DEVELOPMENT OF A COMPREHENSIVE DASHBOARD FOR ACADEMIC MANAGEMENT IN HIGHER LEARNING INSTITUTION

CHOW WAI KEAN

A project report submitted in partial fulfilment of the requirements for the award of Master in Data Management and Analytics

Lee Kong Chian Faculty of Engineering and Science
Universiti Tunku Abdul Rahman

October 2023

DECLARATION

I hereby declare that this project report is based on my original work except for citations and quotations which have been duly acknowledged. I also declare that it has not been previously and concurrently submitted for any other degree or award at UTAR or other institutions.

Signature :

Name : Chow Wai Kean

ID No. : 2206588

Date : 19/12/2023

APPROVAL FOR SUBMISSION

I certify that this project report entitled **"DEVELOPMENT OF A COMPREHENSIVE DASHBOARD FOR ACADEMIC MANAGEMENT IN HIGHER LEARNING INSTITUTION"** was prepared by **CHOW WAI KEAN** has met the required standard for submission in partial fulfilment of the requirements for the award of Master in Data Management and Analytics at Universiti Tunku Abdul Rahman.

Approved by,

Signature : ______________________________

Supervisor : Dr Fatimah Audah binti Md. Zaki

Date : 21/12/2023

Signature : ______________________________

Co-Supervisor : ______________________________

Date : ______________________________

ACKNOWLEDGEMENTS

I would like to thank everyone who had contributed to the successful completion of this project. I would like to express my gratitude to my project supervisor, Dr. Fatimah for her invaluable advice, guidance and her enormous patience throughout the development of the project.

In addition, I would also like to express my gratitude to my loving parents and friends who had helped and given me encouragement and support in completing this project.

ABSTRACT

As the number of students pursuing higher education continues to grow, decision-making for university stakeholders gets more challenging. This project aims to develop a performance dashboard that allows educators and management to monitor and analyse the academic progress and achievements of students from a diverse background and intake. Targeted users of this dashboard are internal users involved actively in formulating strategic business decisions and interested in analysing the pattern or trend of the program. The research methodology used in this project is a modified Knowledge Discovery in Databases (KDD) to achieve the project objective. The steps include domain understanding, data pre-processing, feature engineering, dashboarding, and evaluation. The project had successfully developed comprehensive dashboard to monitor student performance. System usability scale results had indicated that the dashboard developed is user friendly and users agreed that the dashboard able to provide insight for decision making based on student performance.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS / ABBREVIATIONS

BI	business intelligence
CGPA	cumulative grade point average
GPA	grade point average
HEI	higher education institute
KDD	knowledge discovery in databases
KPI	key performance indicator
SQL	structured query language
SUS	system usability scale

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 General Introduction

Higher Education Institutions (HEI), such as college and university, provide tertiary education through various programme such as foundation, undergraduate and postgraduate programmes. Over the years, the number of students pursuing higher education increase significantly and decision making become challenging for the management of the HEI in the view of department level, faculty level and even the university level.

Business Intelligence (BI) evolves from decision support system that able to support the management of university in making business decisions based on data or information collected from business processes such as admission, examination, teaching and learning, and etc (Moss and Atre, 2003). BI is comprised of organizational and technical element as it needs human understanding with critical analytical skill along with the use of necessary supporting tools to extract meaningful insight for decision making (Olszak, 2014). In BI, data collected and available will be utilised and further analysed to reveal insightful information such as trend and pattern. BI in HEI act as strategic catalyst in strategy development, quality assurance, and other improvement at distinct business processes as well as improvement of teaching and learning.

Goal 4 of the Sustainable Development Goals emphasises the importance of quality education in building a better future. This project focused on developing a student performance dashboard that will allow decision makers to monitor and assess various aspects of student performance and formulate appropriate strategies to improve the overall quality of education. Besides, the development and implementation of a comprehensive dashboard aligns with the core values of Malaysia's latest national agenda, Madani, which emphasises the use of innovation to improve the quality of life in the country.

1.2 Problem Statement

Higher education always strives for continuous improvement to stay competitive in providing quality education. It often faces the need to make strategical decisions by analysing the performance in a timely manner. As the number of students increased over time, the increased in diversity cause difficulties for the educators or management to monitor the performance of the student. Even within a same programme, there can be students who enrolled from diverse qualification, demographic and learning abilities. Due to the lack of a comprehensive dashboard, decision makers are unable to effectively identify segments of diverse students who require additional assistance and fail to implement early intervention to improve education quality. Hence, there are needs of a mechanism to assist in monitoring and improving students learning outcomes. According to Gordanier et al. (2019), effectiveness of early intervention to improve student performance can be different for student with different background. Hence, there are needs to identify the segment of students that are in needs and form the suitable early intervention to effectively help those students with poor academic performance, withdrawal from study, or delayed graduation.

Visuals is easier to recognise and process as compared to words (Dewan and Librarian, 2015). Without visualisation, educators or managers will have limited access to monitor the student performance from a high-level view. Besides, it is time consuming and tedious for educators to repetitively collect data or reporting required and perform analysis from scratch whenever they need to analyse the student performance. Raw textual data available in spreadsheet or tables is harder to be understand by human as compared to graphical representation (Islam and Jin, 2019). It has created challenges for stakeholders to identify patterns, trends, and outliers of the student performance effectively. Without the help of data visualisation, meaningful insights are harder to be made for the formulating actionable recommendations that enhance the students' learning.

Data quality is the root cause that contributes to misleading findings in data-driven analytics (Singh & Dwivedi, 2020). The raw data that is acquired from the database can be dirty and lack of sufficient features that indicate certain performance. These unprocessed raw data are not able to reflect the situation effectively for strategies formulation during decision-making. Without proper data processing and visualisation, the stakeholders will face difficulties in identifying potential trends and insights that could be strategized in the decision-making process. Hence, there is a need to utilise tools that pre-process raw data and further virtualize the data using dashboard technique as an initiative to allow stakeholders to monitor the performance of the student effectively and formulate strategy to help student in needs.

1.3 Aims and Objectives

The project aim is to develop performance dashboard that allow educators and management to monitor and analyse the academic progress as well as achievements of the students from diverse background and intake. The dashboard will integrate data and visualise them in a user-friendly and interactive way. The aim of this project is achieved through the following objectives:

1. To perform data pre-processing on the dataset by August 2023.
2. To perform feature engineering for determining significant features that are related to student performance by October 2023.
3. To develop a dashboard for student performance analysis by December 2023.
4. To evaluate the student performance dashboard with a baseline score of 68 on the System Usability Scale (SUS) by December 2023.

1.4 Scope and Limitation of the Study

The scope of this project focuses on developing a student performance dashboard in the context of Universiti Tunku Abdul Rahman (UTAR) which is a higher education institution. The subject for analysis and data collected for this project is related to UTAR Software Engineering students with intake session starting from January 2018. The data collected covers admission information such as student demographic as well as their respective study status and grade point throughout the study.

The targeted users of this dashboard will be internal users of UTAR such as management committee, administrative head, and lecturer of UTAR that are involved actively in formulating strategic business decisions and interested in analysing the pattern or trend of the program. In this project, measurement of performance to be focused is limited to student study status and graduate on time. Prediction on existing student graduates on time is performed with a suitable model to provide forecast on whether the existing student will graduate on time.

Python and Microsoft Power BI are the tools used to perform data processing, data prediction and data visualisation. The output dashboard is to provide analysis and high-level view of indicators that represent the graduate on time and study status over the observation period.

This project does not consider the behavioural context of the student and the grade point of student is aggregated by semester without detail of individual subject. The performance dashboard does not cover the human resource, research, financial and other administrative aspects of the university.

CHAPTER 2

LITERATURE REVIEW

2.1 Literature Review

2.1.1 Performance Dashboard

Performance dashboard is one of the tools used in BI. Performance dashboard refers to visualisation application that integrated with data and allows organization to effectively determine and monitor the current performance of business in various context (Eckerson, 2010). Dashboard act as a convenient method to visualise data and information that is required by the decision makers for making beneficial decision. It also comprised of numerous indicators that allow decision maker to keep track on essential metrics across different processes within the business. Certain characteristics of performance dashboard includes use of visual components, variant of data source, ability of drilling data, dynamic view, display of key performance indicator, user friendly etc. Performance dashboard had been implemented in different industry and area such as hospital healthcare management, financial management etc.

The success of the performance dashboard much rely on the availability of data as it will be the main input to generate meaningful insight for the top management and dirty data can cause misleading analysis. Hence, data management system or information systems in a HEI must be comprehensive to capture all necessary data and information. Data warehouse is an approach to ensure the time variant data across various sources are stored as normalized and dimensional data to ensure data consistency and prevent data redundancy (Santoso and Yulia, 2017).

