SULAWESI UTARA	43	50	
DKI JAKARTA	36	40	
JAWA BARAT	45	29	
JAMBI	59	11	
RIAU	21	13	
KALIMANTAN SELATAN	9	23	
JAWA TIMUR	21	11	
DI YOGYAKARTA	24	8	
JAWA TENGAH	13	14	
SUMATERA BARAT	19	6	
BANTEN	5	14	
SUMATERA SELATAN	13	4	
KEPULAUAN RIAU	9	7	
KEPULAUAN BANGKA BELITUNG	3	11	
ACEH	4	7	
LAMPUNG	3	7	
SUMATERA UTARA	4	1	
KALIMANTAN BARAT	3	2	
BENGGKULU	1	4	
KALIMANTAN TENGAH	2	1	

Figure 3. Number of students from 2023-2024 by Provinces

According to the previous map graph, South Sulawesi was the province with the highest number of new students at UMI in 2023 and 2024. More than 80% of the new students detected from that province. In addition, several other provinces, namely Southeast Sulawesi, Central Sulawesi, and West Sulawesi, were also contributed. Based on this data, Sulawesi is considered the most dominant region for new students from 2023 to 2024, although there was a decrease in the number of students in 2024. On the other hand, there are new students from far away from Sulawesi, namely Aceh, North Sumatra, and East Java. UMI can expand more to areas with a small number of students, which is an opportunity to increase the number of students from the area. In addition, Kalimantan is known area which had many students also joined in UMI. The increasing number of student helped the university to attract the potential student from that area. According to the data statistic, Kalimantan is known as area that had minimum contributing in a number of university in that area (Badan Pusat Statistik, 2025).

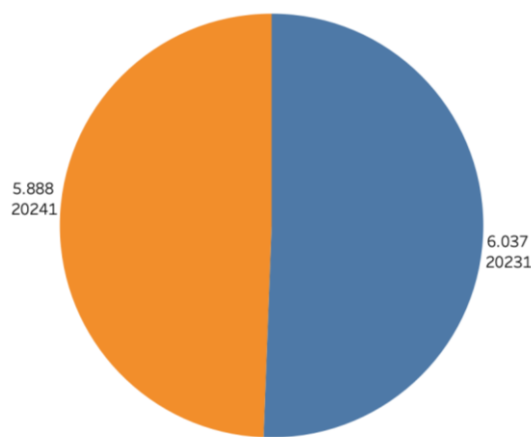


Figure 4. New Student Data from 2023 and 2024 admissions.

Based on the pie chart in Figure 4, the number of students in 2023 is higher than in 2024. Where the number of new students in 2023 was 6,037 and 5,888 for 2024. This is evident from the phenomenal background of the number of new students which experienced a decline in the last year based on data.

Jenjang ..	Code Thn Akademik	
	20231	20241
S1	4.185	3.892
P	1.093	1.043
S2	616	797
S3	123	137
D3	20	19

Figure 5. Number of Student by degree level in 2023 & 2024

Regarding the educational levels chosen at UMI, undergraduate programs have the highest number of applicants, followed by professional programs. The number of new students at the master's and doctoral levels increased from 2023 to 2024, while the number of undergraduate, professional, and

diploma programs decreased. UMI needs to reconsider this to focus on improving the different levels. Further analysis will focus on the undergraduate program.

