Muhammad Zeeshan Raza

Data Infused Strategies for Student Recruitment - A Focus on Data-Backed Decision Making

School of Technology and Innovations
Master's thesis in Industrial Systems Analytics

Vaasa 2024

University of Vaasa
School of Technology and Innovations

Author:	Muhammad Zeeshan Raza
Title of the thesis:	Data Infused Strategies for Student Recruitment - A Focus on Data-Backed Decision Making
Degree:	Master of Science
Discipline:	Industrial Systems Analytics
Supervisor:	Amit K. Shukla
Year:	2024 **Pages:** 67

Abstract

Extracting and getting meaningful information from the data is crucial in today's world when the generation and availability of data has become faster than ever before. You probably have heard that while data may be independent from information, information can never be independent of data. Therefore, data analytics has become essential means of transforming raw data into actionable information.

The need for effective decision-making process and particular marketing strategies in the process of foreign students' recruitment prompts the University student recruiting team to create an easily accessible and comprehensive database from where all necessary information relating to the issue under consideration can be retrieved and analyzed. Thus, the present research supports the adoption of dynamic dashboards as a way of developing the university's branding and marketing strategies when employing a data-focused methodology. Metrics for the created visualizations will be identified in the form of key performance indicators (KPIs) and will be based on the data extracted from the students' application files.

It includes Data Extraction, Data Cleansing, Data Anonymization, Data Transformation, Data Integration, Data Loading, Data Visualization and Data Forecasting as part of study approach. They will help in evaluating the previous data and in using the future information without further alteration of the whole structure. Furthermore, the research will help the study of trends and estimates through the use of artificial intelligence or machine learning models, which will help in developing the necessary answers on future trends and outcomes.

The general objective of this thesis is to enhance the university's strategies for acquiring more students through advancing data analytics to provide for more rational approaches to decision making.

Contents

Table

Figures

Chapter 1. Introduction

While recruiting and enrolling students, universities face several challenges in the current dynamic educational environment. As competition grows more intense and the target population becomes more diverse, more sophisticated and information-driven approaches are needed to attract the right students and retain them. Conventional hiring practices often fail to recognize and target prospective students, comprehend the diversity of applicant profiles, and maximize marketing initiatives (Gilch & Sieweke, 2021). Furthermore, organizations have to utilize such resources as they stay relevant in the market due to advancement in technology and availability of information which is now becoming common. Another issue is the poor quality of data structures thereby making institutions struggle to gather meaningful information from big data, reducing their capacity to make strategic decisions (Smith, 2020).

Future learners are more critical and expect more personalization; this means they expect to be provided with data and other input that will be relevant to their decisions (Johnson & Coleman, 2019). Much attention should be paid to developing strategies that can help institutions overcome the challenging environment characteristic for the industry. It is thus agreed that better effectiveness in student recruitment will be realized through the application of big data & analytics by organizations, where these facets will enhance pattern identification, student behavior prediction, and resource utilization (Brown et al. , 2023).

Another factor increasing the level of concern is that education is becoming more and more internationalized, and the number of foreign students, who move from country to country, is on the rise. Universities effectively have to advertise themselves on the global market, which naturally requires a great deal of strategic thinking and profound understanding of how one can win the hearts and minds of people from different cultures (Jones & Kilburn, 2022). As a result of this global competition, sophisticated analytical tools are necessary to gather as well as process data from various geographical areas and populations. Furthermore, the current global crisis with COVID-19 underlines the necessity of building up the apt digital structures and employing the data-driven strategies. The pandemic has forced academic institutions to shift from traditional face-to-face recruitment fairs and workshops to exclusively online events, as well as online courses and lectures (Smith & Clark, 2021). The institutions that had developed strategies that incorporated data analytic and digital platforms were at an advantage and continued with the recruitment as they adapted to the changes.

This study aims at solving these challenges by providing a strategy for data-centred strategies to student recruitment. Through the application of data analytics in decision-making processes, universities may enhance their efficiency and variety of aims and goals, which in time may result in achieving the goals. This type of research has great importance since this new approach has the potential to revolutionize the process of students recruiting (Iivari, N. , Sharma, S. , & Ventä-Olkkonen, L. , 2020). The use of quantitative approaches may help institutions in achieving better competitiveness, decide the most effective distribution of the resources they have, and increase students' diversity and equality. Consequently, the process of recruiting students was mainly based on the use of quantitative indications and qualitative estimations; oftentimes ignoring such factors that may affect candidate action and choice (Posselt et al. , 2020). The technological advancements in data analysis, and the application of ML algorithms coupled with cloud computing have created opportunities to delve deeper into the students' profile and their interests and academic profile (Kuleto, V. , Ilić, M. , Dumangiu, M. , Ranković, M. , Martins, O. M. , Păun, D. , & Mihoreanu, L. , 2021). Moreover, universities themselves might rely on previous years' results or employ forecasting strategies to identify potential challenges and ensure timely adjustments of their recruitment strategies in accordance with the demand on the market.

This thesis focuses on several important aspects, which are as follows:

- integration of analytical techniques in the practices that concern admission and enrolment of students.
- a list of Key Performance Indicators (KPIs) that are vital to measure recruitment effectiveness.
- the importance of anonymizing data in order to prevent important information leaks.
- specifically, the use of predictive modelling techniques to forecast future outcome.
- the real-life applications of data-driven approaches in actual university environments (Taylor Jr, L. D. , 2020).

To achieve this, this research aims at identifying some of the important studies of application timing including; demographic information of the applicants and students, their educational background and their results as stated in the admission cycles of 2022 and 2023. Notably, the thesis will also seek to identify data-driven solutions for student recruitment that is made up of several activities including, data extraction, cleansing and transformation, integration and loading, visualization and data prediction.

In addition, the suggested approach complies with the general trends observed in the field of higher education, in which analytics and predictive modelling are recognized as indispensable tools for institutional advancement (Zawacki-Richter et al. , 2019). Besides, this research aims to provide valuable implications, strategies, and recommendations to implement changes in student recruitment processes in higher learning institutions (Beech, S. E. , 2020). It emphasizes the need for institutions to embrace utilization of data in decision-making within more intensified competition within the education sector.

1.1 Research Question and Objectives

To address the above goals, we have formulated and identified a research question and several research objectives, which are mentioned below:

Research Question

RQ: How can universities enhance the student recruitment and marketing strategy through data-driven decision making?

Objectives

O1: To facilitate a data driven approach to student recruitment and admission.
O2: To discover and understand the key performance indicators (KPIs) that will offer opportunities for improvement in recruitment approaches.
O3: To establish a more efficient way of gathering and analyzing recruitment information.
O4: To determine whether it is advantageous and possible to migrate and consolidate historical data.
O5: To apply the concept of machine learning in predicting and forecasting the results of recruitment of students.

Chapter 2: Literature Review

When trying to improve student recruiting efforts using data-driven methods, there is a need to look at previous literature in an attempt to understand theoretical principles as well as the findings relevant to the research aim. The purpose of this literature review is to analyze previous research on using data to make decisions about student recruitment, as those works are related to decision-making processes regarding student recruitment based on data. It will draw on important theoretical views to get valuable insights.

2.1 Theoretical Frameworks

2.1.1 Data-Driven Decision-Making in Higher Education

Data-driven decision-making (DDDM) has become a popular topic over the last several years, particularly within the context of higher education institutions (Kaspi, S. , & Venkatraman, S. , 2023). Anderson and Dexter (2005) have made a great contribution to encourage the adoption of this shift in thinking because it has various implications for strategic planning and operational decision making within the educational sector (Kallio, T. J). In their study, they reveal a number of important issues concerning the use of data from many sources to define the effective strategies for student's behaviour and achievement (Alshanqiti, A. , & Namoun, A. , 2020).

Universities, understanding the new dynamics and competitive scenario, are slowly coming to terms with the significant role of 'big data' in tuning their recruitment process and students' success initiatives. Institutions may make strategic decisions by using the integration of data as well as motivating analysing tools to determine complicated patterns and trends (Basu, R. , Lim, W. M. , Kumar, A. , & Kumar, S. , 2023). Such data analysis could shed light on which recruitment strategies are most closely associated with attracting potential students for enrollment, thus, enhancing enrollment ratios and increasing the students' diversity.

Moreover, the implementation of DDDM does not end in the aspect of recruitment but has influenced other areas of higher education operations. University managers have the ability through data to make informed decisions that support organizational excellence, learning outcomes and can extend to assessment of academic programs, resource allocation among others (Dukic, J. , 2021). In this manner, through the use of data, institutions are able to identify potential areas for improvement. They are also able to assess the success of existing management and support strategies, which in turn enables the enhancing of processes so as to better address and respond to the needs of the learners and stakeholders (Stronge, J. H. , & Xu, X. , 2021).

In the existing literature about DDDM in higher education, a significant concern has been made regarding the importance of developing the organisational culture to support data-driven decision making. The authors argued that through integrating the data analytics into strategic planning, schools may find new opportunities for innovation and growth in the development of the new brand ownership and positioning within the ever-evolving educational environment (Mian et al. , 2020).

2.1.2 Predictive Modeling and Machine Learning

Assessment technologies and strategies such as predictive modeling and machine learning are important strategies that have found their way into helping universities enhance their efficiency in student enrollment and retention within the existing and developing learning environment. Arnold and Pistilli (2012) have done a great job in pointing out the huge benefits of another form of analytics known as predictive analytics in this area and reveal the extent of its usefulness in several critical sectors (Nguyen, A. , Gardner, L. , & Sheridan, D. , 2020).

Predictive modeling is often applied to find learners who are at risk of dropping out, so measures can be taken for them to prevent them from dropping (Bustamante et al. , 2021). It is also important to note that with past data concerning academic performance, participation rate and socio-demographic parameters, the predictive model has the possibility of pointing out students who may be facing learning challenges or even lacking interest (Matz, S. C. , Bukow, C. S. , Peters, H. , Deacons, C. , Dinu, A. , & Stachl, C. , 2023). Based on this understanding, it is possible for universities to plan and organize specific support service and learning resources in an effort to minimize certain risks or threats and to maximize positive impacts on learner outcomes.

Additionally, predictive modeling can also be in favor of determining other enrollment trends, which will be useful for many colleges to get the info about the further increase and demographic shifts. Institutional predictive models could generate potential enrollment using the findings evaluated against enrollment figures with overlapping external features, including economic trends and shifts in demographics (Kerr, C. C. , Stuart, R. M. , Mistry, D. , Abeysuriya, R. G. , Rosenfeld, K. , Hart, G. R. , . . . & Klein, D. J. , 2021). Which means that institutions can adapt to this in their own way to change their mode of recruiting. Thus, more active strategy lines can help colleges to respond to the changes of market conditions and sustain competitive advantage on the given market in recruitment and training of highly skilled individuals.

Furthermore, predictive analytics enable the personalization of communication and interaction strategies for students, as it is required by the characteristics and needs of each learner. In turn, the universities are able to employ complex computations to analyze large data samples in order to reveal certain tendencies that may indicate the students' preferences and behavior patterns (Nie, M. , Xiong, Z. , & Yang, G. , 2020). A detailed understanding of the subject allows the tailoring of messages, resources, and support services that will strengthen bonds and enhance the overall student experience.

The reviews of literature on predictive modelling and machine learning underscore the implicit necessity of these concepts of revolutionizing student recruitment and retention procedures. This might help universities to not just develop new ways to influence students for better results but can also predict future results and take necessary action (Dekker, I. , De Jong, E. M. , Schippers, M. C. , & Giesbers, B. , 2020). The possibilities brought by the sheer progress of the technology that we have cannot be approached in terms of goals since there is nothing that is unachievable when it comes to predictive modeling in higher learning institutions.

2.1.3 Marketing and Consumer Behavior Theories

Marketing and consumer behaviour theories have been very relevant in the formulation of strategies and techniques that can be used effectively when it comes to marketing for students in terms of student recruitment. Segmentation targeting and position (STP) technique is highlighted by Kotler and Armstrong (2016) as some of the essential approaches which can be used to address different students and to establish a relationship within the enrollment process (M. ANTARA & N. P. A. Y. DEWI, 2022).

Previous works have explained that through the 'segmentation concept', colleges and universities are able to sort potential students based on their similarities in demographic, interest and other factors It means that understanding the needs and interest of the segment, institutions can adjust the approach and goals of their recruitment activities more effectively (Rosenbloom, D. H. , Kravchuk, R. S. , & Clerkin, R. M. , 2022. As a result, they are tailored in a way that focuses on specific kinds of people; this leads to higher rates of engagement and conversion.

Moreover, targeting tactics will allow colleges to better distribute their funds and adjust their communications with the target population to areas which are most likely to join in. It is significant to note that universities, through identifying specific market groups and analysing their unique characteristics, may determine the essential segments that will aid in channeling resources to effectively influence potential consumers (Vrontis, D. , Makrides, A. , Christofi, M. , & Thrassou, A. , 2021).

It is a marketing concept also called 'positioning' where a particular university deliberately creates its image and builds its reputation with an aim to differentiate the product to competitors and connect with certain target groups. Universities' recruitment and positioning strategies successfully convey Unique Selling Points or university propositions to orient and affect consumers' decision making process to enhance the gains in the recruiting activities (Ebo Hinson, R. , & Mogaji, E. , 2020).

For successful enrollment marketing communication efforts and good student-stakeholder relationships during the enrollment process, it is important to understand the psychographic variables and needs of the target audience in their decision-making process of joining a learning institution. This article, Alwi, S. , Che-Ha, N. , Nguyen, B. , Ghazali, E. M. , Mutum, D. M. , & Kitchen, P. J. , 2020) identified consumer behavior research that

University can use to amplify its message and experiences of its potential students. It also contributes to an increased number of enrollments by setting the emotional connections for the learning curriculum to be in tandem with the goals of students.

In terms of student recruitment, marketing and consumer behavior theories may help institutions understand the needs of the potential students and consequently explain them that why they should join the institutions (Ercantan, O. , & Eyupoglu, S. , 2022). It has been observed that segmentation, targeting, and positioning strategies can be used to ensure that universities achieve enrollment success, and by using psychographic profiles, universities can create a recruiting process that will resonate with different audiences (Djuric, V. , 2023).

2.2 Relevance to Research Problem

These identified theoretical frameworks are useful in addressing the research topic because they offer a comprehensive understanding of data analysis techniques involving students' recruitment. These frameworks are coherent with the elements from data science, predictive modeling and marketing theories, and in general can facilitate a solid foundation for developing a full-proof strategy for universities that enhances efficiency of recruiting strategies.

However, data usage in various decision-making contexts in higher education as endorsed by Anderson and Dexter (2005) recognizes the importance of using data in support of decision-making planning and executing. This paradigm focuses on integrating data from multiple sources in order to better understand the activities and outcomes of students and to use analytical techniques as a way of capturing the complex nature of the information (Hatch, J. A. , 2023). Recognizing that students' expectations do differ across the two-year and four-year segments throughout the range of institutional goals, colleges may be able to achieve institutional goals by implementing this strategy at the college level for overall recruitment and to improve the amount and effectiveness of student success programs.

Besides, the application of predictive modeling and machine learning enables universities to equip themselves with powerful tools for likely outcomes and adjust subsequent actions in general (Namoun et al. , 2020). Administrators may look at enrollment history and even employ statistics to estimate student behaviors, working to pinpoint struggling students and properly address their needs. This way, colleges using this predictive approach can go ahead and solve challenges, allocate resources, and increase student performance (Al-Mamary et al. , 2020).

In addition, by using the values derived from the theories of marketing and consumer behavior like those of Kotler and Armstrong (2016), the college can be able to determine the reasons and processes adopted by potential students. This approach ushers in importance of segmentation targeting along with position in ensuring that the various groups of students are effectively captured within the marketplace. Psychographic analysis, needs and preferences for potential candidates can help universities to enhance the results of their recruitment activities and to create a strong position in recruiting and partnership with students (Goodrich, K. , Swani, K. , & Munch, J. , 2020).

These theoretical frameworks are stressing the importance of the implementation of technology and functionality of tools which helps to find out useful information from big data (Babu, M. M. , Rahman, M. , Alam, A. , & Dey, B. L. , 2024). A framework in the utilisation of ideas from data science, predictive modeling and theories from the realm of marketing can assist universities to maximise the enrolment and retention of an increasingly diverse student body and also help provide support to them. This understanding forms the basis of enhancing its effectiveness and effectiveness of recruiting campaigns in the future and supports university goals of attaining enrollment numbers and promoting institutional objectives (Balzer, W. K. , 2020).

2.3 Empirical Evidence

Support for the data-driven conclusion can be backed by empirical data about student employment. This paper provides examples of how predictive modeling may develop some useful applications for institutions (Hindle, G. , Kunc, M. , Mortensen, M. , Oztekin, & Vidgen, 2020). This study used predictive modeling to quickly pinpoint student candidates who were likely to drop out; institutions could then intervene early with messages of support and sources of help. Accomplishing evidenced-based predictors and historical data, universities effectively recognized potential signs of the disconnection or a more permanent disruption to the studies (Oqaidi, K. , Aouhassi, S. , & Mansouri, K. , 2022). This made them able to tackle any cases of student leakage or poor academic performance in order to reduce student dropout levels.

Case studies at universities with full data analytics systems show how data-driven decision-making affects student recruitment (Luca, M., & Bazerman, M. H., 2021). The University of Phoenix and Georgia State University have seen enrollment and student achievement improvements after deploying data analytics programs. These case studies show how universities can use data to learn about student behaviors, interests, and needs. This strengthens recruitment and student outcomes (Stronge, J. H. , & Xu, X. , 2021).

According to Stronge and Xu (2021), colleges apply data to enhance the process of recruitment and students' performance. A major approach is the implementation of enrollment data analysis procedures, demographic and academic profiles included. This work enables institutions to focus the recruitment efforts by pointing out which students are enrolling and performing. Another aspect is the level of students' activity should be also monitored. Colleges track the library, co-curricular and academic support service utilization among others. Knowing these patterns helps institutions design tools that are aligned with the needs and interests of students and enhances the achievement of better students' outcomes.