2.1.2 Previous work of performance dashboard in HEI

According to previous research conducted by Mutean et al., performance dashboard in HEI or university shall cover some of the essential areas such as student teaching and learning, faculty, finance, research, staff as well as university business processes (Muntean et al., n.d.). These areas can be further visualised by identifying relevant key performance indicators (KPI) that could represent the current performance. KPI also refer to the benchmark in a specific aspect over certain period. For instance, data such as enrolment rate, graduate rate can be used to determine the growth of the faculty, human resources allocation and more.

The typical users of university performance dashboard can be top management such as dean, head of department, manager, and others managerial positions. With performance dashboard, they can perform data-driven decision making without having to find out the status across different department. It also allows them to monitor the progress or achievement of the targeted key performance from time to time.

BINUS university had faced issue of slow access to the information needed by the top management. To overcome that, researcher such as Meyliana et al. had implemented university dashboard in BINUS University to assist the top management of university (Meyliana, Widjaja and Santoso, 2014). From the research, it suggested that KPI and data availability is among the crucial determinant for the success of the dashboard project. The KPI chosen should be revised from time to time as well to ensure it meets the changing needs and pace.

Currently, most of the dashboard proposed are mainly focus on descriptive analytics where visualisation of existing data is performed to review the current status of the organization. However, some researchers such as Nafiisah et al had suggested predictive analytics to be included as future work of the dashboard project (Nafiisah et al., 2021). This view was further supported by Thanh T et al whereby machine learning should be applied to the performance dashboard to enhance predictive analytics and support management in making significant forecasting and decision making (Thanh et al., 2021).

According to Khatibi et al, they had proposed a business intelligence-based model that integrate variety of data sources into their proposed dashboard system, and they utilized the proposed system to conduct and predict meaningful forecast in term of trend of the higher education (Khatibi, Keramati and Shirazi, 2020).

In the research of interactive dashboard conducted by Nafiisah et al (Nafiisah et al., 2021), a four-phase method is adopted. The phases adopted includes analysis, planning, design, and implementation. During analysis phase, relevant approach regarding dashboarding and information system is being studied. Interview had been conducted with the relevant department to understand the needs of the dashboard. Furthermore, existing data is being analysed to understand its context before further processing. In planning phase, relevant metrics and key performance indicators were identified to produce the dashboard prototype. Design phase involves building visualisation that is suitable to represent the demographic and graduate outcomes. Lastly, alumni statistics is being integrated into the dashboard. The tools utilized in this research includes SQL database, Microsoft Power BI, and Python.

According to Denwattana and Saengsai (Denwattana and Saengsai, 2017), architectural design is adopted for the framework of Thailand higher education dashboard. It begins with conceptual design which consists of three bottom-up layer which begin with bottom layer of hardware and software infrastructure, functional processing systems and the top layer of output which are the dashboard and profiles. The conceptual design is then transformed into high-level design which further outlines the specific information systems or resources required for each layer defined in conceptual design. The proposed framework is developed using open-source resources such as MySQL and PHP as a case study in a college of Thailand.

In a Vietnam analytics dashboard research, Thanh et al had proposed a method with five main stages to achieve its objective of supporting decision making through dashboard (Thanh et al., 2021). These five stages represent essential components and activities of the dashboard development which are scores sheet, power BI desktop, load and transform data, design dashboard and lastly publishing and sharing. The first stage involves score sheets that represent the input data such as assessment results which are needed to populate the dashboard. Power BI act as the platform and tools for analytics and visualisation with the use of input data. The input data is being load and transform with the help of Power Query as pre-processing measures before visualisation. During dashboard design, the pre-processed data is then visualised through Power BI. Lastly, the reports and dashboard are being published and shared among the stakeholders.

Table 1 Summary of previous work's methodology and tools

Research	**Methodology**	**Tools**
Nafiisah et al (Nafiisah et al., 2021)	Four-phase method: Analysis, planning, design, and implementation	• SQL database • Microsoft Power BI • Python
Denwattana and Saengsai (Denwattana and Saengsai, 2017)	Architectural design: Conceptual and high-level design	• MySQL • PHP
Thanh et al (Thanh et al., 2021)	Five-stage method: Scores sheet, power BI desktop, load and transform data, design dashboard and lastly publishing and sharing	• Power BI desktop • Microsoft Excel

The previous work conducted had provide a comprehensive methods and workflow for higher education institute to implement dashboard in their respective institute for the purpose of performance monitoring. The tools that adopted in previous work are free and open source which reduce the financial and technical constraint for the implementation.

However, there are certain weakness from the previous work which can be tackle for further enhancement. Feature that chosen for visualisation in previous work are based on literature review as well as relevant experience. The significance of visualised feature can be improved by adopting feature selection technique which can effectively filter highly correlated features based on statistical test. In addition, there are also lack of forecast analysis available in the dashboard of previous works. Forecast analysis can be included by including predictive modelling technique as part of the data pipeline before the visualisation. In this project, feature selection techniques and predictive modelling are adopted with the use of data processing tools such as Python to improve the significance of visualised indicators and provide ability of predictive analytics.

2.1.3 Attributes related to the student performance

Student performance can be affected due to various underlying factors that do not have a significant relationship between each other. According to a systematic review conducted by Abu Saa et al. (2019), it suggested that students' previous grades and performance, learning activity and demographics are among the most identified factors that influence the student performance. Learning activity includes the amount of assessment such as quiz and assignment attempted for the particular course. Besides, students' demographics mainly refers to background such as age, nationality, gender, ethnicity and etc.

The finding of Abu Saa et al. is consistent and supported by the systematic review conducted by Khan and Ghosh (2021) where there are significant number of studies shown quality and background of student can influence their performance. Additionally, Khan and Ghosh review shown that student background and behaviour are the best indicator of student performance before commencement of a course while student background along with student assessment would be the best indicator of student performance after commencement of the course with a higher accuracy.

Educational data mining in early stage of courses assists the lecturer to provide necessary support to the student and better management of the upcoming classes. In the experiment conducted by Chanlekha and Niramitranon (2018), they utilized a combination of non-academic attributes, including demographic information, as well as academic attributes such as current academic performance and past qualifications. These inputs were used to forecast and make predictions of the student performance.

In short, student performance had shown relation to academic and non-academic attributes in various studies. Such attributes such as demographic, result, and graduation information should be taken into consideration when assessing student performance.

2.1.4 Knowledge Discovery in Databases

Knowledge Discovery in Databases (KDD) refers to a process of finding knowledge from the data available. KDD is widely associated with data mining, artificial intelligence, data visualisation and etc. The main objective of KDD is to gain knowledge from raw data that is found in a large database.

KDD is iterative and interactive processes which involves active feedback from user and it is comprised of different stages as follow:

1. Understanding domain: Gather domain knowledge and goal of the application.
2. Determining target dataset: Determine and select the data to be focused for discovery.
3. Data pre-processing: Perform necessary pre-processing tasks such as handling missing values, removing outliers, managing data types, etc.
4. Data transformation: Data reduction steps are conducted to reduce the dimensionality of the data set through feature selection.
5. Data mining: Discovering patterns in data through suitable algorithms and techniques.
6. Interpretation: Analyse and interpret the unveiled pattern or trend which then translated into layman term that can be understandable by all users.
7. Using discovered knowledge: Incorporate the insight gained from the analysis into performance measurement, data reporting or strategy planning.

With appropriate implementation, KDD allows users to gain valuable insights and unveils patterns from large and complex datasets, further enabling data-driven decision-making across various industries. Educational data mining (EDM) is an important part of KDD that analyse educational data with techniques such as machine learning, data mining and etc (Lara et al., 2014). EDM had been adopted in various application such as student grade prediction, visualisation of educational data, student modelling and etc (Romero and Ventura, 2010). Discovered knowledge from KDD can be used to help lecturer or faculty to identify strategy that is necessary to improve student performance in various aspects such as grade, dropout, graduate on time and more. In this project, KDD proposed is different in term of the evaluation phase which focused on evaluating the usability of the performance dashboard instead of the data itself. The change of evaluation phase ensures the dashboard developed is not complex and can be used to extract insight by decision makers easily.

2.1.5 Feature Selection

Feature selection is one of the useful techniques adopted to choose subset of most relevant features in respect to the input data. Curse of dimensionality refers to issue where the amount of data required increases exponentially as the number of data dimension increases (Hastie, Tibshirani and Friedman, 2008). With feature selection, it assists in reducing the dimensionality of the dataset and ease in identifying features that are significant to the target variable or able to provide insight about the dataset.