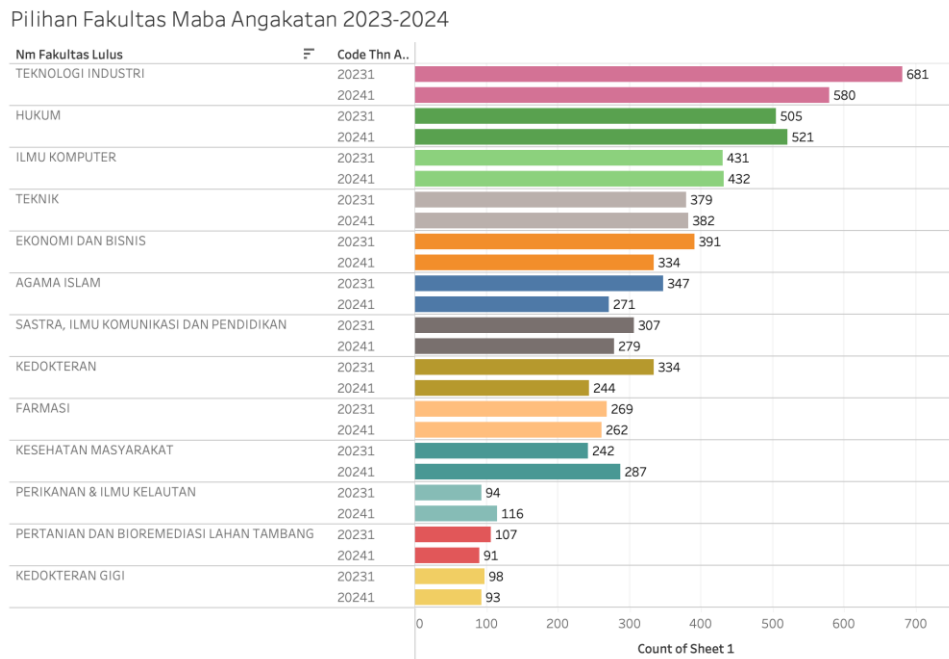
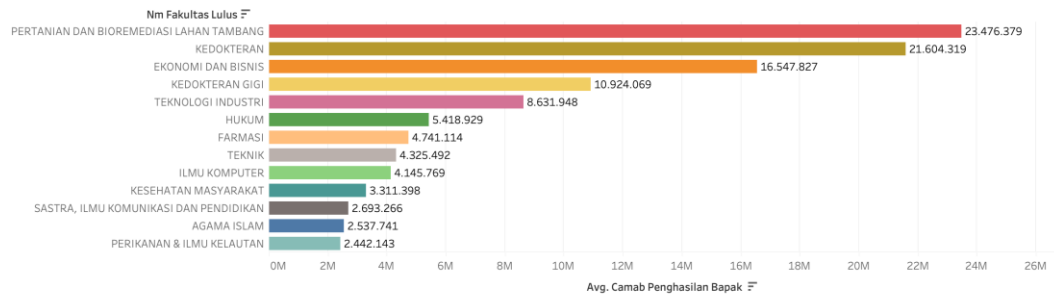


Figure 6. Number of Students by Faculty

Based on the bar graph above, industrial technology is the most popular faculty among undergraduate students in 2023 and 2024. In addition, there are several other top faculties, namely law, computer science, engineering, and economics and business. However, several faculties experienced a decline from 2023 to 2024, including the faculties of industrial technology, economics and business, Islamic religion, literature and communication sciences and education, medicine, pharmacy, agriculture and mining bioremediation, and dentistry. Only five of the 13 faculties experienced an increase in the number of students such as law, computer science, engineering, public health, and fisheries and marine sciences, although there were also faculties that only experienced an insignificant increase. According to Institut Ilmu Kesehatan Bhakti Wiyata, public health is a major that has good job prospect. It is because awareness of the importance of preventive and promotive health is increasing, driven by global issues such as pandemics and climate change, as well as government programs (Institut Ilmu Kesehatan Bhakti Wiyata, 2025).

Rata-rata Penghasilan Bapak per Fakultas



Rata-rata Penghasilan Ibu per Fakultas

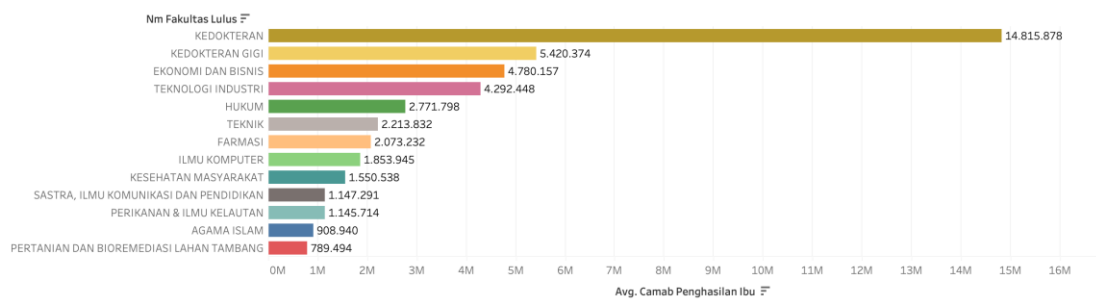


Figure 7. Average of Parent's Income by Faculty

Based on the bar graph above, the income of male parents of new students is in nearly the same order for the 2023 and 2024 academic years. Agriculture, bioremediation, and mining are ranked first, followed by the faculties of medicine, economics and business, dentistry, and others. From this data, high tuition fees (BPP/SPP) do not guarantee a student has a high parental income. The Faculty of Medicine has the highest tuition fees compared to other faculties. On the other hand, the faculties with the lowest average parental income are the Faculties of Islamic Religion and Fisheries & Marine Sciences.