The literature review also points to the evidence that, for the improvement of student yield outcomes in the context of student recruitment, institutions and students can benefit from the use of predictive modeling and data analytics. The utilization of data analytic enhances student enrollment, satisfaction, and engagement (Kurni, M. , Mohammed, M. S. , & Srinivasa, K. G. , 2023, pp. 1709-1719). Based on this study, it is apparent that data analysis is key for institutions that aim at recruiting and promoting their missions in the current YEAR higher learning institution system. Table 1 provides a comparative analysis of the related work in the literature and our methodology.

Table 1. Comparative Table of Theoretical Frameworks for Data-Driven Student Recruitment

References	Theoretical Framework	Key Concepts & Methodology	Application in Student Recruitment	Unique Contribution
Anderson & Dexter (2005), Kallio et al. (2020), Kaspi & Venkatraman (2023)	Data-Driven Decision-Making	Strategic planning, data integration, operational decision-making	Provides insights on student behaviors, preferences, and performance indicators to enhance recruitment strategies	Emphasizes the importance of a data-centric culture in higher education
Arnold & Pistilli (2012), Nguyen et al. (2020), Matz et al. (2023)	Predictive Modeling and ML	Predictive analytics, risk identification, tailored interventions	Forecasts future outcomes, identifies at-risk students, and customizes outreach efforts	Highlights the use of advanced algorithms to predict student needs and improve engagement
Kotler & Armstrong (2016), Vrontis et al. (2021), Ercantan & Eyupoglu (2022)	Marketing and Consumer Behavior Theories	Segmentation, targeting, positioning (STP), psychographic profiling	Uses market segmentation and targeting to optimize recruitment efforts, positions university brand effectively	Focuses on understanding student motivations and decision-making processes to enhance recruitment
This thesis approach	Data-Driven Decision-Making, Comprehensive Data Analytics	Data Extraction, Data Cleansing, Data Anonymization, Data Transformation, Data Integration, Data Loading, Data Visualization, Data Forecasting	Integrates all stages of data handling into a unified framework, utilizing dynamic dashboards and machine learning techniques	Streamlines the data handling process and ensures continuous feedback for ongoing optimization and data-driven decision-making, enhancing recruitment strategies significantly

From the table we see that in this thesis we provide a more holistic approach to student recruitment process. This work is the first to accommodate the three specified theories in a consummate manner for student recruitment. This is so because it incorporates data analysis, predictions, and marketing plans all in one strategy. Our approach emphasizes that the data-driven decision-making examines how crucial it is to apply data from various sources when developing the operational and strategic plans. This establishes the importance of data, which strengthens recruitment approaches and, overall institutional effectiveness in an institution. Other method that is used are predictive modelling and machine learning which can help identify students who will register for the programme. The prediction will allow universities to plan for some issues so that they can enhance recruitment procedures. This theory can assist institutions in personalising and making stronger recruitment communications though informing them of methods of communicating with the various kinds of students in existence.

Chapter 3: Methodology

This section explains the methodology employed for the anonymization, preprocessing, extraction, transformation, and loading of data pertaining to international student recruitment.

Firstly, the data, in some Excel sheets, was divided into different CSV files and preprocessed through Excel formulas and Power Query to meet its most suitable format. Following set up the data format to a structure, it is then loaded to Azure Data Factory to create one common source of information. Later, the dataset is exported and brought into Azure SQL Database in a tabular form and expanded to MS SQL Server to enable easy pulling of data using SQL queries.

The data was further integrated in Power BI, which helped in deriving various performance indicators to gain a better understanding of the recruitment trend. Also, using Python scripting, the data was imported into python pandas data frames for advanced data analysis and modeling using machine learning algorithms. This kind of machine learning is useful for forecasting future trends since it offers preventive discoveries without the need for further transformation or modeling.

This thesis not only provides grounds for the analysis of past occurrences but also allows prospective examination of trends, thus providing a thorough insight into the processes of international student recruitment.

3. 1 Data Source and Collection

The information used in this project was collected from Studyinfo (n. d.). It refers to the official site that contains information on the study programs. It encompasses two types of admissions occurring annually at the university: rolling admissions as well as joint admissions as depicted in the figure 1 below. The process of gathering data from studyinfo is done often, and the obtained information is stored in Excel spreadsheets for further use. This figure further demonstrates the complete methodological process in this study.

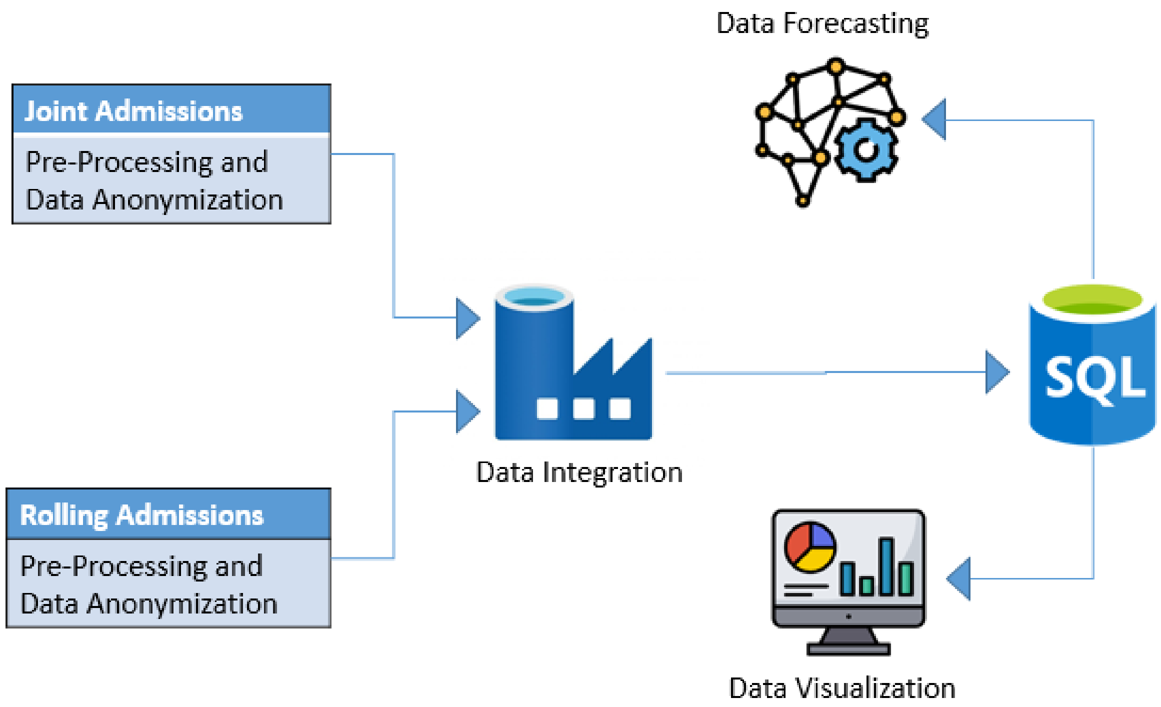


Figure 1. Methodology Flowchart

For the years 2022 and 2023, the university's student recruitment and marketing team provided multiple excel files representing the following data sources:

- Joint Admissions 2022
- Rolling Admissions 2022
- Joint Admissions 2023
- Rolling Admissions 2023

Each data source comprises essential fields, including:

1. **Application submission time:** Enables examination of application submission pattern in relation to various markets, information categorized by year, month, and day.
2. **Student/learner ID*:*** This is a unique ID for creating consistency across different datasets.
3. **Applied programs:** Another way of segregating data so as to review the trends in attraction of the particular subject.
4. **Citizenship:** Gives an idea of the nature of diversity among applicants.

5. **Date of birth:** Assists in the calculation of age for applicants when profiling and aligns marketing strategies as well.
6. **Gender:** Provides a glimpse into the diversity of the applicant pool for marketing to individuals within the organization.
7. **Country of residence:** Shows the location of the applicant at the time of application, which in return helps in designing the marketing and recruitment process.
8. **City of residence:** It just offers more refinements in analyzing the applicant data to better market to specific personas.
9. **Previous education (Graduation year, degree name, degree level, university, country***)*: Helps in understanding the quality of applicants, trends of particular universities, and subject areas of previous degree to determine appropriateness of their programs.
10. **Educational Agency:** Helps in assessing the performance of agencies in previous recruitment campaigns for market segmentation analysis.
11. **Offered Place:** This defines the candidates who are offered a placement so that one can study the trend of conversion for different parameters, including citizenship and program.
12. **Confirmed place:** Helps in defining conversion rates by demographic and program indicators and whether the client accepted the offer.
13. **Registered:** Shows the applicants' enrollment status as either new or returning students, enabling a breakdown of conversion figures from application to enrollment.

This formatted and systematic data is used to research and study the patterns of international student enrollment and thus, make necessary changes to improve the recruitment process and provide suitable programmes. Now, let us consider the actual data of Joint 2022 admission in various excel files. Table 2 lists the corresponding snippet of the data for number of applications. There are around 1242 rows and 10 columns.

Table 2. Joint 2022 All Applications

Application submission time	Anonymised ID	Programme/s	Citizenship	Date of birth	Gender	Citizenship	Name of the degree awarded to you in original language and in English	Name and postal address of the institution	Country where the institution is located
2022-01-18 09:55:35	ANON-263780	Master's Programme in Smart Energy (2 years), 2022 - University of Vaasa, School of Technology and Innovations (1.2.246.562.20.00000000000000002592)	#0: Bangladesh,	01.08.1997	male	Bangladesh	Bachelor of Science in Electrical and Electronic Engineering	American International University- Bangladesh, 408/1, Kuratoli, khilkhet, Dhaka 1229	Bangladesh
2022-01-06 13:04:02	ANON-039676	Master's Programme in Smart Energy (2 years), 2022 - University of Vaasa, School of Technology and Innovations (1.2.246.562.20.00000000000000002592), Master's Programme in Industrial Management (2 years), 2022 - University of Vaasa, School of Technology and Innovations (1.2.246.562.20.00000000000000002590)	#0: Nepal,	06.03.1993	male	Finland	Bachelor in Engineering	Vaasa University of Applied Sciences Wolffintie 30, 65200 Vaasa, Finland	Finland
-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-

Table 3 represents the snippet of application data received by the university in 2022 from all prospective students, providing an overview of the initial interest in the university's programs. Table 3 shows the snippet of admission offers extended to applicants in 2022, indicating how many applicants met the criteria and were offered a place at the university. Table 4 depicts the snippet of admission offers that were accepted by applicants in 2022, highlighting the conversion rate from offers made to offers accepted. Table 5 illustrates the snippet of students who registered and enrolled at the university in 2022, reflecting the final step of the admission process and the actual intake of new students. Such data also exist in other student recruitment data files of Rolling 2022, Joint 2023 and Rolling 2023.

Table 3. Joint 2022 All Offers

Anonymised ID	Programme	Registered
ANON-310055	Master's Programme in Industrial Management (2 years), 2022	Yes
ANON-795006	Master's Programme in Industrial Systems Analytics (2 years), 2022	Yes
ANON-319904	Master's Programme in Industrial Management (2 years), 2022	Yes

Table 4. Joint 2022 All Accepted

Anonymised ID	Programme	Accepted the offer
ANON-121696	Master's Programme in Industrial Management (2 years), 2022	Yes
ANON-310055	Master's Programme in Industrial Management (2 years), 2022	Yes
ANON-795006	Master's Programme in Industrial Systems Analytics (2 years), 2022	Yes

Table 5. Joint 2022 All Registered

Anonymised ID	Programme	Registered
ANON-310055	Master's Programme in Industrial Management (2 years), 2022	Yes
ANON-795006	Master's Programme in Industrial Systems Analytics (2 years), 2022	Yes
ANON-319904	Master's Programme in Industrial Management (2 years), 2022	Yes

3.2. Data Anonymization

Data anonymization serves as a pivotal method for safeguarding sensitive information by converting identifiable data into encrypted identifiers. The core objective is to shield any data that could potentially lead to identification, such as student IDs in our scenario. Various methods of data anonymization are employed today, each offering distinct advantages:

1. **Data Masking:** This technique consists of hiding information within the database and making it unreachable while the database remains usable. Through the use of such approaches as obscuring, we can avoid the instances where our data may be leaked or where some of the data's potential security threats may be present.
2. **Generalization:** Generalization is the process of presenting data to a lesser degree of detail, replacing specific numbers with a set of data. For example, substituting quantitative data with minimum and maximum values that bound it improves data privacy while not impairing usability.
3. **Pseudonymization:** Pseudonymization involves substituting actual identity values with synthetic ones, thus preventing the exposure of sensitive data. Although this technique is not an absolute substitute for all the data, it offers a form of security against data access by unauthorized people.
4. **Data Swapping:** Data swapping involves altering and exchanging data elements while making certain that the original data is not discerned. Thus, combining data provides for the protection of the data's integrity and confidentiality in general while preserving the security of the specific information.

In this thesis, the process of anonymization was crucial to ensure that the student ID information in the data set is not easily recognizable. In order to do this, a formula in excel was used to produce numbers that corresponded with each student id. The formula used contained substrings from the student IDs to maintain the anonymity of respondents while at the same time addressing issues of validity and reliability. The formula is mentioned as follows:

"ANON-" & RIGHT (B2, LEN(B2) – FIND ("1.2.246.562.24.", B2) - 19) (1)

This formula extracted certain parts of the student IDs to anonymize the data without affecting the format or functionality of the dataset. The measures we took included applying data anonymization techniques to ensure that the data was protected and could be used for analysis while preserving the anonymity of the individuals involved.

3.3 Data Pre-processing

The dataset provided was raw and not well structured hence needed a lot of pre-processing before the analysis. To optimize conversion and integrate with Azure Data Factory, Excel sheet data was translated into numerous CSV files. For example, the Joint 2022 admission data was segmented into the following CSVs:

- Joint 2022 all applications
- Joint 2022 all offers
- Joint 2022 all accepted
- Joint 2022 all registered

Following conversion, standardization was enforced by aligning columns uniformly. Therefore, certain columns had to be manipulated to obtain relevant information because some of the data were not presented in a standardized format. For instance, in columns like 'Citizenship', 'Graduation year', 'Previous degree level', 'Previous university', and 'Country of previous university', extraneous characters surrounded the relevant information (e.g., "#0: Ghana,"). A formula was applied to extract the desired data:

=MID (L2, SEARCH ("#0: ", L2) + LEN ("#0: "), SEARCH (",", L2) - (SEARCH ("#0: ", L2) + LEN ("#0: "))) (2)

Likewise, identifying university names from the ‘Name and postal address of the institution’ field is a complicated task due to the inconsistency in data entry. Although no specific pattern was evident, a formula was devised to extract relevant data:

=LEFT(A2, MIN(IFERROR(SEARCH("-", A2), LEN(A2)+1), IFERROR(SEARCH("/", A2), LEN(A2)+1), IFERROR(SEARCH(",", A2), LEN(A2)+1), IFERROR(SEARCH("(", A2), LEN(A2)+1), IFERROR(SEARCH("_", A2), LEN(A2)+1), IFERROR(SEARCH(".", A2), LEN(A2)+1), IFERROR(SEARCH(":", A2), LEN(A2)+1), IFERROR(SEARCH(";", A2), LEN(A2)+1)) - 1) (3)

Additionally, the 'Programme' column in all applications data sources contained multiple applied programs within a single cell, necessitating accurate data splitting. This was achieved using Power Query in Microsoft Excel. Furthermore, discrepancies in the format of the 'Programme' column between data sources required alignment to ensure consistency for effective data integration in Azure Data Factory:

=IFERROR (LEFT(A1, FIND(")", A1)), A1) (4)

Lastly, non-breaking spaces were removed from the entire dataset using the formula:

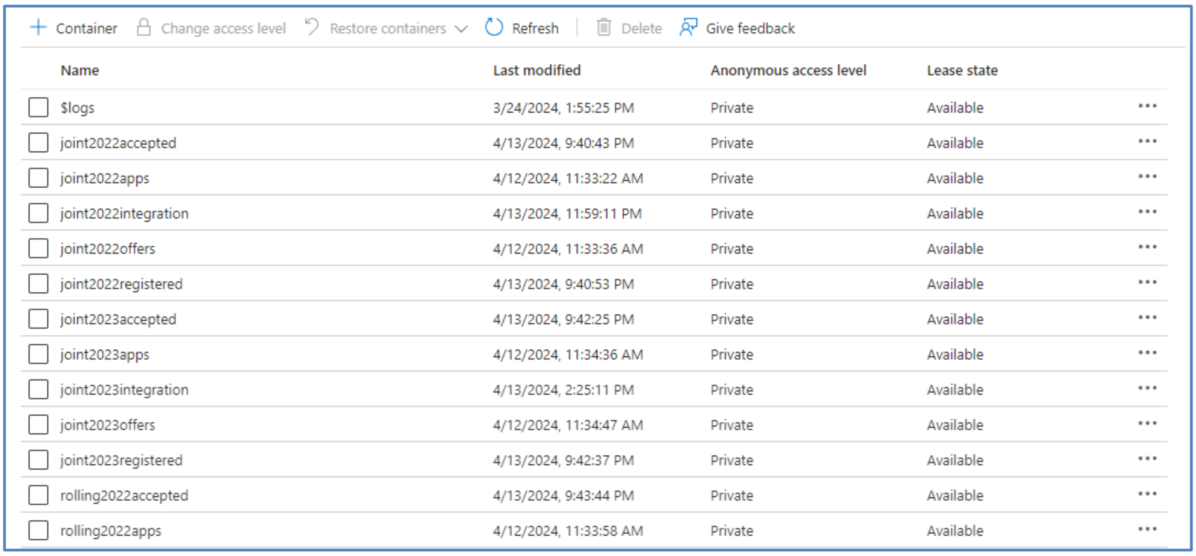
=TRIM(CLEAN(C4)) (5)

The ‘Remove Duplicates’ option under the Data Tools group in Excel was also used to eliminate duplicate rows in order to maintain data quality.

3.4 Data Integration

Data integration remains central and critical to the data processing process since it enables the generation of a unified view and a central system of reference. This makes it easier to make decisions based on data because one is able to have the big picture of how the data is arranged. It helps to avoid the need to manually combine data from multiple sources, provide a complete picture of the customer, and address such concerns as data ownership.