In python, Scikit-Learn is one of the libraries available that allows users to perform feature selection based on the dataset available as a way to reduce the data dimensionality. Scikit-Learn easy to use and compatible with most of the newly developed frameworks and libraries. The main goal of this library is to enhance accuracy of the predictive model and to improve performance of high-dimensional datasets.

One of the univariate feature selection techniques is known as SelectKBest which will rank the features according to scoring and chosen the best k numbers of features. The selection evaluates the relevance of each feature with the target variable based on statistical tests such as chi-square test. For classification task, Chi-square test is used as measurement of differences between prediction and the actual occurrence of events especially in categorical task. Chi-square test plays the role for test of independence which investigate the relationship between each feature and the target.

2.1.6 Data Visualisation Technique

With the growth of data in terms of volume, velocity and variety, data-driven decision making is now expected in education institutions (Schifter et al., 2014). Data-driven decision making in education allows academic staff to understand their students better in various aspects without having to interview each student individually. However, the dataset can be huge and hard to interpret without proper usage of tools and technology.

In the view of challenges in understanding raw data, data visualisation comes into play by allowing users to transform raw data into graphics and visuals that can be understood easily to conduct analysis. The visual can facilitate them in recognising patterns, gaining meaningful insight and formulating decisions. Organisations that wish to elevate to data-driven organisation are now hiring experts with data visualisation among other data related skills.

Data is intangible and with proper visuals, users can understand in a glance of view. Several techniques can be applied in data visualisation to ensure effective delivery of messages to the end user. Visualisation techniques shown in Figure 1 are among the common visuals that are chosen to be used in the dashboard (Shadare et al., 2016).

- Line graph: To visualise correlation between different variables and the comparison along the changes.
- Bar chart: To visualise categorical data based on the proportion with horizontal or vertical bar representation.
- Scatter plot: To visualise values of two numeric fields and their position along the vertical and horizontal axis.
- Pie chart: To visualise the percentage or part of breakdown of different categories.

Visuals or graphics can aid in representing patterns, discovering trends, and relationships among fields that are available in the data. In common practice, colour can be used in different distinct categories for better clarity. In addition, filters such as checkbox and dropdown menu allowed users to drill-through the data.

According to Stephen (Midway, 2020), one of the principles that aid in improving visualisation effectiveness is to use the right software. Visualisation tools come in different ranges of features that suit different business needs and require different levels of technical skills. Previous works discussed in the previous section had suggested that Power BI is among the most used tools for dashboarding.

Gartner as a well-known IT research and consultancy service provider had conducted research for analytics and business intelligence platforms which can be represented using a magic quadrant. Gartner magic quadrant is a research methodology that positions IT services or products according to their completeness of vision and the ability to execute. In the 2023 Gartner Magic Quadrant for Analytics and BI Platforms, Microsoft was positioned as a Leader with the furthest completeness of vision and highest ability to execute which was followed by Tableau (Kurt & Julian, 2023).

According to Gartner, analytics and business intelligence platforms provide aids and support for business users to perform data analysis with minimum technical skills. In their market research, they had considered twelve critical areas including automated insights, analytics catalogue, data preparation, connectivity of data source, storytelling, visualisation, governance, natural language query, reporting, data science integration, metrics storage and collaboration.

Figure 1: Magic Quadrant for Analytics and Business Intelligence Platforms

Source: Gartner

Figure 1 2023 Gartner Magic Quadrant for Analytics and BI Platforms

As a leader for sixteenth consecutive years, Microsoft Power BI has demonstrated seamless alignment with the Microsoft ecosystem products such as Teams and Azure that allows efficient sharing and collaboration. Besides, Microsoft Power BI also allows automated flows via Power Apps along with AI-powered services in the Power BI premium plan despite offering affordable per user price as compared to other vendors and it also has a free version with limited functionalities.

Tableau which is positioned as the second leader holds unique selling points as compared to Microsoft Power BI such as the community of visual analytics developers that specialised in developing skills of Tableau in performing data visualisation. These communities further facilitate the self-learning process for new joiners that wish to utilise Tableau for data visualisation. In addition, Tableau provides flexibility and interoperability which allow users to incorporate Tableau into their existing on-premises solution which enhances the analytics experience across different stakeholders.

The comparison between Power BI, Excel and Tableau in various aspect is summarised as followed:

Table 2 Comparison between Power BI, Excel and Tableau

Aspect	**Power BI**	**Excel**	**Tableau**
Ease of Use	User-friendly for Microsoft apps user	Widely familiar and user-friendly	Require more time to learn and expertise
Data Connectivity	Support different file types, databases and cloud services	Limited data compatibility and inefficient for large dataset	Extensive data source compatibility
Data Visualisation	Interactive dashboards with essential visual	Basic charting available	Extensive customization and flexibility
Pricing	Free and paid versions with feature-based pricing	Paid and included in Microsoft Office suite	Only paid versions with licensing models

Collaboration and sharing	Able to integrate with Microsoft Teams and SharePoint	Limited sharing and collaboration capabilities	Rely on Tableau Server and Tableau Online for sharing
Community and Support	Huge Microsoft community and online resources	Massive active users' community	Active user community with learning resources
Integration with Analytics	Integration with Azure Machine Learning and Power Query	Support basic analytics and lack of machine learning	Integration with various statistical and machine learning tools

Both Microsoft Power BI and Tableau have their own users' group and have their unique selling point. However, Microsoft Power BI can be a better option for small and medium enterprises to explore data visualisation and analysis as it provides a free version for the public while Tableau doesn't. Besides, Microsoft Power BI as part of the Microsoft ecosystem provides better user experience in terms of extensive use cases and cross application support especially for existing organisations that subscribe to Microsoft 365. In addition, Power BI is easier to learn which is suitable for beginners that wish to perform data analysis with lesser time. While Microsoft Excel is widely used for data reporting and offer some similar functions, Power BI is considered as a better option because it offers more visualisation options, extensive data source compatibility, is more interactive, and allows for easier collaboration.

Power BI is one of the tools that equipped with the latest technology enabler, Artificial Intelligence, which can leverage data analytics and visualisation experience. Natural language query, anomaly detection, key influencers, and other AI features are available in Power BI for various use cases. Natural language query enables users to get quick answers by typing a question. Furthermore, anomalies detection assists users in effectively detecting outliers from their data. Key influencers also effectively identify performance-influencing factors and generate segmentation that groups data into specific performance categories.

In this project, Power BI is chosen as an effective tool for a university to implement performance dashboards due to its cost effectiveness and user-friendliness. The free version allows for exploration without financial commitments, and it can be extended to paid versions easily with existing educational subscriptions. Besides, it is equipped with AI features that allows users to perform significant analysis with minimum learning curve.

2.1.7 Usability Testing

Usability and user friendliness are among the crucial factors to implement a successful dashboard. Usability testing is a type of test that evaluates the product by conducting testing with the representative users. During the test, users will be interacting with the dashboard to execute day-to-day basis tasks and provide their feedback for improvement.

System Usability Scale (SUS) is a quick and dirty evaluation that can be used to evaluate usability of an application (Brooke, 1996). SUS is conducted by asking testers to answer a ten questions questionnaire towards the end of the testing. The SUS questions consisted of 5 positive statements and 5 negative statements where each statement is arranged alternatively and assessed with a scale of 1 to 5 according to the level of agreement. To calculate the score, negative statements will be determined by taking the scale minus one, and positive will be by taking five minus the scale. In the end, the total score is calculated by multiplying the sum of all statements by 22.5.

SUS can be conducted easily and it works well for small sample sizes. According to Nielsen (2000), it is suggested that five is a magic number in usability testing. From Nielsen study, it was found that 85% of the usability problem is discovered within five users' input and the learning rate is lower once the test users exceed five. Hence, SUS is suitable to evaluate the user friendliness of the dashboard and gather feedback to enhance the dashboard.

CHAPTER 3

METHODOLOGY AND WORK PLAN

3.1 Introduction

Research methodology refers to the methods or techniques to be implemented for a study that drives it to achieve its objective. With the research methodology implemented, the project is broken down into more manageable phases that will be integrated at the end for the final deliverables. In this project, Knowledge Discovery in Databases (KDD) is adopted as the research methodology. This methodology is adopted in this project with five main phases which are domain understanding, data pre-processing, feature selection, dashboarding and evaluation.