Based on the average mother's income for each faculty, the Faculty of Medicine had the highest income, at IDR 16 million for new students in 2023 and IDR 12 million for 2024. Each year has a different average income for each faculty. In 2023, the average mother's income for new students was around IDR 2 million for the Faculty of Law, Engineering, and Pharmacy.

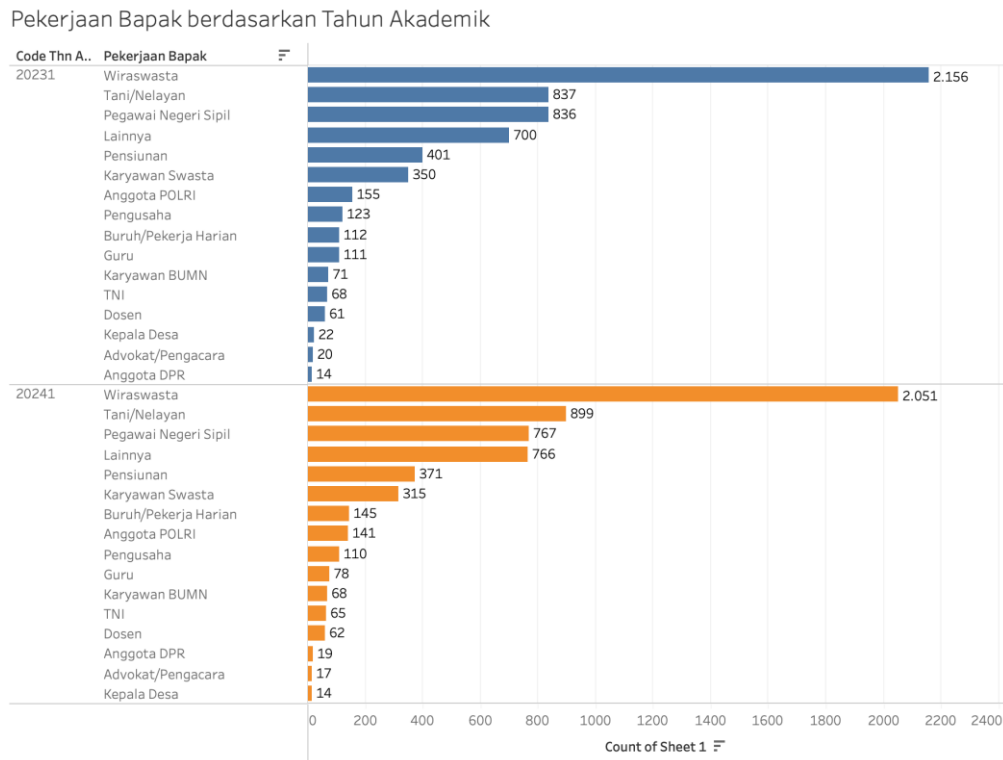


Figure 8. Student Father’s Occupation by Year Admissions

Based on the type of work, self-employment is the most dominant occupation of male parents of new students of Class of 2023 and 2024. Where 30% of the number of new students dominate self-employment as the most common occupation of male parents. Rereferred to previous research, entrepreneur was one of the high number of occupation type from the students (Irwanto, 2024).

Conclusions

The missing values in each of attribute might be cleaned using another way. The admission team in university might encourage the registrants to complete and input correct data to the system. The important attribute which contributed to the analysis have to sign as required field in the form. 41 attributes was left from 52 for further analysis. According to the visualization result, we found that some college students applied to university from several background of area, demographic and etc. The marketing division might apply another strategy which had small number of student on particular areas. For Makassar, it might just maintain the existing strategy and giving some innovation because the antusiasme of student in that area is high from each year’s admission. Kalimantan is known as potential area to promote the university due to the minimum availability of college in that area. Public health is the major that had increasing student supported by job potential in the future so it encourage people to join the faculty. For the parent’s income especially for student’s father, the highest income from all faculties is from *pertanian dan bioremediasi lahan tambang*. According to the tuition fee of the faculty, it is not the highest comparing to others. However, from mother’s salary, most of high income is from medicine faculty. As we know, that faculty has very high tuition fee and it is appropriate with the mom’s income. All in all, comparing both of the parent’s income, the variable can not be concluded that program study can affected or not by the salary of student parent’s earnings. Most of the students in UMI have similar parent’s occupation background which is entrepreneur. University might utilize this knowledge to target some entrepreneurs who have age considered having child in senior high school for next admission. For the next research, gender attribute might be considered to be evaluated in each of faculty

or study program chosen in Universitas Muslim Indonesia to understand which gender can be focused in strategy.