The following are some of the tools that are used for data integration namely, Talend, Airflow, SSIS and others (Darius, 2024). In this thesis, Azure Data Factory (ADF) is applied, which is a very efficient managed service (Rawat, 2018). ADF is known for its data processing capabilities and its ability to perform large-scale data integration tasks. ADF incorporates over 90 data sources and thus offers flexibility in case of future data sourcing requirements. It enables automation of the integration process, data handling and managing of large volumes of data and reduction of integration faults. ADF is super-fast and offer so many connectors. The best advantage of using ADF is that it can also collaborate with other Azure cloud technologies like Azure Databricks, Azure Synapse Analytics and others. The process of integrating data from several sources into one data set is depicted in Figure 2.



Name	Last modified	Anonymous access level	Lease state
$logs	3/24/2024, 1:55:25 PM	Private	Available
joint2022accepted	4/13/2024, 9:40:43 PM	Private	Available
joint2022apps	4/12/2024, 11:33:22 AM	Private	Available
joint2022integration	4/13/2024, 11:59:11 PM	Private	Available
joint2022offers	4/12/2024, 11:33:36 AM	Private	Available
joint2022registered	4/13/2024, 9:40:53 PM	Private	Available
joint2023accepted	4/13/2024, 9:42:25 PM	Private	Available
joint2023apps	4/12/2024, 11:34:36 AM	Private	Available
joint2023integration	4/13/2024, 2:25:11 PM	Private	Available
joint2023offers	4/12/2024, 11:34:47 AM	Private	Available
joint2023registered	4/13/2024, 9:42:37 PM	Private	Available
rolling2022accepted	4/13/2024, 9:43:44 PM	Private	Available
rolling2022apps	4/12/2024, 11:33:58 AM	Private	Available

Figure 2: Data Integration

Data Integration encompasses several key steps:

1. **Instantiating Azure Cloud Resources:**
    - Resource Group: A container for organizing and managing all resources.
    - Storage Account: Facilitates storage of student application data.
    - Azure Data Factory: Dedicated to data integration tasks.
    - Server: Essential for creating an Azure SQL Database.
    - Azure SQL Database: Stores, processes, and enables querying of data.
2. **Moving Data to Azure Storage Account:**
    - Converted CSV files are transferred to Azure storage for enhanced manageability and accessibility by Azure Data Factory.
    - Containers within the storage account are created to organize and store the CSV files efficiently.

3.5 Creation of Linked Services in Azure Data Factory

Linked services serve as the bridge between Azure Data Factory (ADF) and other resources, enabling seamless data transfer and integration. In this thesis, we established two key linked services as shown in figure 3:

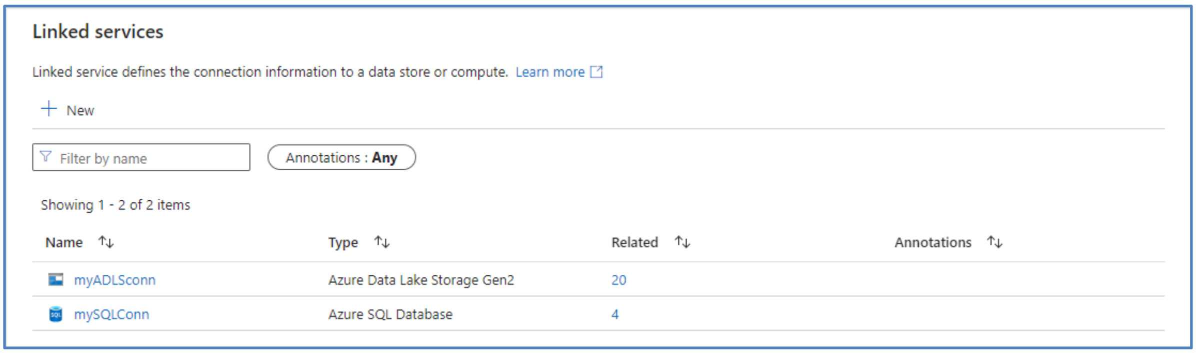


Figure 3: Creation of Linked Services in Azure Data Factory

1. *Linked Service between ADF and Azure Data Lake Storage (ADLS):*
    - This linked service facilitates the extraction of data from Azure Data Lake Storage. Data stored in containers within ADLS is accessed by ADF for further processing.
2. *Linked Service for Loading Data into Azure SQL Database:*
    - Another linked service was created to load integrated data into Azure SQL Database. Once processed and transformed by ADF, the data is loaded into Azure SQL Database for storage and querying.

Both linked services play a critical role in the data integration pipeline, enabling efficient movement of data between various Azure services. This ensures data accessibility and availability for downstream processing and analysis.

3.6 Creation of Datasets in Azure Data Factory

Once the linked services are established, the next step involves creating datasets to pinpoint the exact location of the data. As shown in figure 4, each dataset corresponds to the location of CSV files stored in Azure Data Lake Storage (ADLS) containers.

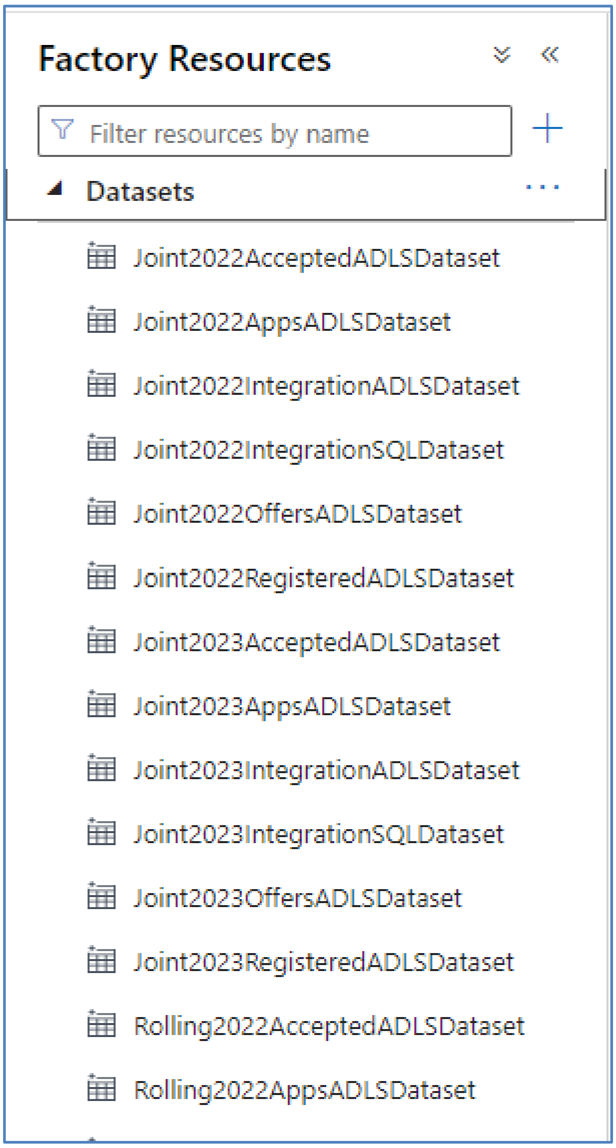


Figure 4: Creation of Datasets in Azure Data Factory

3.7 Creation of Data Pipeline and Dataflows in Azure Data Factory

With the necessary prerequisites in place, we proceed to create a data pipeline named "Student_Recruitment_Integration," encompassing four distinct dataflows, each tailored to a specific set of admission files. For instance, in the case of Joint 2022 admission files, the dataflow consists of the following activities:

3.7.1 Source Creation:

- Two sources are created within the dataflow, each pointing to the location of all applications and all offering data sources in ADLS.

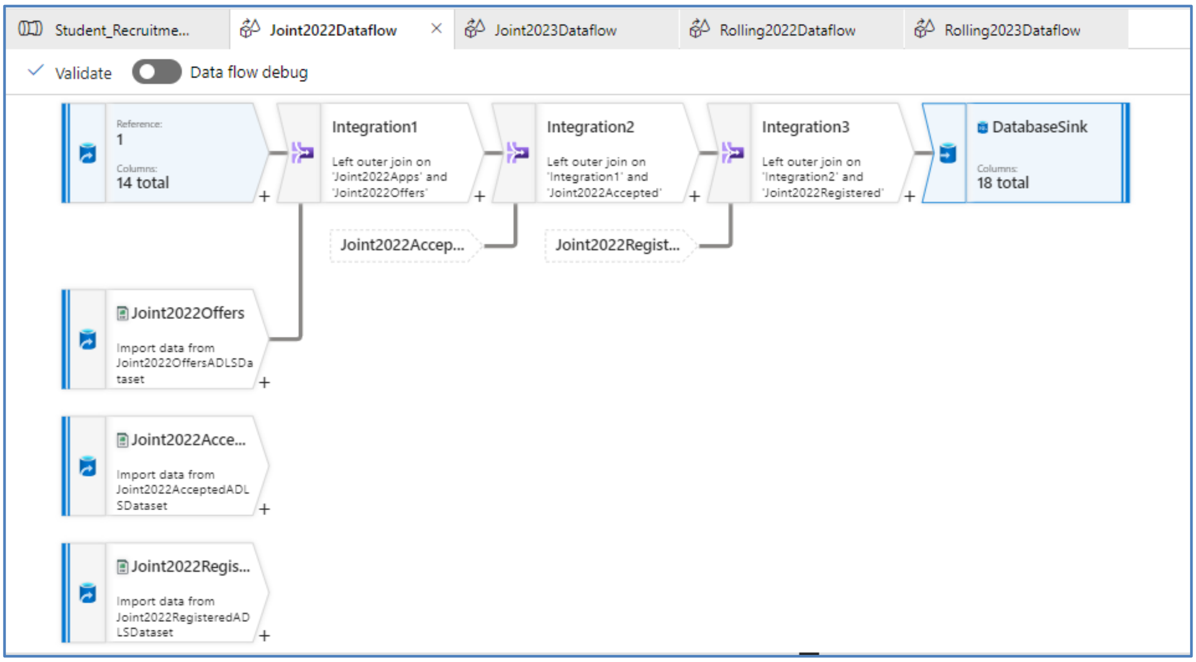


Figure 5: Joint 2022 Dataflow

3.7.2 Join Activity:

- The join activity, named "Integration1," combines the data from all applications and all offered sources. The left join uses all applications data as master data on the left and all offers data on the right. The composite keys “Anonymised ID” and “Programmes” are used to link the tables to avoid any duplication of records.

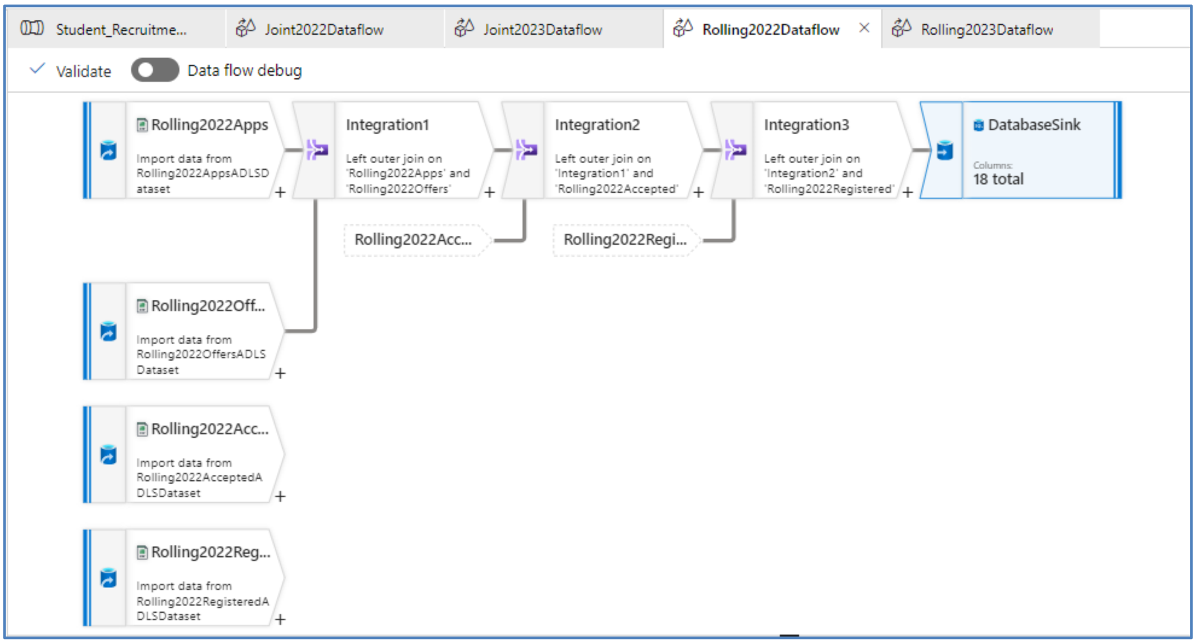


Figure 6: Rolling 2022 Dataflow

3.7.3 Similar Configuration for Other Data Sources:

- The same join configurations are also used for all accepted and all registered data sources and this promotes uniformity in the dataset.

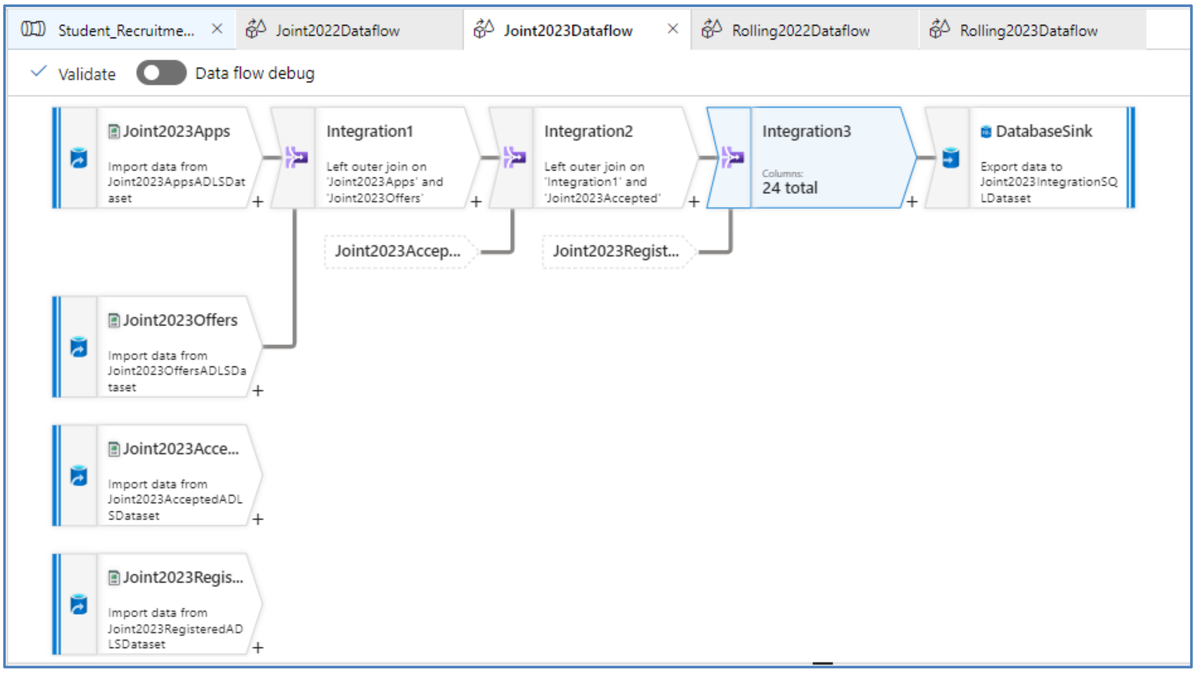


Figure 7: Joint 2023 Dataflow

3.7.4 Sink Activity:

- Lastly, the integrated data is imported into Azure SQL Database by using the Sink activity.

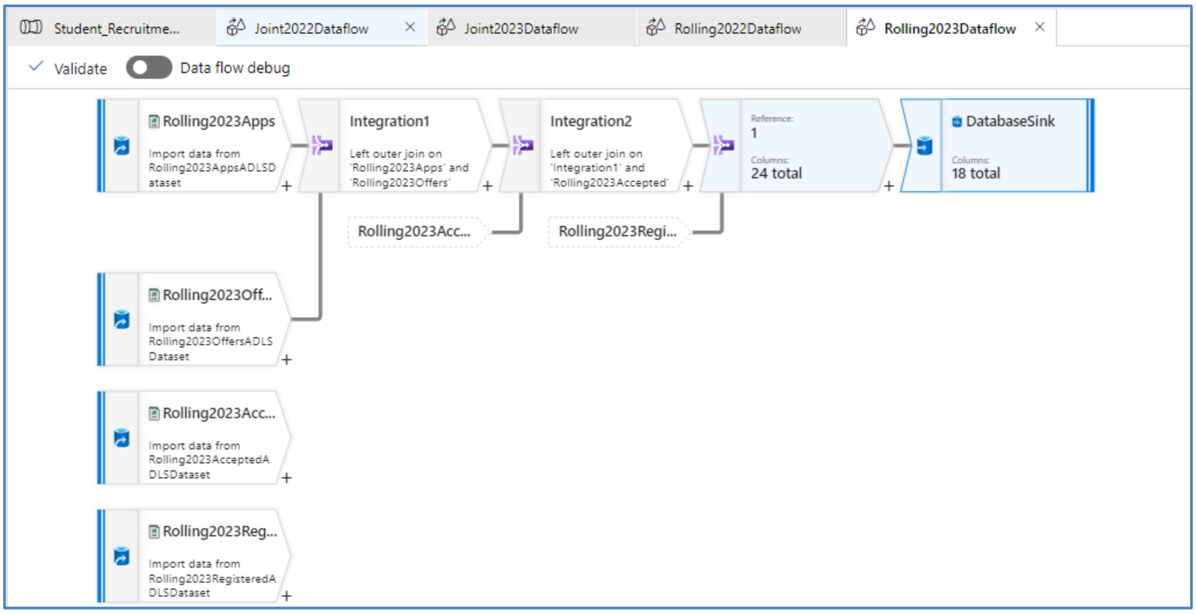


Figure 8: Rolling 2023 Dataflow

This process is repeated for other student recruitment data files, namely the Rolling 2022 dataset, the Joint 2023 dataset, and the Rolling 2023 dataset to ensure standardization and consistency in data management across the different admission sets.