3.2 KDD Adopted in Project

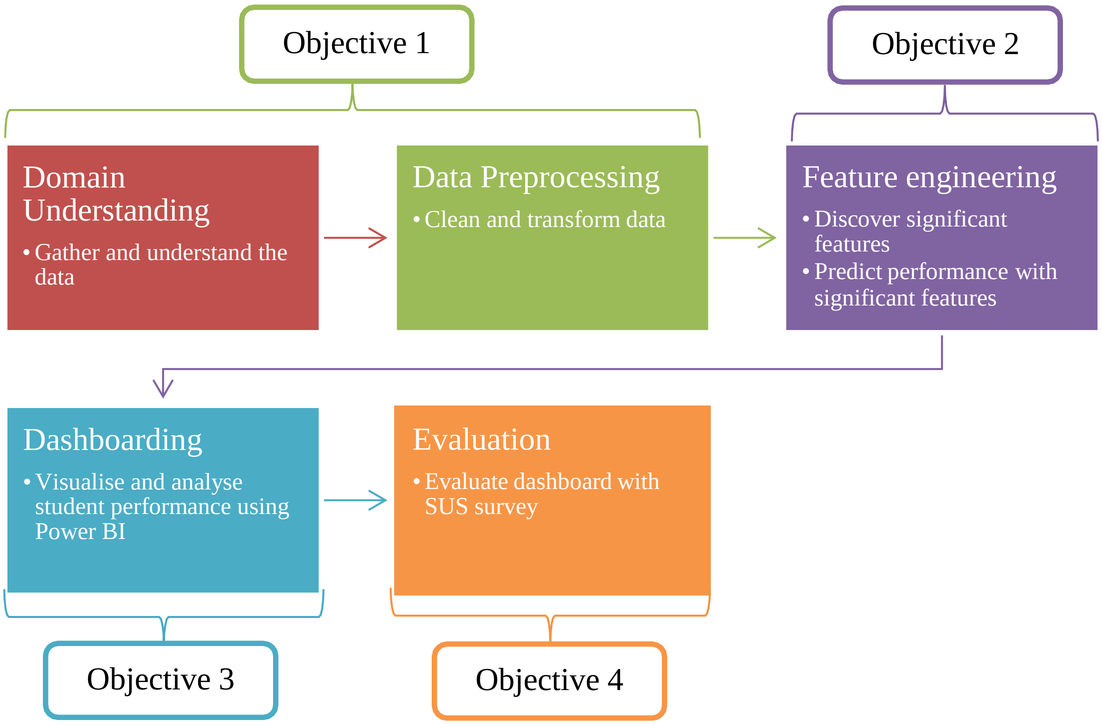

Figure 2 Research Phases

Figure above show the modified KDD phases that is suitable for the execution of this project to achieve each objective specified.

3.2.1 Domain Understanding

At the beginning of project, the project objective and goal are formulated based on problem discovered and title specified. During domain understanding phase, literature reviews of various article and resources are conducted based on relevant topics which includes business intelligence, dashboard design principle, higher education key performance indicator, data management and analytics tools and etc. Review on these areas is necessary to provide meaningful input for designing the dashboard solution that able to facilitate the decision-making of higher education institute. A list of user requirements is outlined after discussion with academic staff that are actively involved in decision making and interested to monitor student performance.

Table 3 User Requirements

No	Requirements
1	The dashboard shall allow user to monitor the distribution of student in term of demographic.
2	The dashboard shall allow user to monitor the academic performance of student.
3	The dashboard shall allow user to monitor the distribution of student in term of graduate on time status.
4	The dashboard shall allow user to monitor the distribution of student in term of study status.
5	The dashboard shall allow user to filter the visual chart based on different intake, study status and qualification.
6	The dashboard shall provide user with forecast trend based on graduate on time status for students.
7	The dashboard shall be user friendly and easy to understand.

In this project, student data are requested from the faculty and provided with the assist of the university’s IT team. The dataset requested is based on UTAR Software Engineering students that starts from 2018-2023 intake. The data provided is in excel format that consists of 9782 rows with 20 columns. Due to privacy concern, the data provided is encoded without revealing the actual identity of the student. In addition, glossary is provided as reference for coded column such as country, study status, English test, qualification and etc.

Table 4 Data Dictionary

No.	Column Description	Column Name	Data Type
1	Encoded Student ID	Student code	Numeric
2	Study Level	Level	Categorical
3	Study Course	Course	Categorical
4	Intake Session	Sess. Joined	Categorical: ['201801', '201805', '201810', '201901', '201905', '201910', '202001', '202005', '202010', '202101', '202105', '202110', '202201', '202206', '202210', '202301', '202305']
5	Entry qualification for bachelor	Quali.	Categorical: ['FS', 'AL', 'KT', 'UE', 'ST', 'LL', 'FA']
6	Previous stream in foundation	Foundation Stream	Categorical: ['S', 'NA', 'A']
7	First choice of course in foundation	Foundation Course	Categorical: ['SE', 'NA', 'MH', 'LI', 'AS', '3E', 'ME', 'CS', 'CN', 'PH', 'EE', 'AM', 'ET', 'IB', 'CI', 'CL', 'EV', 'QS', 'MK', 'CT', 'GV', 'MS', 'PY', 'BE', 'BI', 'GD', 'AR', 'EC', 'FM', 'CD', 'BA', 'BM', 'PS', 'AT', 'EN', 'IA', 'BP', 'DA', 'FN', 'GS', 'BF', 'BT', 'IN', 'AC', 'DE', 'RK', 'DT']
8	Student Age	Age	Numeric
9	Student Gender	Gender	Binary: ['F', 'M']
10	Nationality	Country	Categorical: ['MY', 'SY', 'CN', 'BN', 'PK', 'NP', 'ID', 'BD', 'NA', 'IR']
11	Student Race	Race	Categorical: ['C', 'O', 'M', 'I']

12	English test type attended	Eng. Test	Categorical: ['MU', 'OL', 'NA', 'IE', 'TI', 'OT', 'TO', 'CE']
13	English test result obtained	Eng. Result	Categorical: ['3', '4', 'B', '5', '6', 'NA', '7', '2', '1', 'C', '4.5', 'A', 'D', '9', '3.5', 'A+', '65', '5.5', 'A-', 'B+', 'C+', 'E', '8', 'PS', '155', '6.5', '3B', '0', 'B1']
14	Study Status	Status	Categorical: ['Graduate', 'Withdraw', 'Active', 'Deferred']
15	Current year of study	Yrs. Std	Numeric: 1-3
16	Current semester of study	Sem	Numeric: 1-3
17	Current Grade Point Average	GPA	Numeric: 0 – 4.0
18	Current Cumulative Grade Point Average	CGPA	Numeric: 0 – 4.0
19	Current credit hours earned	Total Credit hours earned	Numeric: 0 - 126
20	Graduation session	Grad. Session	Categorical: ['202010', '202101', '202210', '202110', 'NA', '202201', '202105', '202206', '202301']

3.2.2 Data Pre-processing

Pandas is an open-source library that allowed time consuming and repetitive task such as data cleaning, exploratory data analysis, data transformation and etc to be done efficiently. In data pre-processing phase, the excel dataset is loaded into DataFrame using python pandas library with Jupyter Notebook.

After data being loaded, drop_duplicates function is used to remove any duplicating rows. Insignificant column such as Level and Course are dropped as both of this column contains values that is same for every record. As the data collected is not encoded properly, student code is replaced with incremental numeric index to ensure each student records are consistent.

Validation on null values is performed to ensure the data accuracy and consistency before visualisation. By using drop function, rows with null values on Sem and Yrs. Std are dropped as each record should belong to a specific trimester. Pandas astype function is used to convert all dataframe column into suitable data type for further processing and aggregation.

To ensure the data can be fitted into feature selection model and predictive models, Pandas factorize is used to encode column with multiple labels into categorical variables that represented by index. As each student is represented by multiple rows, group by function is used to aggregate each student code to be represented by a single record.