References

- Ameer, T., & Fatahi Valilai, O. (2025). Predictive exploratory data analysis of shopfloor CNC machine operation through a machine learning model. *Journal of Open Innovation: Technology, Market, and Complexity*, 11(2). <https://doi.org/10.1016/j.joitmc.2025.100559>
- Badan Pusat Statistik. (2025, February 19). *Jumlah Perguruan Tinggi1, Dosen, dan Mahasiswa2 (Negeri dan Swasta) di Bawah Kementerian Pendidikan Tinggi, Sains, dan Teknologi Menurut Provinsi, 2024*. <https://www.bps.go.id/id/statistics-table/3/Y21kVGRHNXZVMEI3S3pCRllyMHJRbnB1WkVZemR6MDkjMw==/jumlah-perguruan-tinggi1-dosen-dan-mahasiswa2-negeri-dan-swasta-di-bawah-kementerian-pendidikan-kebudayaan-riset-dan-teknologi-menurut-provinsi-2022.html>
- Bany Mohammed, A., Al-Okaily, M., Qaism, D., & Khalaf Al-Majali, M. (2024). Towards an understanding of business intelligence and analytics usage: Evidence from the banking industry. *International Journal of Information Management Data Insights*, 4(1). <https://doi.org/10.1016/j.jjime.2024.100215>
- Dang, H. C. T., Thi Nguyen, P. L., Thi Le, P. L., Hoang Nguyen, T. T., & Vu, B. T. (2025). Impact of big data analytics capabilities on sustainable performance of Vietnamese retail companies: The mediating role of innovation. *Journal of Open Innovation: Technology, Market, and Complexity*, 11(3). <https://doi.org/10.1016/j.joitmc.2025.100569>
- Desiani, A., Dewi, N. R., Fauza, A. N., Rachmatullah, N., Arhami, M., & Nawawi, M. (2021). Handling Missing Data Using Combination of Deletion Technique, Mean, Mode and Artificial Neural Network Imputation for Heart Disease Dataset. In *Science and Technology Indonesia* (Vol. 6, Issue 4). <https://doi.org/11.26554/sti.2221.6.4.333-312>
- Downing, N. J. (2025). Missing value imputation in environmental, social, and governance data: an impact on emissions scores. *Finance Research Letters*, 85. <https://doi.org/10.1016/j.frl.2025.107818>
- Emmanuel, T., Maupong, T., Mpoeleng, D., Semong, T., Mphago, B., & Tabona, O. (2021). A Survey on Missing Data in Machine Learning. *Journal of Big Data*, 8(1). <https://doi.org/10.1186/s40537-021-00516-9>
- Fattah, I. A., Prabowo, H., Tjhin, V. U., & Rahim, R. K. (2025). The interplay between business analytics capabilities and decision-making performance in Indonesia's public sector. *Digital Business*, 5(2). <https://doi.org/10.1016/j.digbus.2025.100132>
- Griha Tofik Isa, I., Novianti, L., Elfaladonna, F., Agustri, S., Manajemen Informatika, J., & Negeri Sriwijaya, P. (2023). *Exploratory Data Analysis (EDA) dalam Dataset Penerimaan Mahasiswa Baru Universitas XYZ Palembang*.
- Institut Ilmu Kesehatan Bhakti Wiyata. (2025, December). *Kuliah Kesehatan Masyarakat: Prospek Kerja Lulusannya*. <https://iik.ac.id/blog/2025/11/26/kuliah-kesehatan-masyarakat-prospek-kerja-lulusannya/#:~:text=Mengapa%20Lulusan%20Kesehatan%20Masyarakat%20Banyak,dan%20preventif%2C%20bukan%20hanya%20kuratif.>
- Irwanto, I. (2024). Data on Undergraduate Students' Self-Regulation in Online and Blended Learning Environments During the COVID-19 Pandemic in Indonesia. *Data in Brief*, 53. <https://doi.org/10.1016/j.dib.2024.110066>
- Jabir, S. R., Azis, H., & Mansyur, S. H. (2024). *Enhancing The Quality of College Decisions Through Decision Tree and Random Forest Models*. https://journal.unm.ac.id/index.php/JESSI/article/view/1225?TSPD_101_R0=085ba2cd96ab200005eb449f3cf77e728bdcd2c71d77a5a50be6b62f2ff1105f5d54a0ca4156dc1208e60be126144