By using these data pipelines and dataflows, we are able to integrate data in a more efficient manner while at the same time load the data into Azure SQL Database for analysis and use.

3.8 Data Loading

After the data is integrated with Azure Data Factory, the integrated data is then stored in Azure SQL Database and generates four tables, where each table contains the integrated data for one single admission file. Azure SQL Database is a powerful and reliable service provided by Azure Cloud to help manage relational database systems. Indeed, Azure SQL Database is known for its scalability and availability enabling the processing of large data sets and generation of results almost instantly.

In addition, the integrated data is also stored in SQL Server Database, which has similar functionalities to Azure SQL Database. To connect to SQL Server is beneficial as it integrates the on-premises and the cloud in one connection. This approach offers the versatility of the two and makes it easier to write SQL for executing different kinds of business tasks and operations.

3.9 Data Visualization

KPIs are well-understood when applied with data visualization as a tool to comprehend such indicators. Since we had the data in the MS SQL Server, we took advantage of the visualisation power of Power BI to realize the potential of the data. Upon this, the data naturally organized into multiple tables within Power BI after import. However, before creating visuals, it was important to prepare and analyze the data with the aid of Power BI’s Data Analysis Expression (DAX). This was done with the use of several DAX measures within the Data View tab of Power BI to come up with new measures and metrics.

Moreover, Power Query was very helpful in the process of cleaning and reshaping the data as required, to make it suitable for analysis. Since the data was sorted out in the programs for each applicant, we began to examine these programs as separate entities and at the same time, find value in the other related fields.

To establish the total number of applicants in different data sheets, we applied some predetermined DAX measures. These measures were used to lay the foundation for our subsequent data analysis and visualization work, and provided a solid basis for our understanding of performance metrics and trends in our dataset.

```
Unique Students per Country J22 =
CALCULATE(
   DISTINCTCOUNT('Joint2022IntegrationTable'[Anonymised ID]),
   ALLEXCEPT('Joint2022IntegrationTable', Joint2022IntegrationTable[Citizenship])
)
 Unique Students per Age J22 =
CALCULATE(
   DISTINCTCOUNT('Joint2022IntegrationTable'[Anonymised ID]),
   ALLEXCEPT('Joint2022IntegrationTable', Joint2022IntegrationTable[Age J22])
)
 Unique Students per City J22 =
CALCULATE(
   DISTINCTCOUNT('Joint2022IntegrationTable'[Anonymised ID]),
   ALLEXCEPT('Joint2022IntegrationTable', Joint2022IntegrationTable[City of Residence
J22])
)
 Unique Students per CountryofRes J22 =
CALCULATE(
   DISTINCTCOUNT('Joint2022IntegrationTable'[Anonymised ID]),
   ALLEXCEPT('Joint2022IntegrationTable', Joint2022IntegrationTable[Country of
residence])
)
```

```
Unique Students per Gender J22 =
CALCULATE(
   DISTINCTCOUNT('Joint2022IntegrationTable'[Anonymised ID]),
   ALLEXCEPT('Joint2022IntegrationTable', Joint2022IntegrationTable[Gender])
)
```

To calculate and visualize applicants who were offered, accepted and registered in the programme, following DAX measures were used:

```
Count Offered J22 = COUNTROWS(
   FILTER(
      Joint2022IntegrationTable,
      Joint2022IntegrationTable[Offer] = "Yes"
   )
)

Count Accepted J22 = COUNTROWS(
   FILTER(
      Joint2022IntegrationTable,
      Joint2022IntegrationTable[Accepted the offer] = "Yes"
   )
)

Count Registered J22 = COUNTROWS(
   FILTER(
      Joint2022IntegrationTable,
      Joint2022IntegrationTable[Registered] = "Yes"
   )
)
```

It was also needed to extract the city out of the ‘city and country’ column. Different applicants have used different ways of writing city and country as there was not any fixed format available for the at studyinfo. Following DAX measure has used to extract city by looking at different data entries, but still it has some limitations.

```
City of Residence J22 =
VAR FullText = Joint2022IntegrationTable[City and country of residence]
VAR Delimiters = {";", ",", " ", "and", "/", "-"}
VAR FirstDelimiterPosition =
   MINX(
      FILTER(
         ADDCOLUMNS(
            Delimiters,
            "Pos", FIND([Value], FullText, 1, LEN(FullText) + 1)
         ),
         [Pos] > 0
      ),
      [Pos]
   )
RETURN
IF(
   FirstDelimiterPosition = LEN(FullText) + 1,
   FullText,
   LEFT(FullText, FirstDelimiterPosition - 1)
)
```

Following DAX measure was used to calculate age from the 'Date of Birth' column.

Age J22 = DATEDIFF(Joint2022IntegrationTable[Date of birth], TODAY(), YEAR)

To calculate day, month and year out of application submission time, following DAX measures were applied.

Application Submission Day J22 = DAY(Joint2022IntegrationTable[Application submission time])

Application Submission Month J22 = MONTH(Joint2022IntegrationTable[Application submission time])

Application Submission Year J22 = YEAR(Joint2022IntegrationTable[Application submission time])

Power query was also used to replace values to further clean the data and to change the data types at any stage. For example, data columns in the data were changed to German format using 'Locale' functionality in Power Query. Several charts have been used to create data visualizations like tree map, pie chart, clustered column chart, clustered bar chart, funnel chart, line chart etc (See Chapter 4). Drilldown capabilities were also applied in visualizations to see the details behind the data.

3.10 Data Forecasting

The following code is used for predicting whether a student will register for the programme or not using multiple features like Programme, Citizenship, Data of Birth, Gender, Country of Residence, Offered and Accepted. We are using Random Forest Classifier to make predictions. While other models like SVM and Neural Networks could be used, we have used Random Forest because it fits our scenario as it combines multiple decision trees to make predictions. This helps in reducing the overfitting issue as well. It requires less data preparation and can handle huge datasets. It also helps in identifying feature importance, which we have done in our code.

https://github.com/zeeshanu6/MSc-Thesis-ML-Code.git

In the code, we have first imported multiple libraries and modules required for analysis and making predictions. We are reading the CSV file for Joint 2022 admissions and dropping some columns which are not required. The values of 'Yes' and 'No' are replaced by 1 and 0 respectively in the 'Offer', 'Accepted' and 'Registered' column. Likewise, we have the done the same process with 'Gender' column. Since 'Country of residence' and 'Citizenship' column will have similar values, we have used the same label encoding for them to convert categorical data into numerical data. Age has also been calculated from 'Date of birth' column. We are mapping different programmes in this admission dataset with certain values to play around with numerical data.

Finally, we are choosing our features and target for this dataset. After that, we are splitting the data for testing and training purposes. We are using Random Forest machine learning model and evaluating the accuracy of our model. We are also calculating the values of

Coefficient of Determination, Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). In the end, we are visualizing the feature importance's in this Random Forest model, and decision-making process of one of the trees in the Random Forest model.

Figure 9 shows the detailed steps of the proposed methodology in the thesis in the form of flowchart.

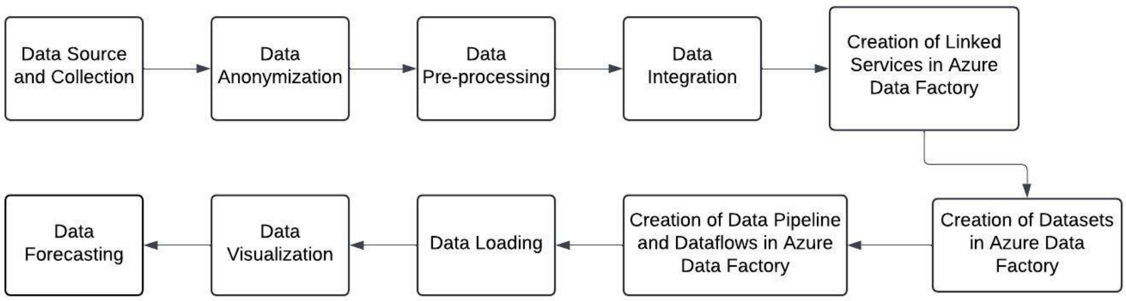


Figure 9: Methodology Steps

Chapter 4: Result

The purpose of this research was to contribute to the improvement of the student recruitment and marketing of universities by applying analysis and decision-making techniques. Many resources and tools have been used for cleaning, integrating, and analyzing data. It is through various levels of data cleaning, data merging, and data mining that various findings were made regarding the recruitment trends and model for future student enrollment.

This study supports data-driven decision-making to increase student's enrollment and marketing. This is because when such data is integrated and analyzed, universities get to learn about the behaviors and patterns of the applicants and this helps the universities to enhance their recruitment strategies and thus improve on the admission rates. Thus, the application of new data tools and methods guarantees the reliability and practical value of these findings, which makes it possible to build effective strategies and make the right decisions.

4.1 Citizenship

Figure 10 shows the distribution of applicants, which provides information on their national origin and gives an idea of the countries from which they come. This information is rather important for the assessment of the demographic background and the international orientation of the applicants. Besides Finland, the other countries from which most of the applicants are from Pakistan, Bangladesh, Sri Lanka, Nigeria, and Ghana. For Joint admissions, we can notice that there is an exponential increase of 910%, 671%, 575%, 525% and 272% in the number of applicants from Pakistan, Bangladesh, Sri Lanka, Nigeria and Ghana respectively. For Rolling admissions, the scenario is somewhat same but citizens from Pakistan and Bangladesh are the leading ones.

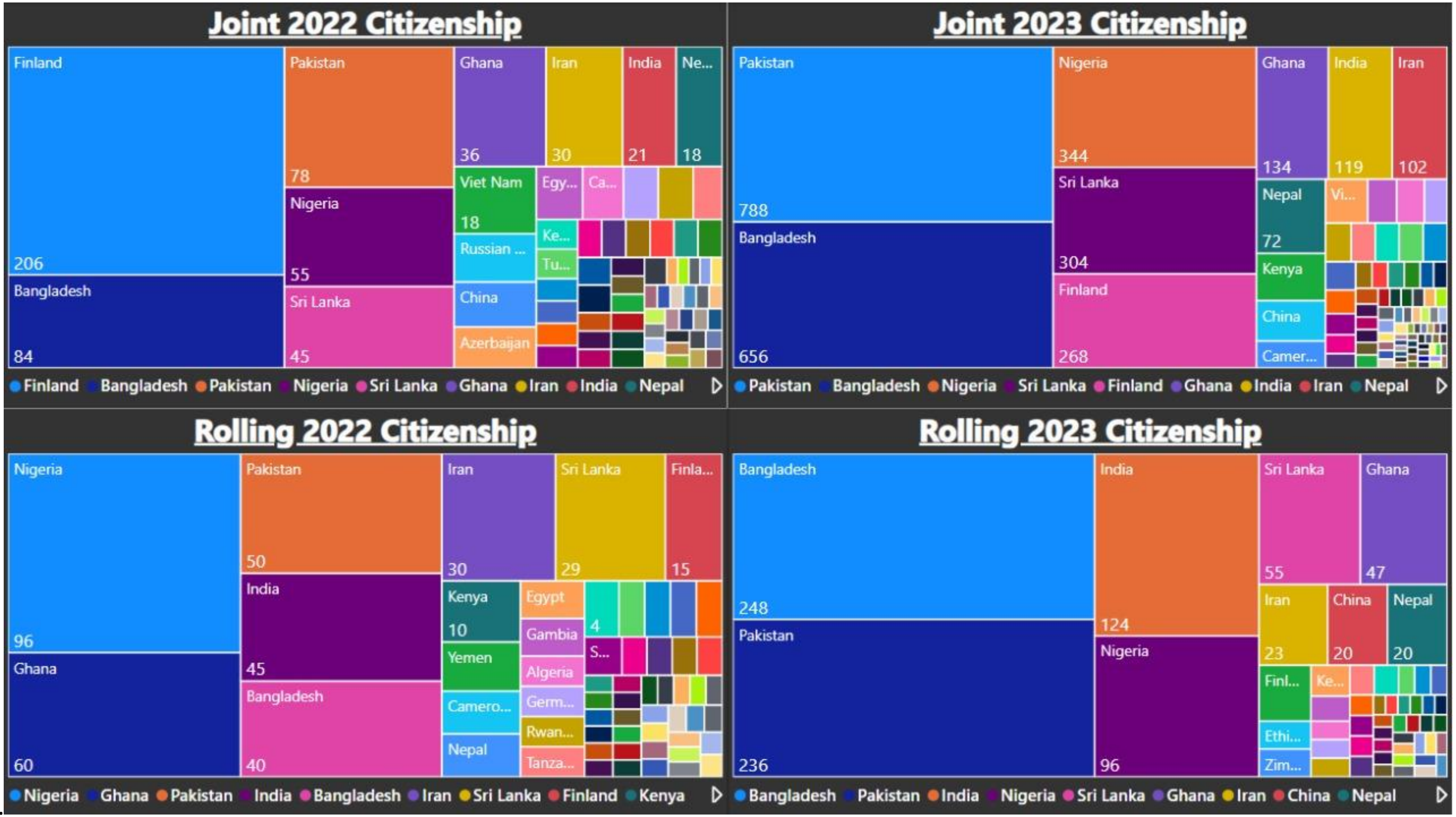


Figure 10: Joint vs. Rolling Admissions Citizenship Distribution

4.2 Age Distribution

The age distribution of applicants for joint and rolling admissions in 2022 and 2023 is depicted in figure 11, which also highlights changes and patterns in the age distribution of the application pool throughout these years. The most common age of applicants is 27 years and around it. The second most common age of applicants is 26 years. The situation is quite similar for both years of Joint applications. Overall, the greatest number of applicants are from 20s in all age distributions.

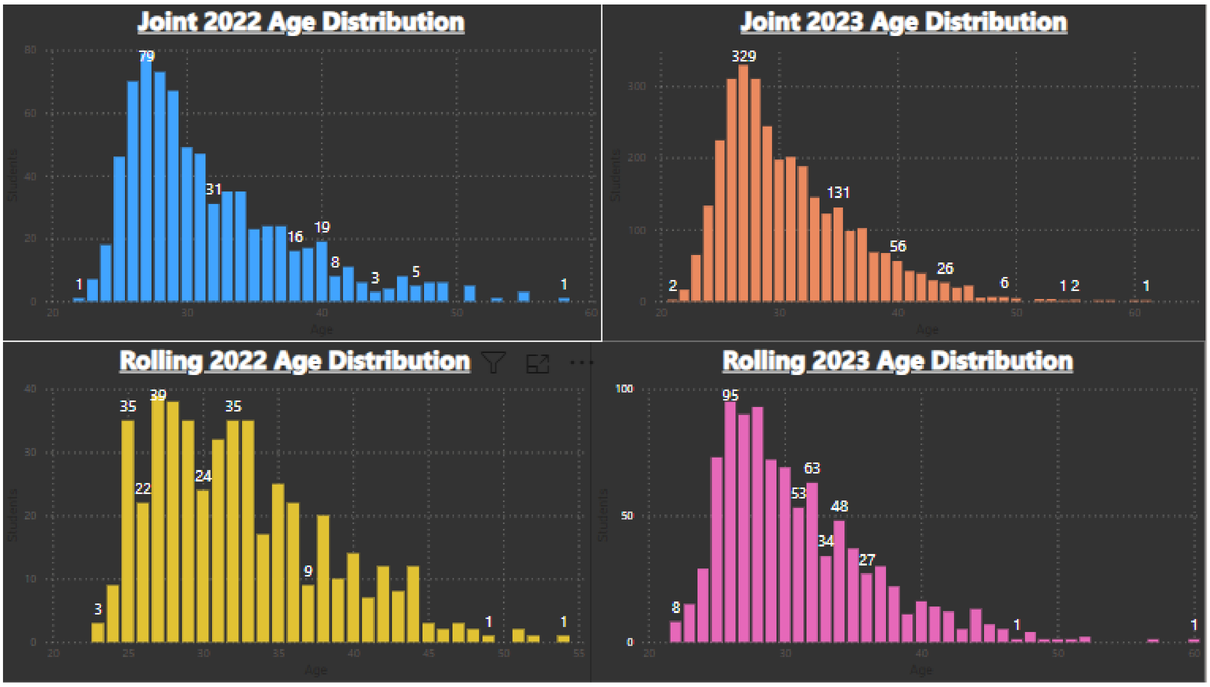


Figure 11: Age distribution insights across applicants

4.3 Country of Residence

The country of residence distribution of candidates for joint and rolling admissions in 2022 and 2023 is shown in figure 12, which also highlights changes and trends in the residence of applicants over these times. Other than Finland, most of the applicants are the residents of Pakistan, Bangladesh, Sri Lanka, Nigeria and Ghana. For Joint admissions, we can notice that there is an exponential increase in the number of applicants from Pakistan, Bangladesh, Nigeria, Sri Lanka and Ghana. For Rolling admissions, the scenario is somewhat same, but residents of Pakistan and Bangladesh are the leading ones.