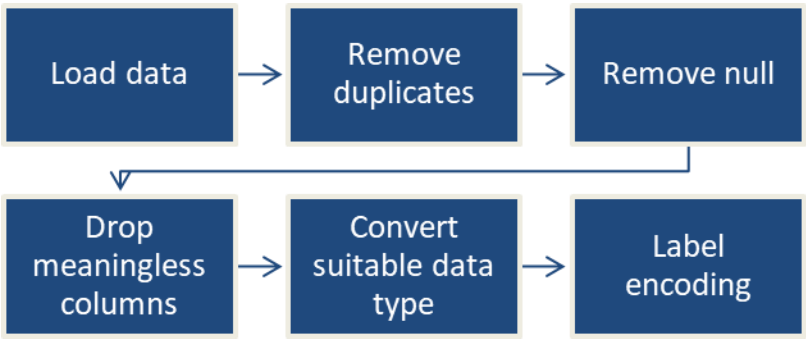

Figure 3 Data Pre-processing Steps

```
{'Sess. Joined': Index(['201801', '201805', '201810', '201901', '201905', '201910', '202001',
       '202005', '202010', '202101', '202105', '202110', '202201', '202206',
       '202210', '202301', '202305'],
      dtype='object'),
 'Quali.': Index(['FS', 'AL', 'KT', 'UE', 'ST', 'LL', 'FA'], dtype='object'),
 'Foundation Stream': Index(['S', 'NA', 'A'], dtype='object'),
 'Foundation Course': Index(['SE', 'NA', 'MH', 'LI', 'AS', '3E', 'ME', 'CS', 'CN', 'PH', 'EE', 'AM',
       'ET', 'IB', 'CI', 'CL', 'EV', 'QS', 'MK', 'CT', 'GV', 'MS', 'PY', 'BE',
       'BI', 'GD', 'AR', 'EC', 'FM', 'CD', 'BA', 'BM', 'PS', 'AT', 'EN', 'IA',
       'BP', 'DA', 'FN', 'GS', 'BF', 'BT', 'IN', 'AC', 'DE', 'RK', 'DT'],
      dtype='object'),
 'Gender': Index(['F', 'M'], dtype='object'),
 'Country': Index(['MY', 'SY', 'CN', 'BN', 'PK', 'NP', 'ID', 'BD', 'NA', 'IR'], dtype='object'),
 'Race': Index(['C', 'O', 'M', 'I'], dtype='object'),
 'Eng. Test': Index(['MU', 'OL', 'NA', 'IE', 'TI', 'OT', 'TO', 'CE'], dtype='object'),
 'Eng. Result': Index(['3', '4', 'B', '5', '6', 'NA', '7', '2', '1', 'C', '4.5', 'A', 'D', '9',
       '3.5', 'A+', '65', '5.5', 'A-', 'B+', 'C+', 'E', '8', 'PS', '155',
       '6.5', '3B', '0', 'B1'],
      dtype='object'),
 'Status': Index(['GR', 'WT', 'EN', 'WN', 'WF', 'WJ', 'WC', 'WY', 'EC', 'WP', 'DF', 'WD',
       'WM'],
      dtype='object'),
 'Grad. Session': Index(['202010', '202101', '202210', '202110', 'NA', '202201', '202105',
       '202206', '202301'],
      dtype='object'),
 'Degree Class': Index(['Distinction', 'Degree', 'Merit', 'Fail'], dtype='object')}
```

Figure 4 Index Reference for Categorical Data

3.2.3 Feature Engineering

Feature engineering involves processes of choosing, transforming and extracting important features from the raw data. The existing features available are lacking the ability to indicate certain performance such as whether the student is graduating on time. Besides, existing features available are insufficient to develop the dashboard that fulfil the user requirements. Hence, new columns are created to measure duration of study based on intake session and their respective graduate session available in the dataset. Student with less or equal to 3 years of study is then labelled as graduating on time in an additional column. To further categorise the student according to their academic performance, additional column is created to store whether the student fail for any course and also the honours classification category based on the cut off CGPA specified.

Table 5 New Features

No.	Column Description	Column Name	Data Type
1	Duration of study before graduate	Study Time	Numeric
2	Graduate on time	GOT	Binary: 0 (Not graduate on time), 1 (Graduate on time)

3	Graduate honours classification	Degree Class	Categorical: ['Distinction', 'Degree', 'Merit'] 3.67 - 4.0 = Distinction, 3.0 - 3.6699=Merit 2.0 - 2.99=Degree
4	Fail any course(s)	Fail	Binary: 0 (None), 1 (Had)

Feature selection can effectively reduce the complexity and dimension of the dataset by choosing only the significate features that is related to the specific target. With pandas corr function, the pairwise correlation between each column in the dataset is obtained and plotted as a heatmap with seaborn library. The correlation for GOT as target is extracted to reveal highly correlated features which represented the higher correlation coefficient.

Table 6 Pairwise Correlation Coefficient with GOT

Column	**Correlation Coefficient**
GOT	1.000000
Study Time	0.891488
Fail	0.460116
Sem	0.407225
CGPA	0.354928
Age	0.352821
Sess. Joined	0.315544
GPA	0.304550
Degree Class	0.222894
Gender	0.150417
Race	0.099596
Total Credit hours earned	0.093122
Foundation Course	0.019958
Country	0.018406
Grad. Session	0.014964
Eng. Result	0.009106
Foundation Stream	0.004773

Eng. Test	0.003094
Quali.	0.002993

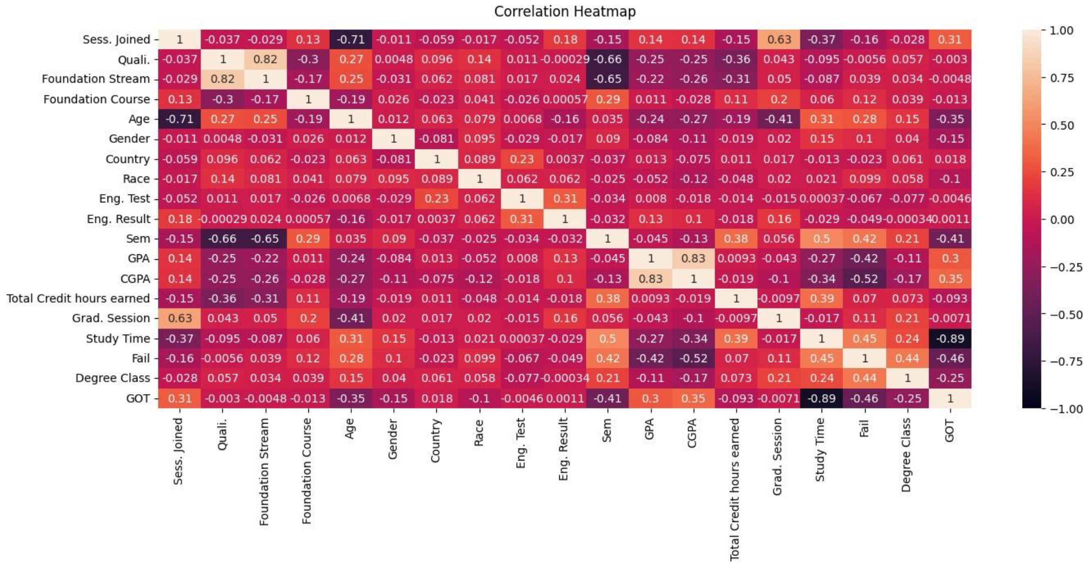

Figure 5 Correlation heatmap for GOT

SelectKBest is used to determine the top 5 features that related to GOT by considering the Chi-square statistics. With k=5, 'Age', 'Race', 'Sem', 'Fail', 'Degree Class' are selected as the top 5 features that are related to GOT. Seaborn pair plot is used to create useful visualisation that plot pairwise relationships between features within a dataset. It also allows the distribution of feature to be visualised easily with different selected features as hue. The highly correlated features identified also function as a reference to be visualised in the dashboard.