- 800c18eee08eb5bfa9f6c28aca1f0dcbf5c0f4645c0419c6ff6a52be03da4c0a7b68e80571a5ee868b0d988d1c8930b31d521421578c4a7be933174da7c4556f0c69205044588123639
- Jabir, S. R., Tenripada, A. U., Asis, M. A., Widyawati, D., & Faradibah, A. (2022). *Buletin Sistem Informasi dan Teknologi Islam Pengembangan Solusi Perawatan Kesehatan Terhadap Autism Spectrum Disorder (ASD) Menggunakan Pendekatan Data Analysis* INFORMASI ARTIKEL ABSTRAK. 3(2), 157–166. <https://jurnal.fikom.umi.ac.id/index.php/BUSITI/article/view/1397>
- Katyal, A., Sharma, P. K., & Kannan, M. (2025). *Exploratory Data Analysis (EDA) on Undergraduate Data Science Students Through R Programming*. <https://doi.org/10.21203/rs.3.rs-7422204/v1>
- Kazadi Mbamba, C., Keymer, P., Alvi, M., Topalian, S. O. N., Ud Din, F., & Batstone, D. J. (2025). Enhancing data quality in wastewater processes: Missing data imputation with deep Variational Autoencoders and genetic algorithms. *Computers and Chemical Engineering*, 199. <https://doi.org/10.1016/j.compchemeng.2025.109123>
- Li, G., Zheng, Q., Liu, Y., Li, X., Qin, W., & Diao, X. (2024). A Classification Method for Incomplete Mixed Data Using Imputation and Feature Selection. *Applied Sciences (Switzerland)*, 14(14). <https://doi.org/10.3390/app14145993>
- MZ, Y., Bororing, J. E., Rahayu, S., & Ramadhani, T. A. (2022). Aplikasi Dashboard Visualisasi Data Calon Mahasiswa Baru menggunakan Metabase. *Edumatic: Jurnal Pendidikan Informatika*, 6(1), 116–125. <https://doi.org/10.29408/edumatic.v6i1.5483>
- Nguyen, H. V., & Byeon, H. (2024). A Hybrid Self-Supervised Model Predicting Life Satisfaction in South Korea. *Frontiers in Public Health*, 12. <https://doi.org/10.3389/fpubh.2024.1445864>
- Orji, U. E., Ukwandu, E., Obianuju, E. . A., Ezema, M. . E., Ugwuishiwu, C. . H., & Egbugha, M. . C. (2022). *2022 5th Information Technology for Education and Development*. <https://arxiv.org/pdf/2305.19297>
- Paramita, A. S., & Ramadhan, A. (2024). An Unsupervised Learning and EDA Approach for Specialized High School Admissions. *Journal of Applied Data Sciences*, 5(2), 316–325. <https://doi.org/10.47738/jads.v5i2.178>
- Popoola, P. A., Tapamo, J. R., & Assounga, A. G. H. (2025). Effective and efficient handling of missing data in supervised machine learning. *Data Science and Management*, 8(3), 361–373. <https://doi.org/10.1016/j.dsm.2024.12.002>
- Septiani, N., & Iqbal, M. (2024). Analisis Tren Pendaftaran Siswa Menggunakan Big Data. *Bulletin of Information Technology (BIT)*, 5(4), 366–370. <https://doi.org/10.47065/bit.v5i2.1744>
- Srivastava, N. K., Pandey, P., & Mishra, V. (2023). Exploratory Data Analysis (EDA) Based on Demographical Features for Students' Performance Prediction. In A. Dagur, K. Singh, P. S. Mehra, & D. K. Shukla (Eds.), *Artificial Intelligence, Blockchain, Computing and Security Volume 1* (1st edition, p. 1028). CRC PRes.
- Tri, I., Yanto, R., & Handayani, O. P. (n.d.). *Visualization of Data Inventory Using Visual Data Mining (VDM) and Exploratory Data Analysis (EDA) Methods*. Retrieved December 25, 2025, from <https://joiv.org/index.php/joiv/article/view/4075/1429>
- UMI. (n.d.). *Profil & Sejarah UMI*. 2025. Retrieved June 24, 2025, from <https://umi.ac.id/struktur-umi/profil-sejarah-umi/>
- Universitas Ahmad Dahlan. (2023, November 7). *Tantangan PTS Hadapi Penurunan Mahasiswa Baru dan Upaya Mengatasinya*. <https://lldikti5.kemdikbud.go.id/home/detailpost/tantangan-pts-hadapi-penurunan-mahasiswa-baru-dan-upaya-mengatasinya>
- Xiao, H., Cao, R., Chen, Z., Hong, C., Wang, J., Yao, M., Fan, L., & Luo, T. (2025). Handling missing data in large-scale TBM datasets: Methods, strategies, and applications. *Intelligent Geoengineering*, 2(3), 109–125. <https://doi.org/10.1016/j.ige.2025.07.001>