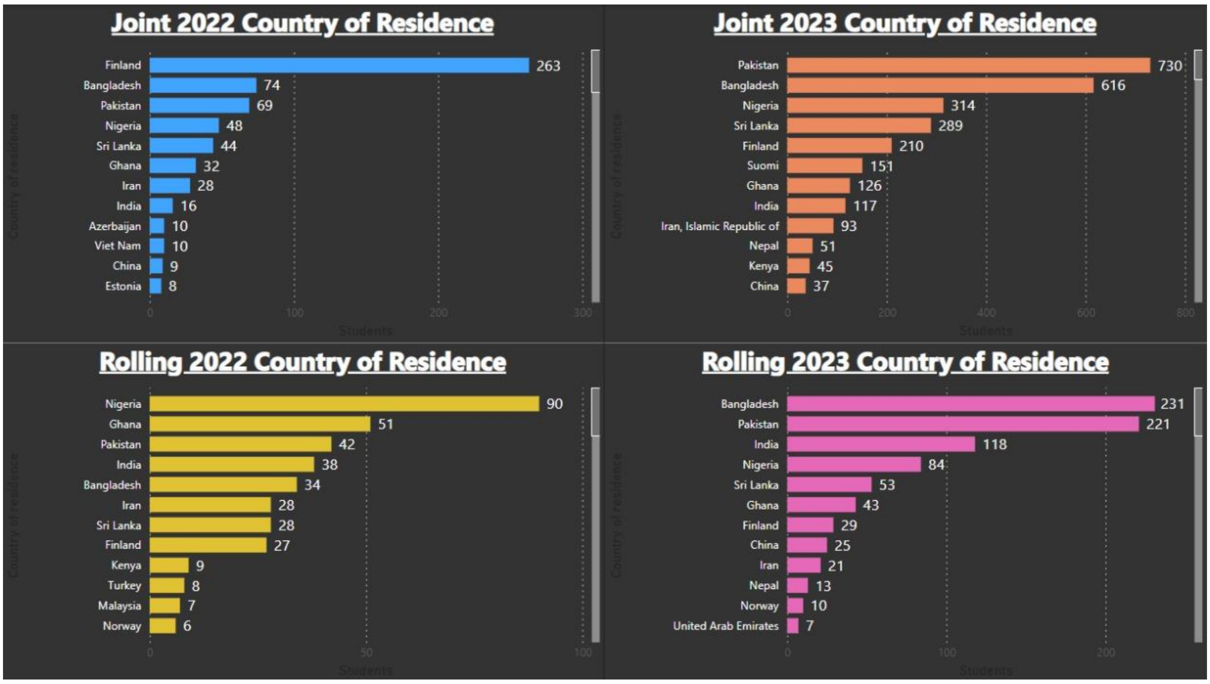


Figure 12: Diversity in the country of residence among applicants

4.4 Joint 2023 City of Residence

Figure 13 shows the trends of applications around multiple cities for Joint 2023. To watch these trends for a particular country, we use the drill-down function in Power BI.

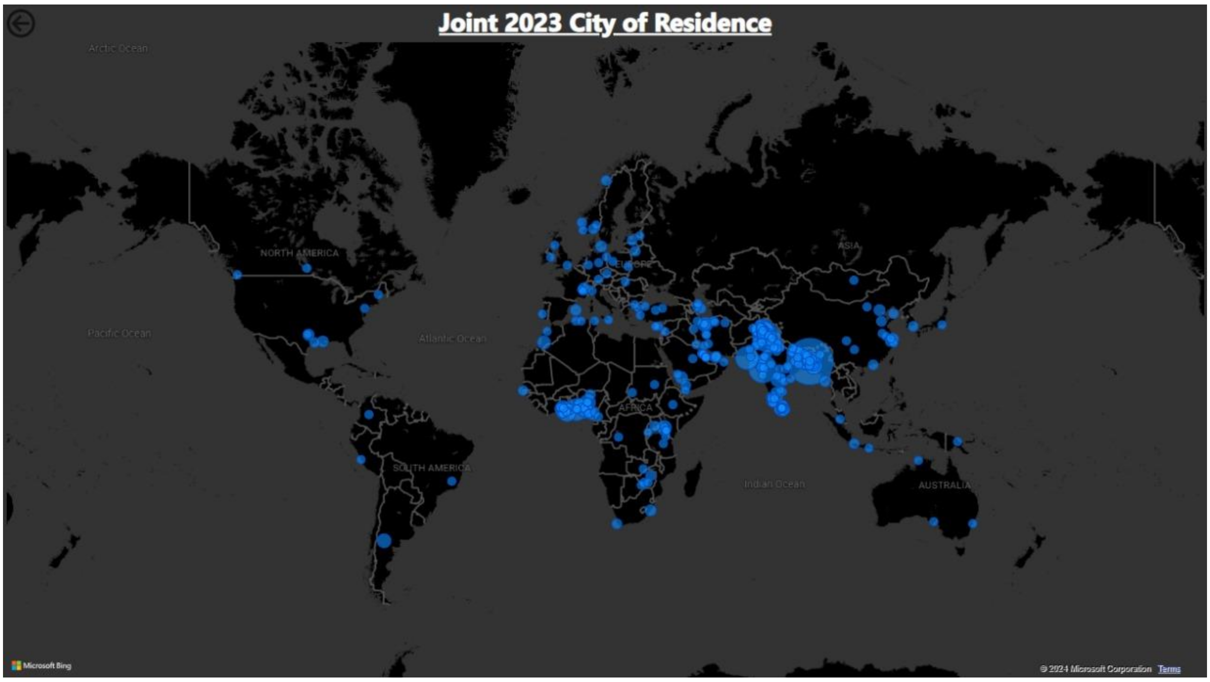


Figure 13: City residence diversity in Joint 2023

4.5 Admission Conversion Rate

The percentage of applicants that moved through the application, acceptance, and registration phases is depicted in figure 14, which also highlights any trends or modifications in the efficiency of the admissions process during these years. The admission conversion rates for joint and rolling admissions in 2022 and 2023 are also shown. The highest conversion rate where maximum students registered for the programme from all applications belongs to Joint 2022 admission where conversion rate is 9.19% whereas the lowest conversion rate belongs to Joint 2023 admissions with conversion rate of 0.84%.

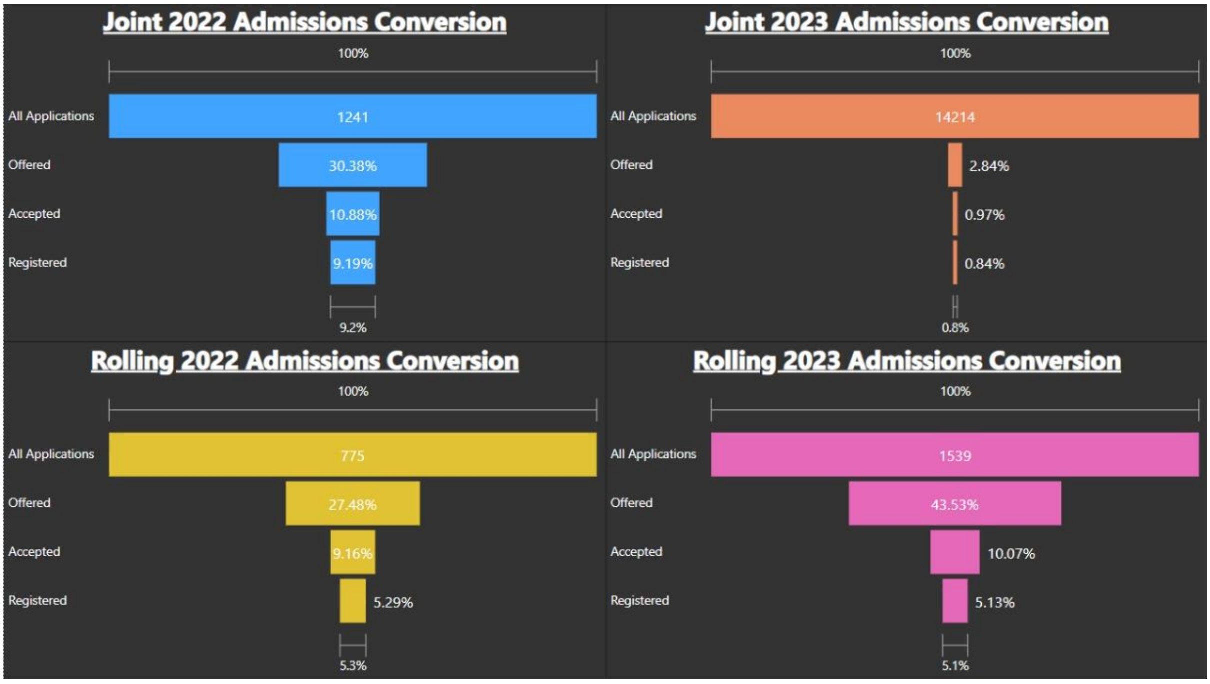


Figure 14: Admission conversion rates analysis

4.6 Application Submission Time

The distribution of application submission times for joint and rolling admissions in 2022 and 2023 is depicted in figure 15, which also highlights patterns and trends in the times at which candidates submitted their applications for each of these years' admission categories.

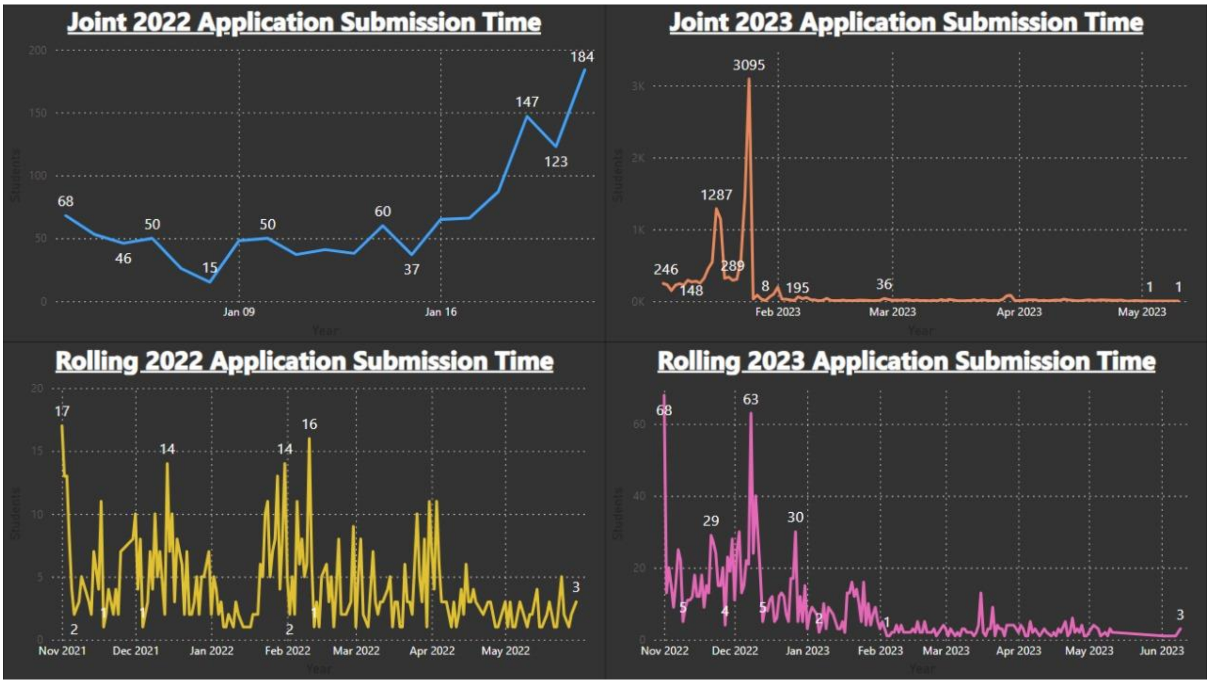


Figure 15: Application submission time trends

4.7 Programmes

Figure 16 illustrates trends and shifts in program popularity and candidate preferences throughout the years, showing the distribution of applicants across various academic programs for both combined and rolling admissions in 2022 and 2023. The programmes of MSc International Business, MSc Strategic Business Development and MSc Industrial Management are among the most applied programmes by applicants.

Figure 16: Program preferences among applicants

4.8 Genders

In order to see if there are any changes or trends in the gender representation of the applicant pool throughout various years and admission types, figure 17 displays the distribution of applicants by gender for 2022 and 2023 joint and rolling admissions. We can clearly see that males are dominant over females with respect to the number of applications for both years.

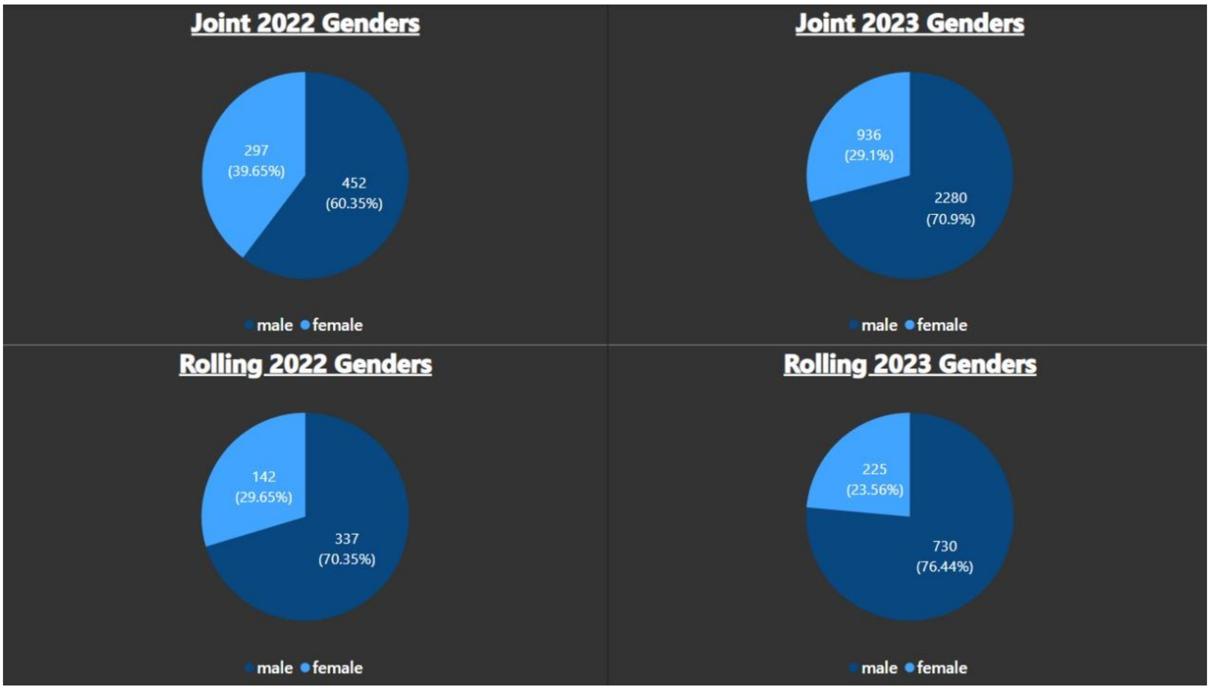


Figure 17: Gender diversity in applicant pool

4.9 Application Submission Time Country for Joint 2022

Figure 18 shows how different nations' application submission times for 2022 joint admissions vary, revealing regional differences. We have selected the countries with most applications for analysis for their application submission time. These results depict that as soon as the deadline for submitting application gets closer, the number of applications is increasing.

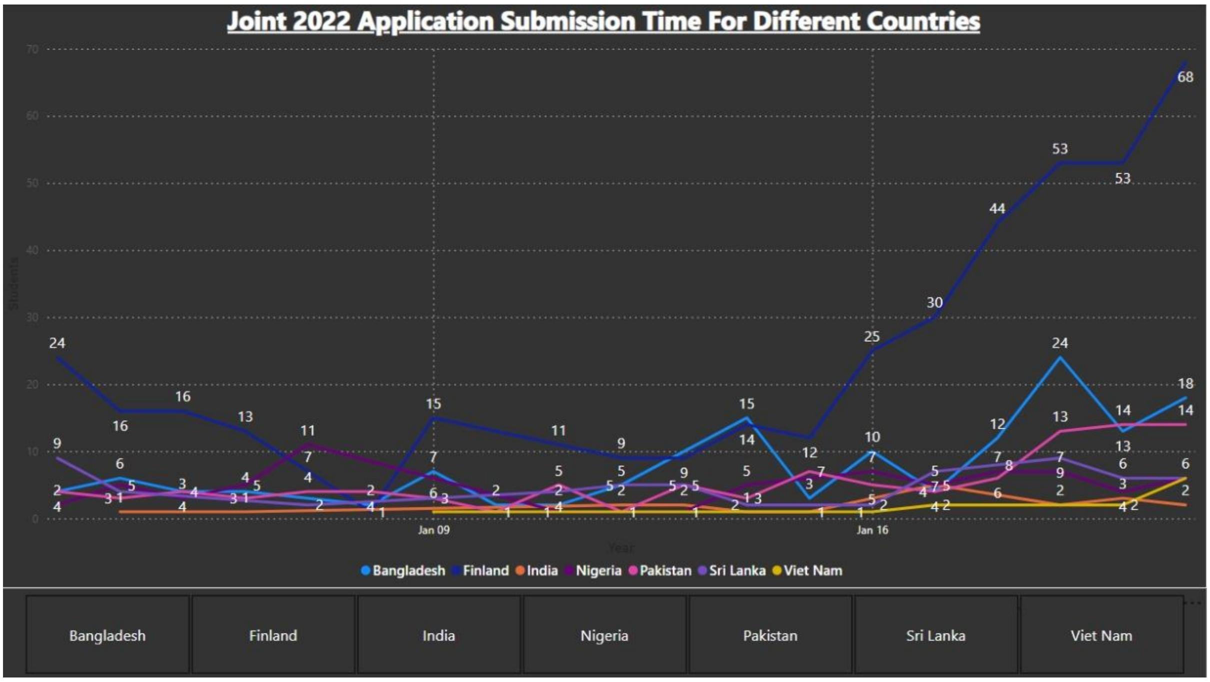


Figure 18: Application submission time vs. Country for Joint 22

4.10 Application Submission Time vs Country for Joint 2023

The distribution of application submission times for joint admissions in 2023 across various countries is shown in figure 19, which provides information about any patterns or shifts in application behavior and timeliness from the year prior. The greatest number of applications from all these countries are seen in the month of January. For other months, the number of applications is quite less than January and there are not many variations.