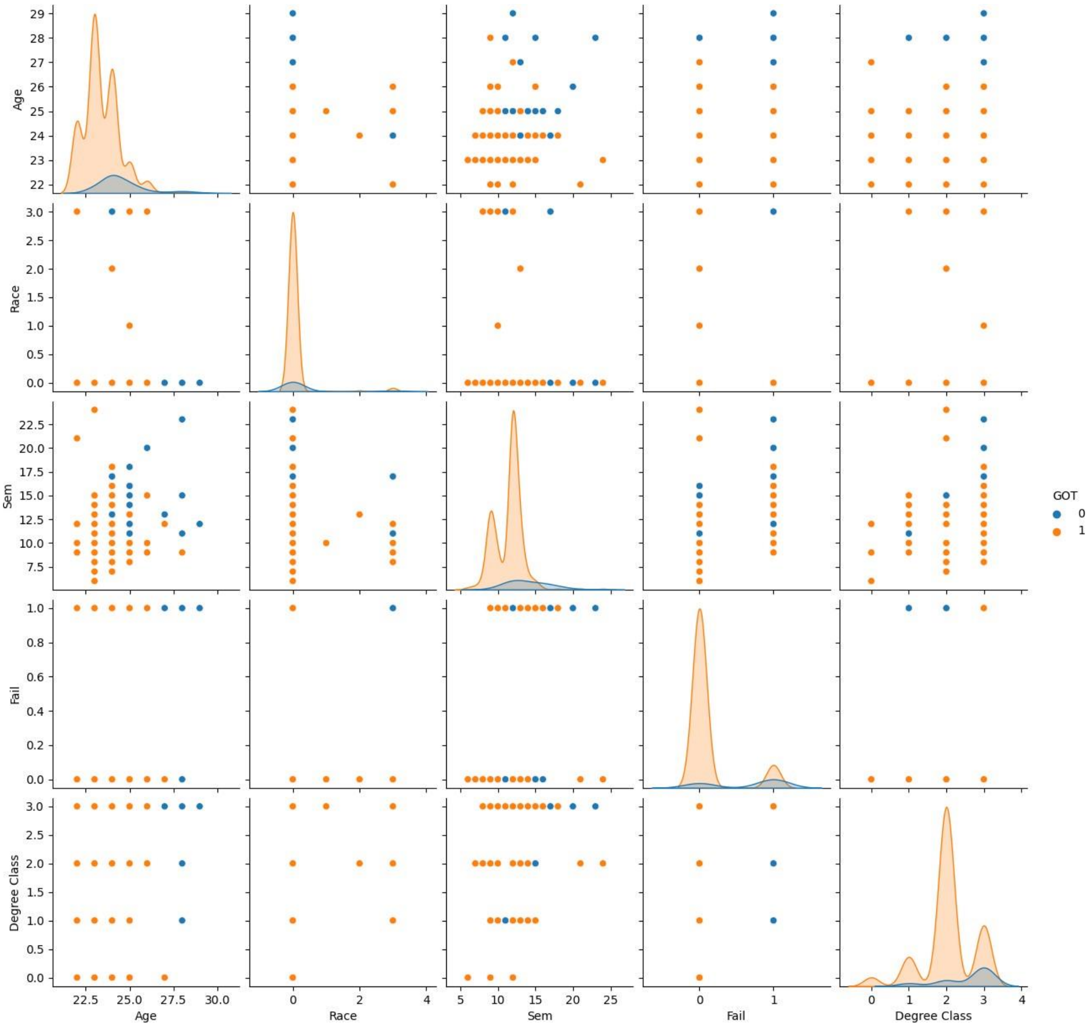

Figure 6 Pair plot with GOT as hue

After feature selection, a k-nearest neighbour, logistic regression and decision tree model is trained to predict whether an existing student will graduate on time based on their existing performance and background. The predictive models are trained using data with the significant features selected based on the result from SelectKBest. Performance of each trained model is then evaluated using confusion matrix and classification report in term of accuracy, precision, recall and f1-score as followed.

Table 7 Performance of Logistic Regression

	Precision	**Recall**	**F1-score**	**Support**
0	0.46	0.72	0.56	25
1	0.92	0.8	0.86	106
Accuracy			**0.79**	131

Table 8 Performance of K-Nearest Neighbours

	Precision	**Recall**	**F1-score**	**Support**
0	0.51	0.72	0.60	25
1	0.93	0.84	0.88	106
Accuracy			**0.82**	131

Table 9 Performance of Decision Tree

	Precision	**Recall**	**F1-score**	**Support**
0	0.53	0.72	0.61	25
1	0.93	0.85	0.89	106
Accuracy			**0.82**	131

Upon comparison, decision tree shown the best performance as compared to logistic regression and k-nearest neighbours. Hence, decision tree model is used for prediction of graduate on time and the predicted results on existing student are exported into a csv file for purpose of visualisation in the dashboard.

3.2.4 Dashboarding

In this phase, the pre-processes data exported from the python are used to visualise the student performance using Power BI. Firstly, the data is being loaded into Power BI environment using its Get Data function that support different file format and connection. In this project, excel and csv are the type of file format that need to be loaded.

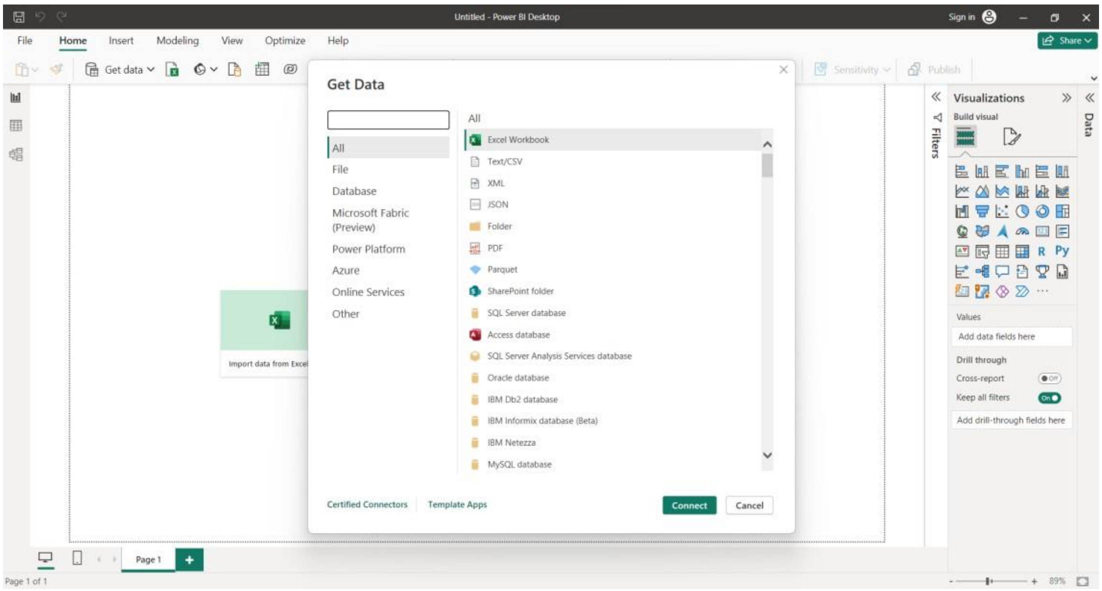

Figure 7 Loading data using Power BI

In the next step, some data transformation such as changing data type and renaming columns are performed using the query windows to ease the visualisation in the later step to provide meaningful visuals effectively. The changes are applied upon clicking the Close & Apply to save all the steps made to the query.

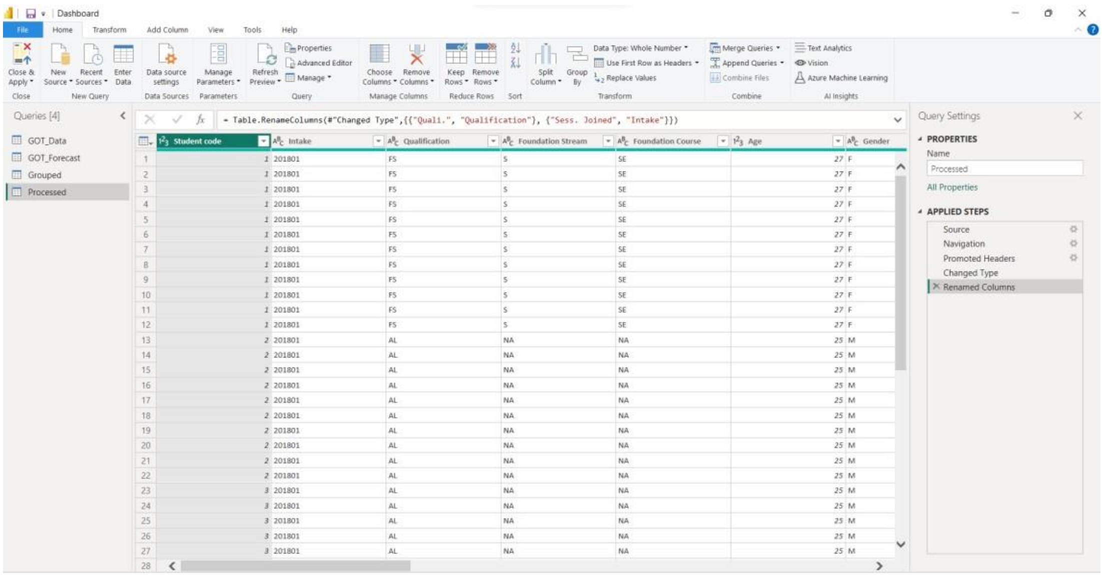

Figure 8 Transforming data using Power BI

After data are successfully loaded, the data is being visualised through various graphical representation such as pie chart, filled map, line chart, bar chart, scatter chart, tree map, funnel chart and more. Each graphical element is customized by using suitable data, legend, axis, and etc.

There are four 16:9 ratio dashboards developed using Power BI where each of this dashboard serves the analysis of different scope and prospective. The first dashboard provides an overall distribution of the students' that are enrolled into the programme according to their study status, gender, median of CGPA, performance, nationality, intake and qualification.