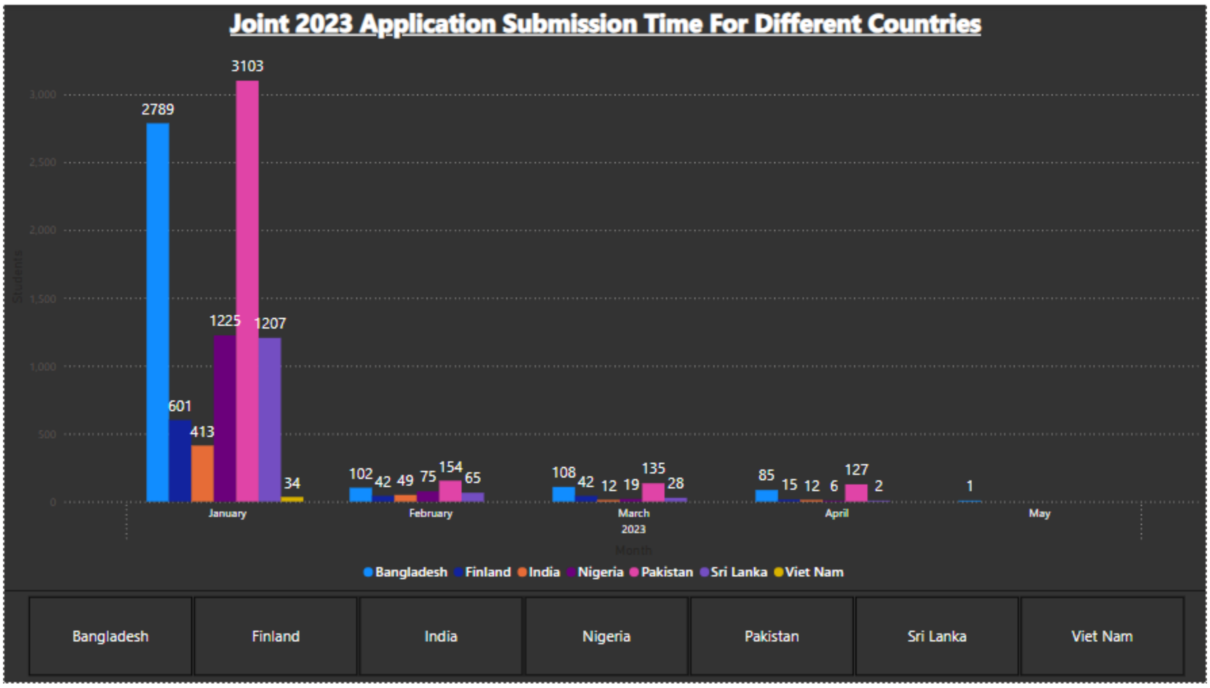


Figure 19: Application submission time vs. Country for Joint 23

4.11 Application Submission Time vs Country for Rolling 2022

Figure 20 provides insights into the temporal trends across various geographic regions by displaying the timeline of application submissions for rolling admissions in 2022 across a variety of countries. Applicants from Nigeria has applied the most in the month of February. Applicants from Pakistan and Bangladesh look more active in the months of December and January. Applicants from Finland and Sri Lanka has applied more from November to January. Overall, the number of applications somewhat closer to each other for all months between November to May.

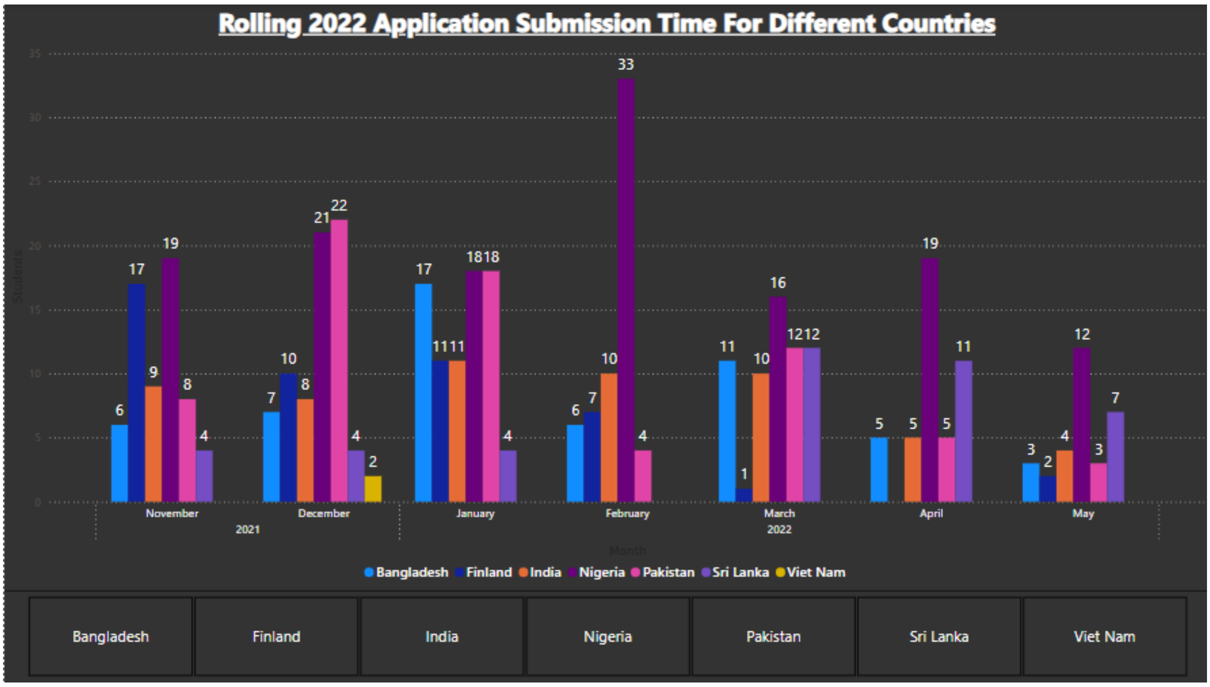


Figure 20: Application submission time vs. Country for Rolling 22

4.12 Application Submission Time vs Country for Rolling 2023

Figure 21 illustrates temporal changes across geographic areas by illustrating the timing of application submissions for rolling admissions in 2023 across several countries. Most number of applications are in the months of November and December. From there on, the number of applications can be seen reducing with time.

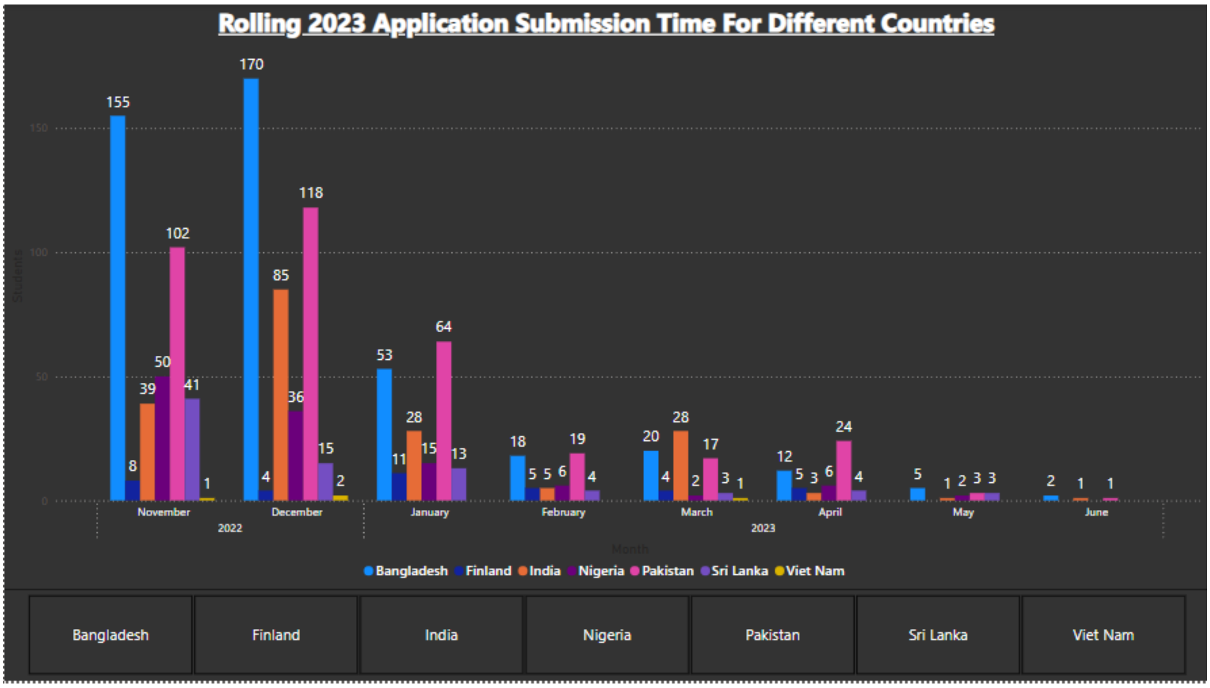


Figure 21: Application submission time vs. Country for Rolling 23

4.13 Previous University

The figure 22 shows the distribution of applicant's prior institutions for joint admissions in 2022 while figure 23 shows the snippet of the same data as a list. It offers information about the educational backgrounds and career paths of potential students, which may be useful for developing recruitment tactics and developing academic programs.

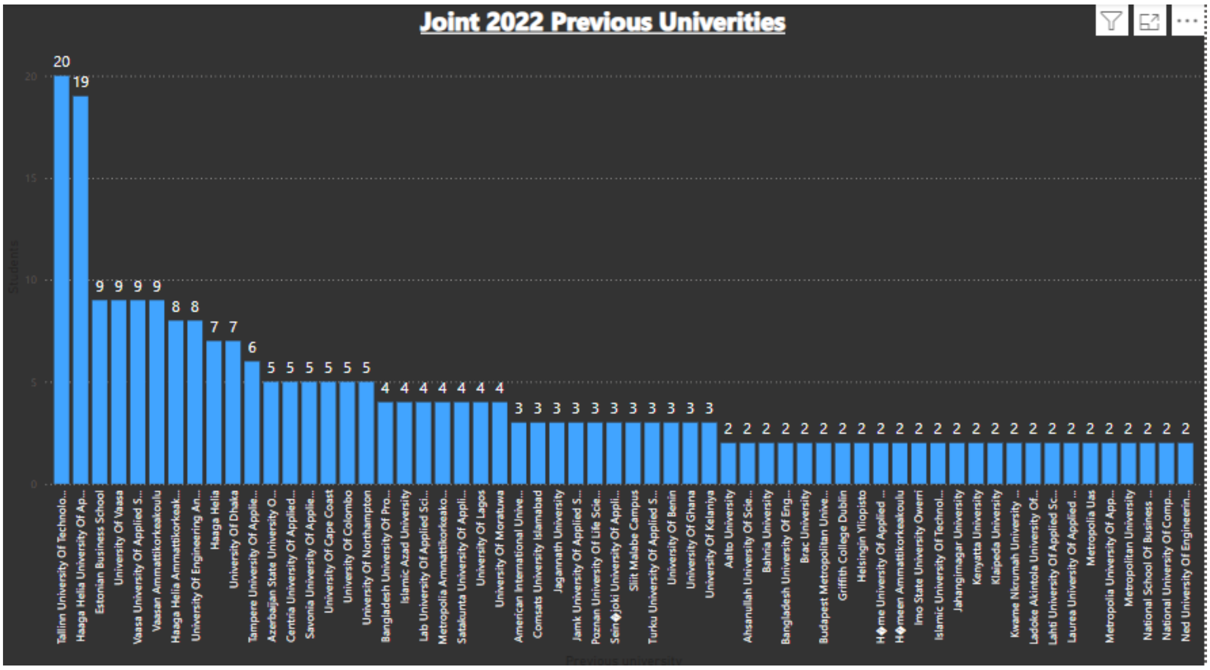


Figure 22: Joint 2022 Previous universities attended

Previous University	No of Applicants
Tallinn University Of Technology	20
Haaga Helia University Of Applied Sciences	19
Estonian Business School	9
University Of Vaasa	9
Vaasa University Of Applied Sciences	9
Vaasan Ammattikorkeakoulu	9
Haaga Helia Ammattikorkeakoulu	8
University Of Engineering And Technology	8
Haaga Helia	7
University Of Dhaka	7
Tampere University Of Applied Sciences	6
Azerbaijan State University Of Economics	5
Centria University Of Applied Sciences	5
Savonia University Of Applied Sciences	5
University Of Cape Coast	5
University Of Colombo	5
University Of Northampton	5
Bangladesh University Of Professionals	4
Islamic Azad University	4
Lab University Of Applied Sciences	4
Metropolia Ammattikorkeakoulu	4
Satakunta University Of Applied Sciences	4
University Of Lagos	4
University Of Moratuwa	4
American International University	3
Comsats University Islamabad	3
Jagannath University	3
Jamk University Of Applied Sciences	3
Poznan University Of Life Sciences	3
Sliit Malabe Campus	3
Turku University Of Applied Sciences	3

Figure 23: Joint 2022 Applicants by Previous University

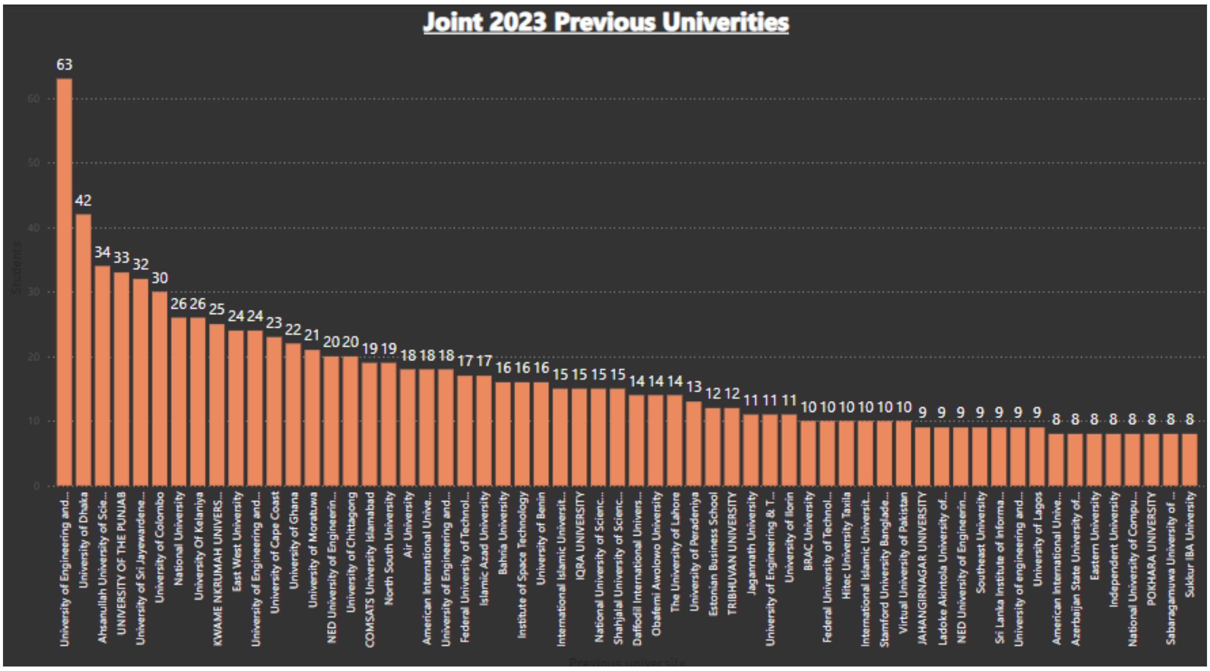


Figure 24: Joint 2023 Previous universities attended

Previous University	No of Applicants
University of Engineering and Technology	63
University of Dhaka	42
Ahsanullah University of Science and Technology	34
UNIVERSITY OF THE PUNJAB	33
University of Sri Jayewardenepura	32
University of Colombo	30
National University	26
University Of Kelaniya	26
KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY	25
East West University	24
University of Engineering and Technology Lahore	24
University of Cape Coast	23
University of Ghana	22
University of Moratuwa	21
NED University of Engineering and Technology	20
University of Chittagong	20
COMSATS University Islamabad	19
North South University	19
Air University	18
American International University-Bangladesh	18
University of Engineering and Technology Taxila	18
Federal University of Technology	17
Islamic Azad University	17
Bahria University	16
Institute of Space Technology	16
University of Benin	16
International Islamic University Islamabad	15
IQRA UNIVERSITY	15
National University of Sciences and Technology	15
Shahjalal University of Science and Technology	15
Daffodil International University	14

Figure 25: Joint 2023 Applicants by Previous University

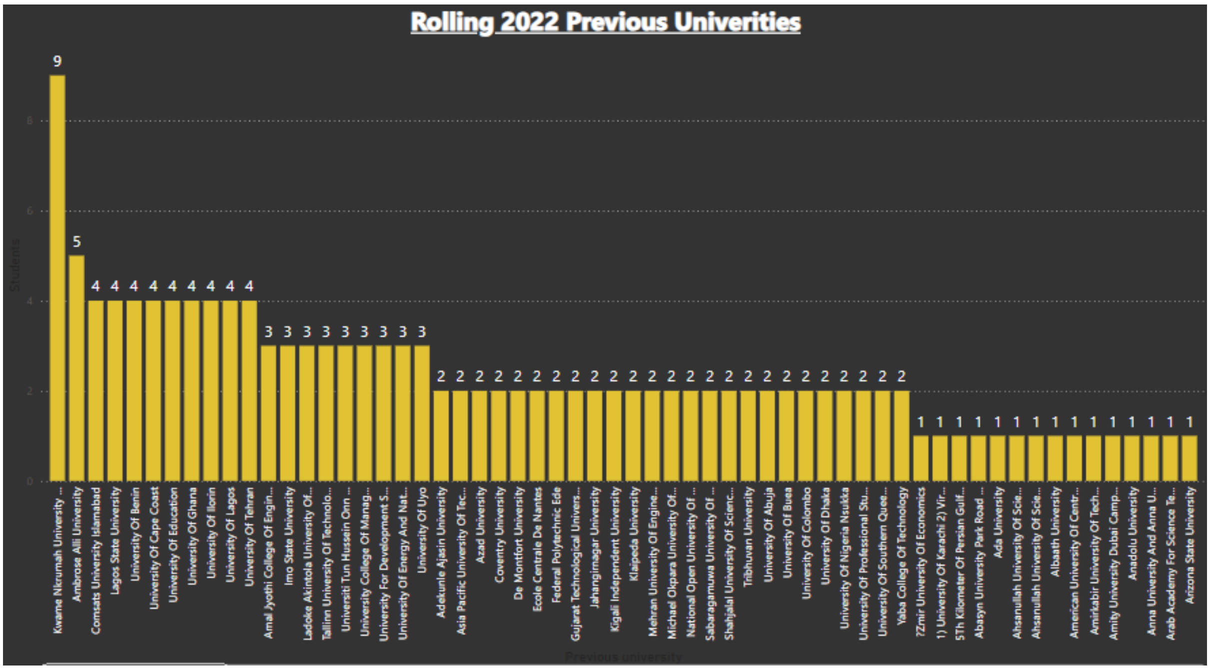


Figure 26: Rolling 2022 Previous universities attended

Previous University	No of Applicants
Kwame Nkrumah University Of Science And Technology	9
Ambrose Alli University	5
Comsats University Islamabad	4
Lagos State University	4
University Of Benin	4
University Of Cape Coast	4
University Of Education	4
University Of Ghana	4
University Of Ilorin	4
University Of Lagos	4
University Of Tehran	4
Amal Jyothi College Of Engineering	3
Imo State University	3
Ladoke Akintola University Of Technology	3
Tallinn University Of Technology	3
Universiti Tun Hussein Onn Malaysia	3
University College Of Management Studies	3
University For Development Studies	3
University Of Energy And Natural Resources	3
University Of Uyo	3
Adekunle Ajasin University	2
Asia Pacific University Of Technology & Innovation	2
Azad University	2
Coventry University	2
De Montfort University	2
Ecole Centrale De Nantes	2
Federal Polytechnic Ede	2
Gujarat Technological University	2
Jahangirnagar University	2
Kigali Independent University	2
Klaipeda University	2

Figure 27: Rolling 2022 Applicants by Previous University

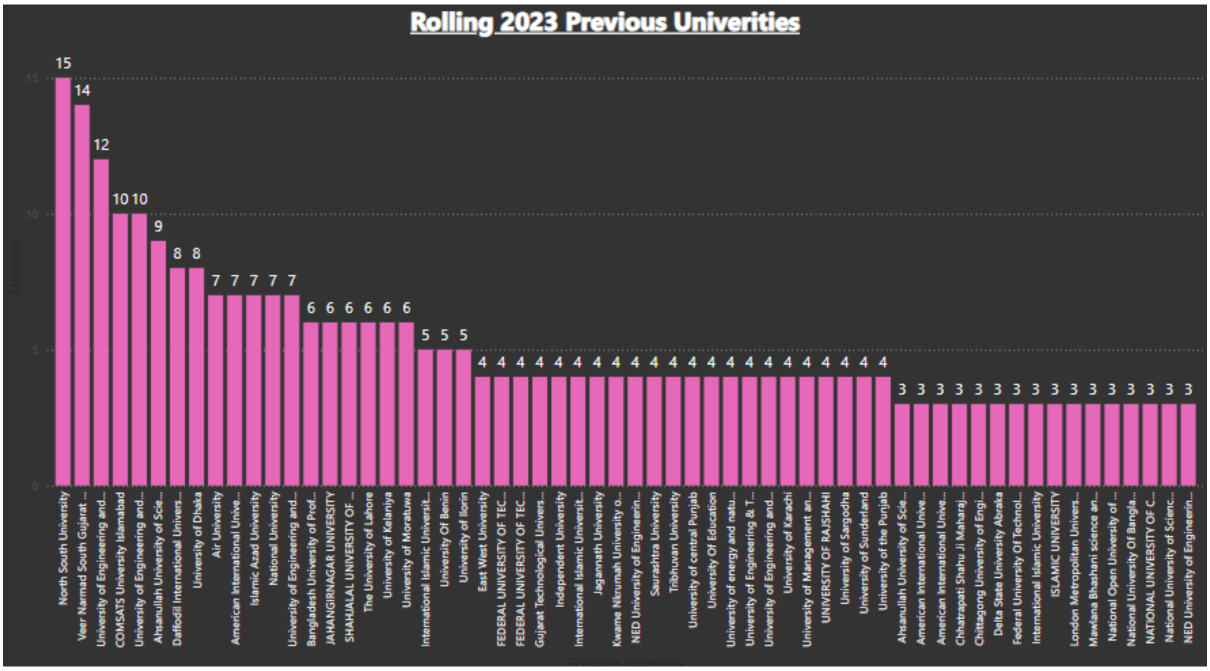


Figure 28: Rolling 2023 Previous universities attended

Previous University	No of Applicants
North South University	15
Veer Narmad South Gujarat University	14
University of Engineering and Technology Lahore	12
COMSATS University Islamabad	10
University of Engineering and Technology	10
Ahsanullah University of Science and Technology	9
Daffodil International University	8
University of Dhaka	8
Air University	7
American International University-Bangladesh	7
Islamic Azad University	7
National University	7
University of Engineering and Technology Taxila	7
Bangladesh University of Professionals	6
JAHANGIRNAGAR UNIVERSITY	6
SHAHJALAL UNIVERSITY OF SCIENCE AND TECHNOLOGY	6
The University of Lahore	6
University of Kelaniya	6
University of Moratuwa	6
International Islamic University Chittagong	5
University Of Benin	5
University of Ilorin	5
East West University	4
FEDERAL UNIVERSITY OF TECHNOLOGY	4
FEDERAL UNIVERSITY OF TECHNOLOGY OWERRI	4
Gujarat Technological University	4
Independent University	4
International Islamic University Islamabad	4
Jagannath University	4
Kwame Nkrumah University of Science and Technology	4
NED University of Engineering & Technology	4

Figure 29: Rolling 2023 Applicants by Previous University

4.14 Forecasting Results

We are extracting and printing the feature importance's in this Random Forest model. This plot, as shown in figure 30 helps us understand which features are most influential in predicting whether a student will register for the program or not.

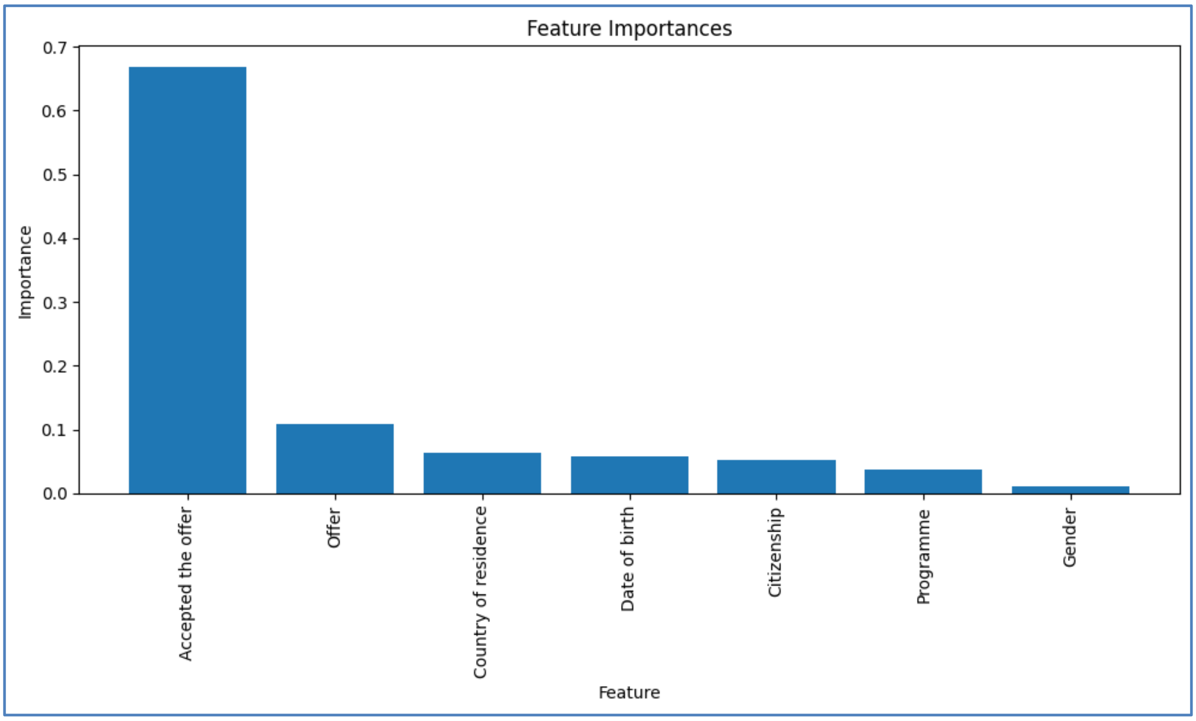


Figure 30: Feature Importance

We also created the visualization from the code which allows us to see the whole structure and decision-making process of one of the trees in the Random Forest model as shown in figure 31. This can be useful for understanding how the model arrives at its predictions, which is particularly valuable for explaining the model's behavior.

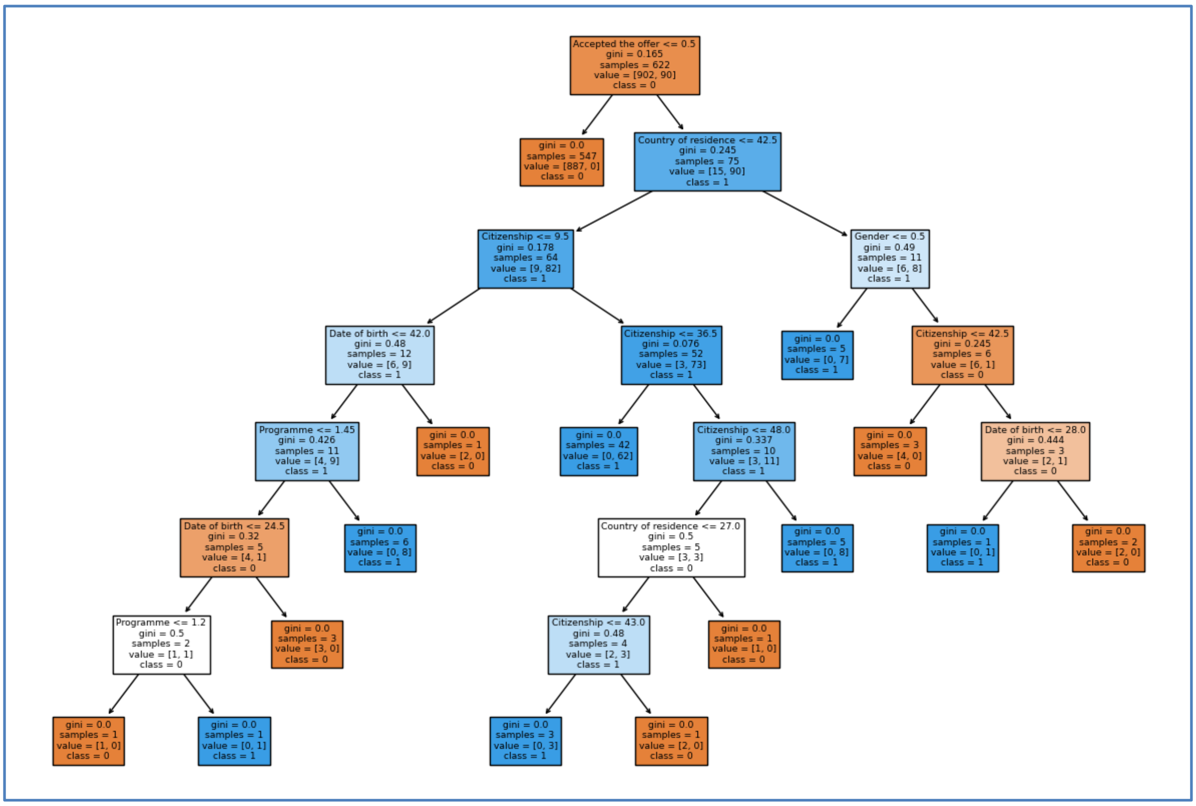


Figure 31: Random Forest Model Structure

This tree basically represents the whole structure by which Random Forest is working and predicting. Each step in this tree represents a feature and evaluates a decision. Based upon that decision, further evaluation occurs accordingly in the next steps. For instance, in the root node, the model is trying to evaluate whether a student will accept the offer or not with 622 training samples and impurity level of 0.165. The value factor shows that 902 students did not register while 90 students registered. Moreover, we can see acceptance of offer being the most influential factor, followed by geographic and demographic information.

The Random Forest model has given us the accuracy of 0.97, which indicates that this model will have 97% of chance in predicting correctly that whether the student will register for the programme or not.

Chapter 5: Discussion and Conclusion

This thesis aimed to enhance data-supported student recruiting and marketing methods for the university. The primary objective is to explore how universities can integrate data-driven approaches to improve recruitment processes, ultimately addressing the research question of “How can universities enhance the student recruitment and marketing strategy through data-driven decision making? “The conclusion made from this study shows that data collection, analysis and deployment used in this thesis enhance the efficiency and sustainability of formal student recruitment through real-time dashboards and data forecasting based on the data generated by these applications.

The present study shows that the application of the data-driven approach in developing recruitment-related processes provides universities with a competitive understanding of students’ behaviors, preferences, and outcomes. Using original quantitative data regarding the past and present, institutions can understand when and where to invest and how-to best market to their target demographics. This is in line with the theoretical underpinning of DDDM that incorporates the use of data from different sources in the strategic planning as well as operations of the business entities (Kaspi and Venkatraman, 2023). From the case of its application in recruitment, data analytics also leads not only to the optimization of the marketing plans, but also the strategic development of the institution as a whole. Additionally, the goal of this study is to provide insightful conclusions, practical solutions, and suggestions for altering the admission procedures of higher education institutions.

5.1 Predictive Modeling and Machine Learning

- The results obtained from data forecasting will help student and recruitment team at the university to have a rough prediction that how many students will register for the programme.
- For example, students who are projected to be at risk of dropout can be easily detected so that universities may be able to offer them requisite amenities and assistance in advance (Bustamante & Garcia-Bedoya, 2021).
- These results are giving us the accuracy of 97%, which is a great score.
- We can use any admission dataset to make predictions using this model.

- Enacting this strategy is beneficial because it prevents students from dropping out and enhances students' performance.
- This evidence supports the findings of previous research works as it revealed that predictive analytics has a profound influence on the student recruitment, as well as on the student retention (Nguyen et al. , 2020).

The research also affirms how the concept of marketing and consumer behavior can serve as significant theoretical frameworks for designing efficient recruitment strategies. Here, STP: segmentation, targeting, and positioning helps to identify different students' segments and subsets to be targeted by universities. Overall, this work proves that institutions that are able to develop target recruitment initiatives based on the psychographic characteristics, needs, and preferences of students have a better chance of creating meaningful appeals (Kotler & Armstrong, 2016). Targeted communication strategies employed based on analytical findings ensures more engagements with the intended students boosting the levels of registration.

The findings of this research reveal that various sources of results are rather useful when it comes to the practical application of data to recruit more students. For instance, the example of more detailed case-studies of institutions that have adopted integrated data analytics platforms highlight annual changes in enrolment returns and convert rate as well as changes in student retention rates and risk indicators. These institutions have employed data analysis in the assessment of the behaviors and preferences of the students in the learning process, which then helps the institutions in formulating strategies for recruiting students and enhancing the rates of their retention (Luca & Bazerman, 2021). The results confirm the understanding of data analytical methods as effective strategies useful for university recruiters who want to enhance institution's performance to meet its objectives within the context of increased competition of higher education.

This research sheds light on the belief of the usefulness of analytics, knowledge of predictive analysis, and marketing concepts useful in conducting improved student recruitment strategies. Through such approaches, universities can be better equipped to make informed decisions as well as enhance the methods used in marketing their institutions and consequently increase recruitment as well as retention standards in their learning institutions. The research proves that implementing data driven approaches not only enhances the efficiency of recruiting but also contributes to the tendencies expressed in the development of higher education where data analysis and predictive modeling are viewed as valuable assets for institutional change (Zawacki-Richter et al. , 2019).

In this project, the recruitment and marketing of the university students was done through proper analysis of data. To achieve better results, a big data set from a variety of sources as well as the most effective data analysis instruments were employed to obtain insights and improve the existing method of recruiting.

The main goal was an attempt to create a single source of dynamic data that is used for the purpose of creating useful dashboards for data extraction and analysis. One of the outcomes to emerge from the application of this method involved the creation of KPIs from the student application files in order to enhance decision-making. These sources revealed that the university retained applicants from principally Pakistan, Bangladesh, Sri Lanka, Nigeria, and Ghana thus broadening its international outreach.

The survey found that most applicants were 27 years old, offering a demographic target for marketing. The greatest conversion rate was 9.19% in Joint 2022 admissions and the lowest was 0.84% in Joint 2023 admissions, suggesting admissions process improvements. Some of the most famous programmes among applicants include MSc International Business, Strategic Business Development, and Industrial Management.

The majority of applications were male, suggesting that increased efforts are needed to encourage female participation. Using machine learning, enrollment patterns were forecasted and notable trends were identified to provide potential institutional planning guidance. They identified temporal variations of application usage in different regions, which might help refine strategies used by recruiters. Lastly, in the context of the study, synthesizing and discussing the accumulating masses of data confirmed that data-driven decision-making might help enhance the processes of students' enrollment. Universities may increase enrollment through recruiting optimization if these institutions analyze candidate data and trends. In light of the results of this study, practitioners would benefit in order to enhance the further development and internationalization of the university.

5.2 Key Limitations

5.2.1 Data Quality

- Firstly, the reliability and validity of the finding depend on the availability of data and the quality of the data which may be distorted due to several factors like missing data or measurement errors.
- It is still possible that there are some built-in biases and/or gaps in information procured from the admissions cycles of 2022 and 2023.
- Moreover, timeliness also refers to how relevant the data is when presented, which is particularly important for the recruiting process in organizations with high turnover rates.
- Such limitations suggest that a more careful approach should be employed when interpreting results and validating the findings, and it is equally important to understand the results that may stem from data quality issues.

5.2.2 Generalizability

- While the insights and recommendations offered here are valuable, they might not be universally applicable to all higher education institutions.
- Institutional characteristics such as size, location, and student demographics could significantly influence the feasibility and effectiveness of implementing data-driven strategies.
- For instance, larger institutions may have more resources and expertise to invest in sophisticated data analytics tools, whereas smaller colleges may face constraints in acquiring and maintaining such technologies.
- These are important considerations to bear in mind concerning important limitations of the study, implying that the credibility of the data and results analysed should be deemed highly relevant, given that variations in data quality can significantly affect the findings.