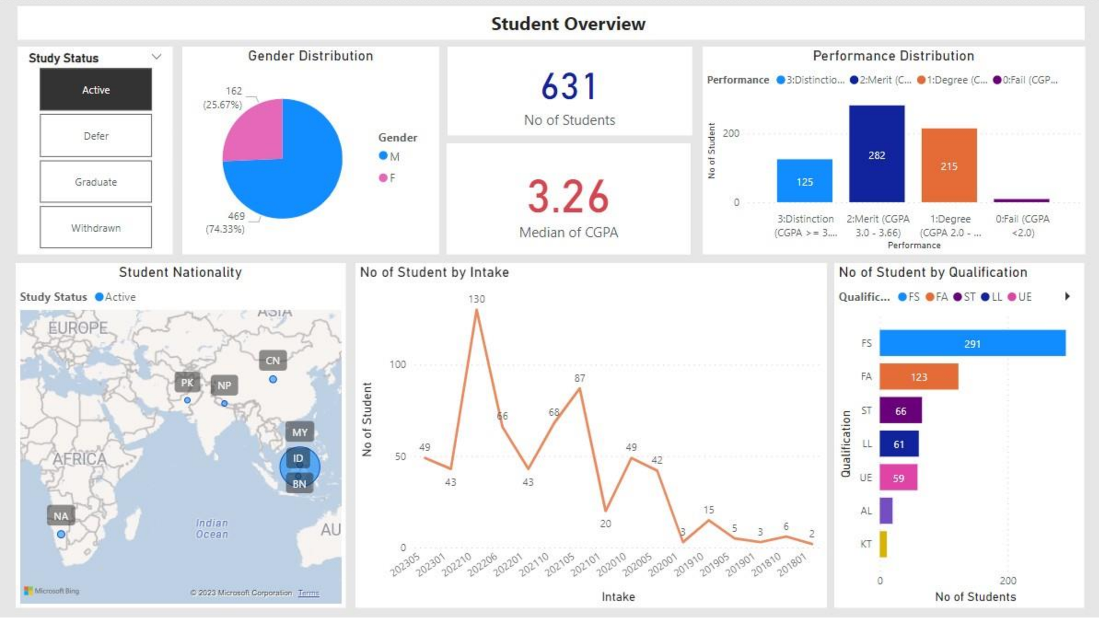

Figure 9 First dashboard – Student Overview

The second dashboard focus in visualizing the student performance in term of median CGPA by different intake and qualification. A student list is shown to display the student information upon drilling down in the graph. Scatter chart is used to visualised CGPA and credit hours earned with qualification as legend to effectively identify outliers or student with poor performance. It allows the user to filter the dashboard by qualification and study status.

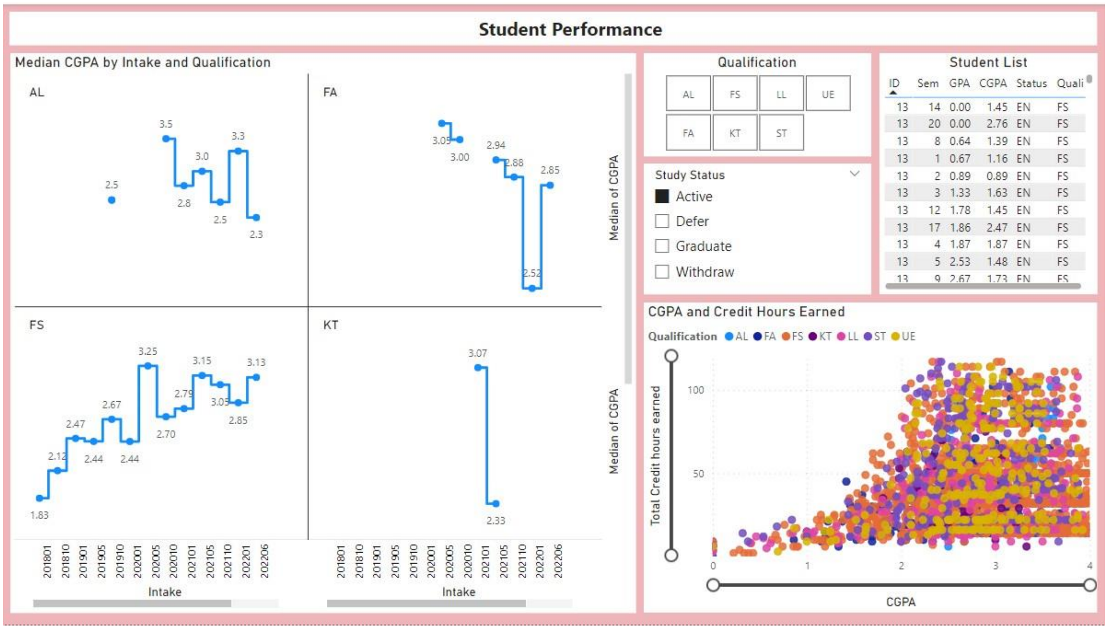

Figure 10 Second dashboard – Student Performance

Figure 7

The third dashboard provides graphical representation on graduate on time in term of number of students by race, number of students by intake, number of students by degree class, number of students by qualification, forecast for existing student, student list of forecasts and the key influencers. It allows the user to filter the dashboard by whether graduating on time.

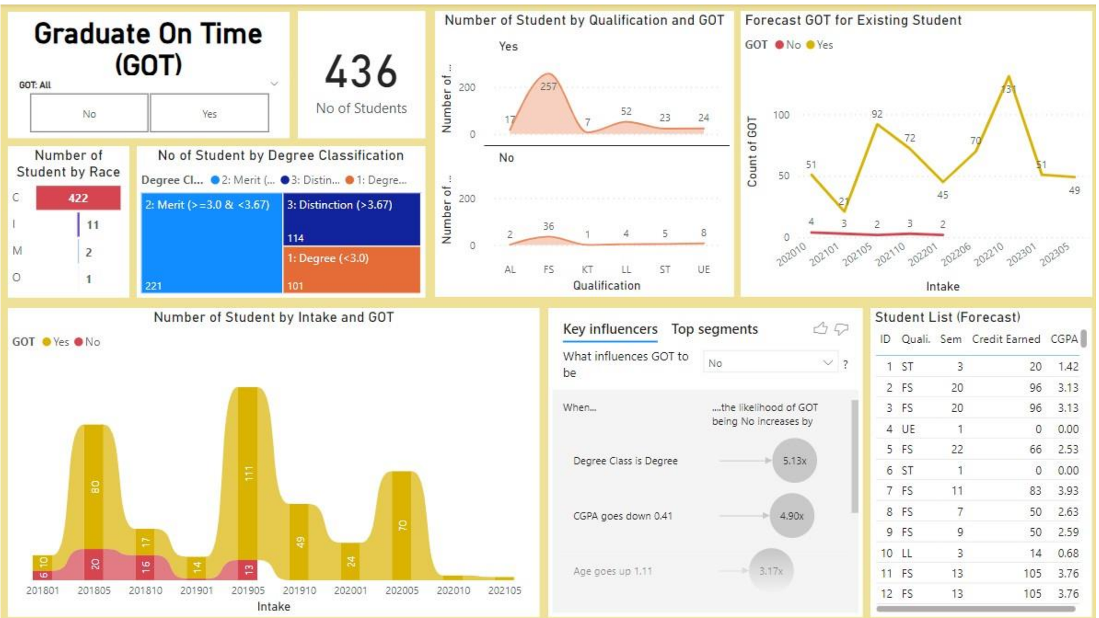

Figure 11 Third dashboard – Student Graduate on Time

The last dashboard provides information on the study status in term of student count by intake and study status, key influencers, segments, and the student list that represent student information. It allows the user to filter the dashboard by intake and study status.