5.2.3 Ethical Considerations

- Another issue that requires further consideration is the ethical implications of Big data such as privacy concerns and dataset biases along with the capabilities that are needed in implementing them.

- Despite attempting to follow best practices and protect the confidentiality of clients described in the study, there can be potential ethical issues tied to data gathering, assessment, and synthesis.
- Institutions have to act carefully when they embrace the trend of using data for purposes of gaining competitive advantages while not infringing on the rights and freedoms of possible learners.
- Additionally, the possibility of successfully employing data-driven approaches can be a function of available technology in the institution and the capability of staff in the management of data, the need for staff development and organisational investment in infrastructure to underpin data projects.

5.2.4 Studyinfo Considerations

- A lot of changes in the architecture of studyinfo is required to make sure the accuracy and integrity of data. This platform should make sure to restrict applicants from providing random data as much as possible.
- There should be select boxes instead of text fields for the universal data. For example, there should be a select box available to provide a list for cities rather than a text box.
- There should be different fields for the name and postal address of the previous education institution.
- The programme names should be consistent over the years. Also, there should be a specific delimiter applied at the end of each programme because we need to analyze data for each programme.
- The system must stop applicants from entering wrong data in every field. Different rules must be applied for each field through regular expressions.

This thesis has explored the question of data-driven approaches to student recruitment in an attempt to respond to the learning institutions' difficulties in leading the identification, targeting, and acquisition of student prospects. By identifying the features of data analytics for decision making, key KPIs, ethical issues, and technological factors, this study has synthesized valuable lessons for the institutions willing to improve their recruitment mechanisms. The strengths of this study lie in the collected data, while the limitations refer to data availability, generalizability, and ethical constraints of big data in higher education recruitment; nevertheless, it highlights the value of approaches based on data in transforming the recruitment landscape. By employing data and numerous statistical models, universities can get better insights about the students’ population, their preferences and their activities, thus facilitating the effective enrollment of prospective students.

As future work progresses, it is crucial for institutions to stay committed to investing in both technological assets and educate their employees in the adoption of data-driven approaches. Therefore, constant practice and research in the higher education context are equally important to develop the study results to the next level, and to fortify the practice for better recruitment standards following the new market tendency and consistent alternative rules. Finally, by adopting approaches of using analytical data in universities, the organizations not only can increase their overall competitiveness and optimize the distribution of resources but can also contribute to the development of the inclusiveness and diversity of the student population. This thesis can be seen as a call for change for the institutions to enhance the use of innovation and data in the recruitment process for the higher education systems making it efficient, fair and focusing on student’s needs. The future of this study encourages to explore efficient forecasting machine learning approaches to predict the future intake and related factors in the admission.

References

Gilch, P. M., & Sieweke, J. (2021). Recruiting digital talent: The strategic role of recruitment in organisations' digital transformation. German Journal of Human Resource Management, 35(1), 53-82. https://journals.sagepub.com/doi/abs/10.1177/2397002220952734

Fodor, S., Szabó, I., & Ternai, K. (2021). Competence-oriented, data-driven approach for sustainable development in university-level education. Sustainability, 13(17), 9977. https://www.mdpi.com/2071-1050/13/17/9977

Iivari, Sharma, & Ventä-Olkkonen (2020). Digital transformation of everyday life–How COVID-19 epidemic changed youth basic education and why information management research should care. International information management 55:102183. https://www.sciencedirect.com/science/article/pii/S0268401220310264

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education–where are the educators?. International Journal of Educational Technology in Higher Education, 16(1), 1-27. https://link.springer.com/article/10.1186/s41239-019-0171-0

Kuleto, V., Ilić, M., Dumangiu, M., Ranković, M., Martins, O. M., Păun, D., & Mihoreanu, L. (2021). Exploring opportunities and challenges of artificial intelligence and machine learning in higher education institutions. Sustainability, 13(18), 10424. https://www.mdpi.com/2071-1050/13/18/10424

Taylor Jr, L. D. (2020). Neoliberal consequence: Data-driven decision making and the subversion of student success efforts. The Review of Higher Education, 43(4), 1069-1097. https://muse.jhu.edu/pub/1/article/761668/summary

Posselt, J., Hernandez, T. E., Villarreal, C. D., Rodgers, A. J., & Irwin, L. N. (2020). Evaluation and decision making in higher education: Toward equitable repertoires of faculty practice.

Higher Education: Handbook of Theory and Research: Volume 35, 1-63. https://link.springer.com/content/pdf/10.1007/978-3-030-11743-6_8-1.pdf

Singer, C. G. (2023). Educational Data Mining: An Application of a Predictive Model of Online Student Enrollment Decisions (Doctoral dissertation, Arizona State University). https://search.proquest.com/openview/eae78fb3debf367aaf3d96a51584e76e/1?pq-origsite=gscholar&cbl=18750&diss=y

Wang, J., Yang, Y., Wang, T., Sherratt, R. S., & Zhang, J. (2020). Big data service architecture: a survey. Journal of Internet Technology, 21(2), 393-405. https://jit.ndhu.edu.tw/article/view/2261

Beech, S. E. (2020). Adapting to change in the higher education system: International student mobility as a migration industry. In Exploring the Migration Industries (pp. 68-83). Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9780429199660-4/adapting-change-higher-education-system-international-student-mobility-migration-industry-suzanne-beech

Gaftandzhieva, S., Hussain, S., Hilcenko, S., Doneva, R., & Boykova, K. (2023). Data-driven Decision Making in Higher Education Institutions: State-of-play. International Journal of Advanced Computer Science and Applications, 14(6). https://www.academia.edu/download/103956949/Paper_42_Data_driven_Decision_Making_in_Higher_Education_Institutions.pdf

Bell, E., Bryman, A., & Harley, B. (2022). Business research methods. Oxford university press. https://books.google.com/books?hl=en&lr=&id=hptjEAAAQBAJ&oi=fnd&pg=PP1&dq=The+purpose+of+this+literature+review+is+to+analyze+previous+research+on+using+data+to+make+decisions+about+student+recruitment.+It+will+draw+on+important+theoretical+views+to+get+valuable+insights.&ots=Ddkg36x11w&sig=oCzfwteOvt-4xmTPEMFfQtaXHuU

Du, L., & Li, Q. (2020). A data-driven approach to high-volume recruitment: Application to student admission. Manufacturing & Service Operations Management, 22(5), 942-957. https://pubsonline.informs.org/doi/abs/10.1287/msom.2019.0779

Kaspi, S., & Venkatraman, S. (2023). Data-Driven Decision-Making (DDDM) for Higher Education Assessments: A Case Study. Systems, 11(6), 306. https://www.mdpi.com/2079-8954/11/6/306?utm_campaign=releaseissue_systemsutm_medium=emailutm_source=releaseissueutm_term=doilink47

Kallio, T. J., Kallio, K. M., & Blomberg, A. (2020). From professional bureaucracy to competitive bureaucracy–redefining universities' organization principles, performance measurement criteria, and reason for being. Qualitative research in accounting & management, 17(1), 82-108. https://www.emerald.com/insight/content/doi/10.1108/QRAM-10-2019-0111/full/html

Alshanqiti, A., & Namoun, A. (2020). Predicting student performance and its influential factors using hybrid regression and multi-label classification. Ieee Access, 8, 203827-203844. https://ieeexplore.ieee.org/abstract/document/9250366/

Basu, R., Lim, W. M., Kumar, A., & Kumar, S. (2023). Marketing analytics: The bridge between customer psychology and marketing decision-making. Psychology & Marketing, 40(12), 2588-2611. https://onlinelibrary.wiley.com/doi/abs/10.1002/mar.21908

Dukic, J. (2021). Exploring the Impact of Performance-Based Funding Policy Reform: The Role of Institutional Research in Supporting Data-Driven Decision-Making. https://ir.lib.uwo.ca/oip/229/

Stronge, J. H., & Xu, X. (2021). Qualities of effective principals. ASCD. https://books.google.com/books?hl=en&lr=&id=k_kuEAAAQBAJ&oi=fnd&pg=PP1&dq=Through+the+utilization+of+data,+institutions+may+discover+areas+for+development+and+evaluate+the+effectiveness+of+current+efforts,+allowing+them+to+continuously+enhance

+their+processes+in+order+to+better+meet+the+needs+of+their+students+and+stakeholders.&ots=hv3eIzqnwD&sig=KvwAi5FQKd16oV5HXDhtbeEHH9I

Mian, S. H., Salah, B., Ameen, W., Moiduddin, K., & Alkhalefah, H. (2020). Adapting universities for sustainability education in industry 4.0: Channel of challenges and opportunities. Sustainability, 12(15), 6100. https://www.mdpi.com/2071-1050/12/15/6100

Nguyen, A., Gardner, L., & Sheridan, D. (2020). Data analytics in higher education: An integrated view. Journal of Information Systems Education, 31(1), 61. http://jise.org/Volume31/n1/JISEv31n1p61.html

Matz, S. C., Bukow, C. S., Peters, H., Deacons, C., Dinu, A., & Stachl, C. (2023). Using machine learning to predict student retention from socio-demographic characteristics and app-based engagement metrics. Scientific Reports, 13(1), 5705. https://www.nature.com/articles/s41598-023-32484-w

Bustamante, D., & Garcia-Bedoya, O. (2021). Predictive academic performance model to support, prevent and decrease the university dropout rate. In Applied Informatics: Fourth International Conference, ICAI 2021, Buenos Aires, Argentina, October 28–30, 2021, Proceedings 4 (pp. 222-236). Springer International Publishing. https://link.springer.com/chapter/10.1007/978-3-030-89654-6_16

Kerr, C. C., Stuart, R. M., Mistry, D., Abeysuriya, R. G., Rosenfeld, K., Hart, G. R., ... & Klein, D. J. (2021). Covasim: an agent-based model of COVID-19 dynamics and interventions. PLOS Computational Biology, 17(7), e1009149. https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1009149

Nie, M., Xiong, Z., Zhong, R., Deng, W., & Yang, G. (2020). Career choice prediction based on campus big data—mining the potential behavior of college students. Applied Sciences, 10(8), 2841. https://www.mdpi.com/2076-3417/10/8/2841

Dekker, I., De Jong, E. M., Schippers, M. C., De Bruijn-Smolders, M., Alexiou, A., & Giesbers, B. (2020). Optimizing students' mental health and academic performance: AI-enhanced life crafting. Frontiers in Psychology, 11, 1063. https://www.frontiersin.org/articles/10.3389/fpsyg.2020.01063/full

ANTARA, M., & DEWI, N. P. A. Y. (2022). Segmenting, Targeting and Positioning: The Marketing Strategy of Jatiluwih Organic Red Rice, Bali. Denpasar: Pelawa Sari. https://www.academia.edu/download/78305991/Buku_Beras_Merah_Antara.pdf

Rosenbloom, D. H., Kravchuk, R. S., & Clerkin, R. M. (2022). Public administration: Understanding management, politics, and law in the public sector. Routledge. https://www.taylorfrancis.com/books/mono/10.4324/9781003198116/public-administration-david-rosenbloom-robert-kravchuk-richard-clerkin

Vrontis, D., Makrides, A., Christofi, M., & Thrassou, A. (2021). Social media influencer marketing: A systematic review, integrative framework and future research agenda. International Journal of Consumer Studies, 45(4), 617-644. https://onlinelibrary.wiley.com/doi/abs/10.1111/ijcs.12647

Ebo Hinson, R., & Mogaji, E. (2020). Co-creation of value by universities and prospective students: Towards an informed decision-making process. In Higher Education Marketing in Africa: Explorations into Student Choice (pp. 17-46). Cham: Springer International Publishing. https://link.springer.com/chapter/10.1007/978-3-030-39379-3_2

Alwi, S., Che-Ha, N., Nguyen, B., Ghazali, E. M., Mutum, D. M., & Kitchen, P. J. (2020). Projecting university brand image via satisfaction and behavioral response: Perspectives from UK-based Malaysian students. Qualitative Market Research: An International Journal, 23(1), 47-68. https://www.emerald.com/insight/content/doi/10.1108/QMR-12-2017-0191/full/html

Ercantan, O., & Eyupoglu, S. (2022). How do green human resource management practices encourage employees to engage in green behavior? Perceptions of university students as prospective employees. Sustainability, 14(3), 1718. https://www.mdpi.com/2071-1050/14/3/1718

Djuric, V. (2023). Searching for a better segment: segmentation of student body through the use of online search data to improve university enrollment (Doctoral dissertation, University of Liverpool). https://livrepository.liverpool.ac.uk/id/eprint/3177178

Dubey, R., Gunasekaran, A., Childe, S. J., Blome, C., & Papadopoulos, T. (2019). Big data and predictive analytics and manufacturing performance: integrating institutional theory, resource-based view and big data culture. British Journal of Management, 30(2), 341-361. https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-8551.12355

Hatch, J. A. (2023). Doing qualitative research in education settings. State university of New York press. https://books.google.com/books?hl=en&lr=&id=78POEAAAQBAJ&oi=fnd&pg=PT7&dq=This+paradigm+emphasizes+the+need+of+combining+data+from+many+sources+and+utilizing+analytical+tools+to+obtain+a+deeper+understanding+of+student+behaviors,+preferences,+and+performance+indicators.+&ots=-egGHP97sK&sig=B0TRjwa6D5g83ij8byBl9Elys-U

Namoun, A., & Alshanqiti, A. (2020). Predicting student performance using data mining and learning analytics techniques: A systematic literature review. Applied Sciences, 11(1), 237. https://www.mdpi.com/2076-3417/11/1/237

Al-Mamary, Y. H. S., Abdulrab, M., Alwaheeb, M. A., & Alshammari, N. G. M. (2020). Factors impacting entrepreneurial intentions among university students in Saudi Arabia: testing an integrated model of TPB and EO. Education+ Training, 62(7/8), 779-803. https://www.emerald.com/insight/content/doi/10.1108/ET-04-2020-0096/full/html

Fang, H. (2023). Effects of Marketing Mix on Chinese Students' College Choice (Doctoral dissertation, Concordia University Chicago). https://search.proquest.com/openview/859b4a7df0028bdc7215d59a7f953494/1?pq-origsite=gscholar&cbl=18750&diss=y

Goodrich, K., Swani, K., & Munch, J. (2020). How to connect with your best student prospects: saying the right things, to the right students, in the right media. Journal of Marketing Communications, 26(4), 434-453. https://www.tandfonline.com/doi/abs/10.1080/13527266.2018.1514319

Balzer, W. K. (2020). Lean higher education: Increasing the value and performance of university processes. Productivity Press. https://www.taylorfrancis.com/books/mono/10.4324/9781351216944/lean-higher-education-william-balzer

Babu, M. M., Rahman, M., Alam, A., & Dey, B. L. (2024). Exploring big data-driven innovation in the manufacturing sector: evidence from UK firms. Annals of Operations Research, 333(2), 689-716. https://link.springer.com/article/10.1007/s10479-021-04077-1

Hindle, G., Kunc, M., Mortensen, M., Oztekin, A., & Vidgen, R. (2020). Business analytics: Defining the field and identifying a research agenda. European Journal of Operational Research, 281(3), 483-490. https://www.sciencedirect.com/science/article/pii/S0377221719308173

Oqaidi, K., Aouhassi, S., & Mansouri, K. (2022). Towards a students' dropout prediction model in higher education institutions using machine learning algorithms. International Journal of Emerging Technologies in Learning (Ijet), 17(18), 103-117. https://www.learntechlib.org/p/223063/

Luca, M., & Bazerman, M. H. (2021). The power of experiments: Decision making in a data-driven world. Mit Press. https://books.google.com/books?hl=en&lr=&id=XDsZEAAAQBAJ&oi=fnd&pg=PR7&dq=case+studies+conducted+at+institutions+that+have+successfully+implemented+complete+data+analytics+systems+provide+persuasive+evidence+of+the+profound+influence+of+data-driven+decision-

making+on+student+recruitment.+&ots=77W2VQedm7&sig=v0PpYdRwkiA3yLrIjpXFs6AtMco

Stronge, J. H., & Xu, X. (2021). Qualities of effective principals. ASCD. https://books.google.com/books?hl=en&lr=&id=k_kuEAAAQBAJ&oi=fnd&pg=PP1&dq=These+case+studies+illustrate+how+colleges+may+utilize+data+to+obtain+valuable+information+on+student+habits,+interests,+and+requirements.+This,+in+turn,+improves+recruiting+strategies+and+promotes+student+achievement.&ots=hv3eIzvetE&sig=yovtNYYXobUkrHmbYHwSb45qdEY

Studyinfo. (n.d.). Comprehensive details regarding study programs. Retrieved from https://studyinfo.fi.

Kurni, M., Mohammed, M. S., & Srinivasa, K. G. (2023). Predictive Analytics in Education. In A Beginner's Guide to Introduce Artificial Intelligence in Teaching and Learning (pp. 55-81). Cham: Springer International Publishing. https://link.springer.com/chapter/10.1007/978-3-031-32653-0_4

Darius, P. S., Sowjanya, K., Manju, V. N., Saha, S., Mitra, P., Majumder, P., ... & Prabhu, S. M. (2024). From Data to Insights: A Review of Cloud-Based Big Data Tools and Technologies. Big Data Computing, 86-110. https://www.taylorfrancis.com/chapters/edit/10.1201/9781032634050-5/data-insights-preethi-darius-krishna-sowjanya-manju-sanchari-saha-paramita-mitra-pratham-majumder-suneetha-shreekanth-prabhu

Rawat, S., & Narain, A. (2018). Understanding Azure Data Factory: Operationalizing Big Data and Advanced Analytics Solutions. Apress. https://books.google.com/books?hl=en&lr=&id=ZnWADwAAQBAJ&oi=fnd&pg=PR3&dq=Azure+Data+Factory+(ADF)+is+applied,+which+is+a+very+efficient+managed+service+&ots=r1_ajALN6-&sig=6V8icJL2-sOh-0F85GOjAqPER4o