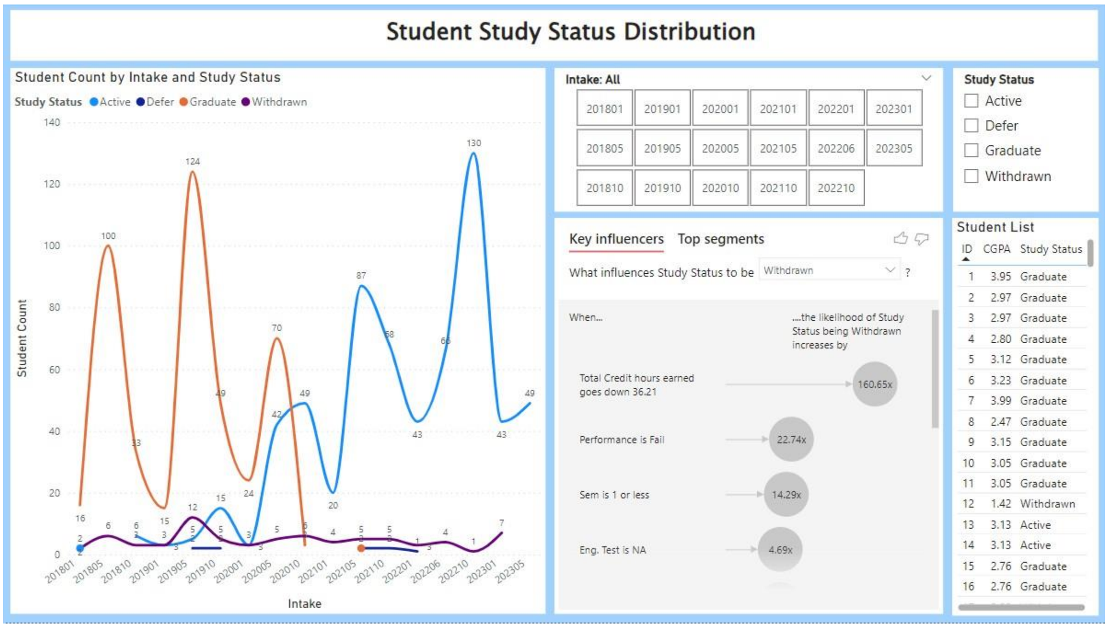

Figure 12 Fourth dashboard – Student Study Status Distribution

3.2.5 Evaluation

System Usability Scale is used to measures the user friendliness and usability of the dashboard developed. Multiple users from the faculty including head of department, head of programmes and lecturers were participated in the testing. Those users are active decision maker that interested in analysing student performance.

During the evaluation, the dashboard is shown to the users and allows them to explore on their own. A google form that outlined 10 questions that referenced from the SUS questionnaires is created and distributed to the users after they had tested the dashboard.

3.3 Work Plan

The project is started in June 2023 and completed by December 2023. Various phases of the research methodology were carried out as planned throughout the project.

In June 2023, domain knowledge required to develop the performance dashboard are gathered and analysed through literature review. Data needed to populate the dashboard are then collected from the higher education institution. Data preparation and wrangling tasks are performed on the collected data including data cleaning, data transformation, data reduction and data validation before the data is ready to be used for visualisation. Feature engineering is performed to reveal the significant factors that relate to the student performance.

The design and development of the dashboard are done between August to October 2023. During this phase, the dashboard development is performed to visualise the features selected that are correlated with the student performance. The performance dashboard is revised after gathering feedback and evaluated using System Usability Scale before the end of this project in December 2023.

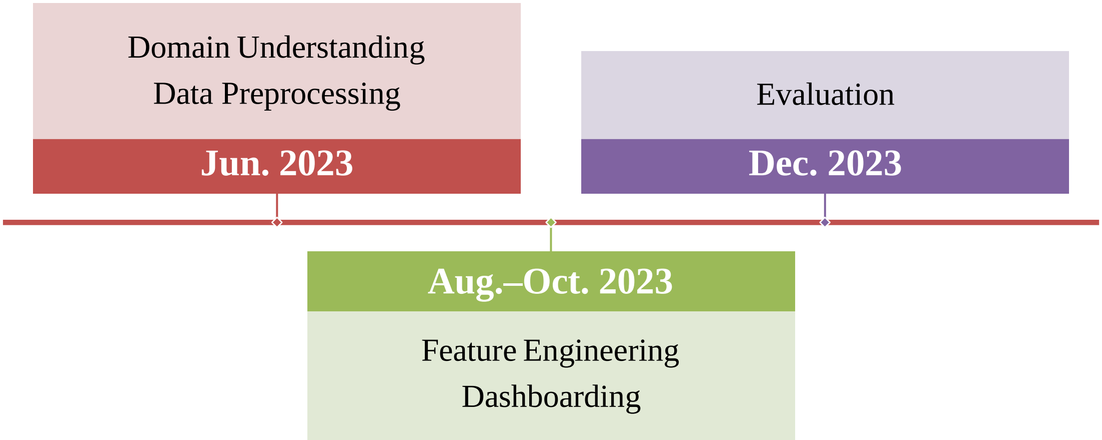

Figure 13 Project Timeline

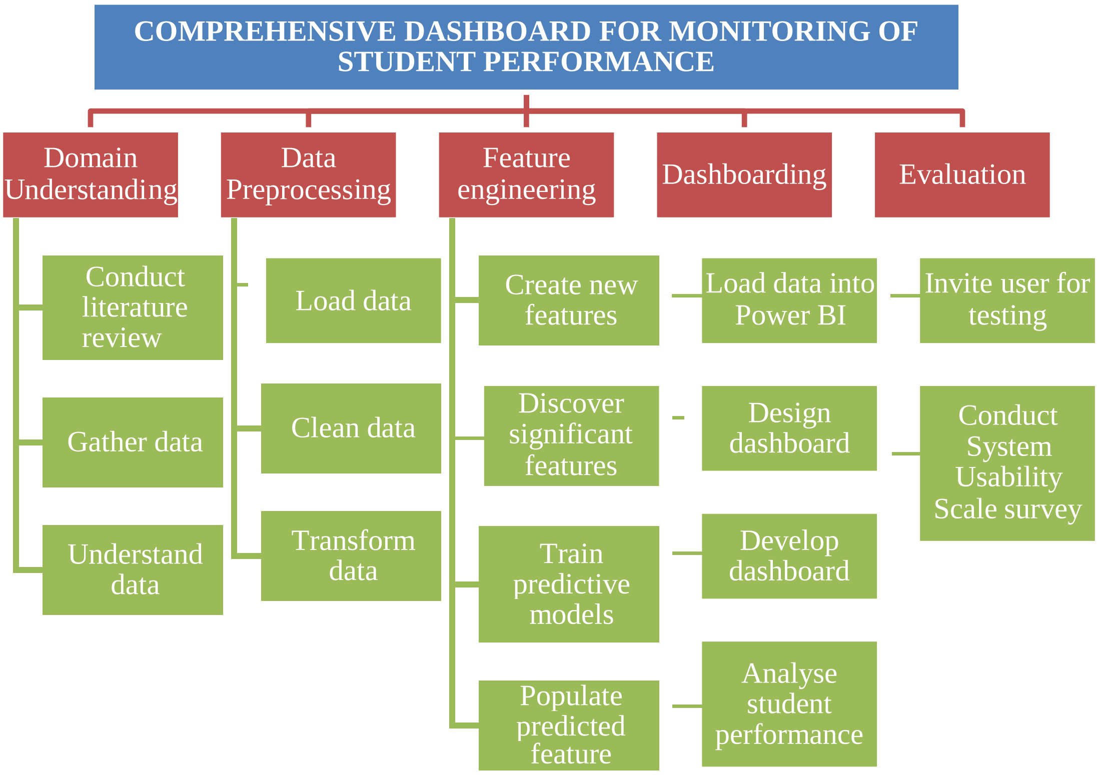

Figure 14 Work Breakdown Structure

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Dashboard Analysis

Dashboards developed with Power BI are utilized for analysing the trend of student performance in different aspect as support in decision making. The output dashboard provides insight based on student performance and allows decision maker to formulate beneficial strategies.

4.1.1 First Dashboard – Student Overview

As an overall, there are total of 1150 students' records being summarised into the overview dashboards which represent all study status which are active, deferred, graduated, and withdrawn student since 2018 intake. According to the pie chart, there are more male students (75.4%) than female students (24.6%). Male students remain as the majority group for all students with different study status. However, the number of new female students has climbed year over year, while the number of new male students has declined in the most recent intake.

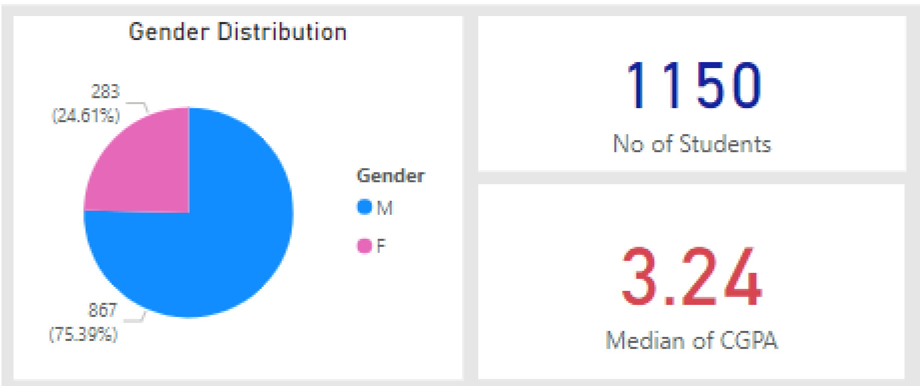

Figure 15 Pie chart and data cards

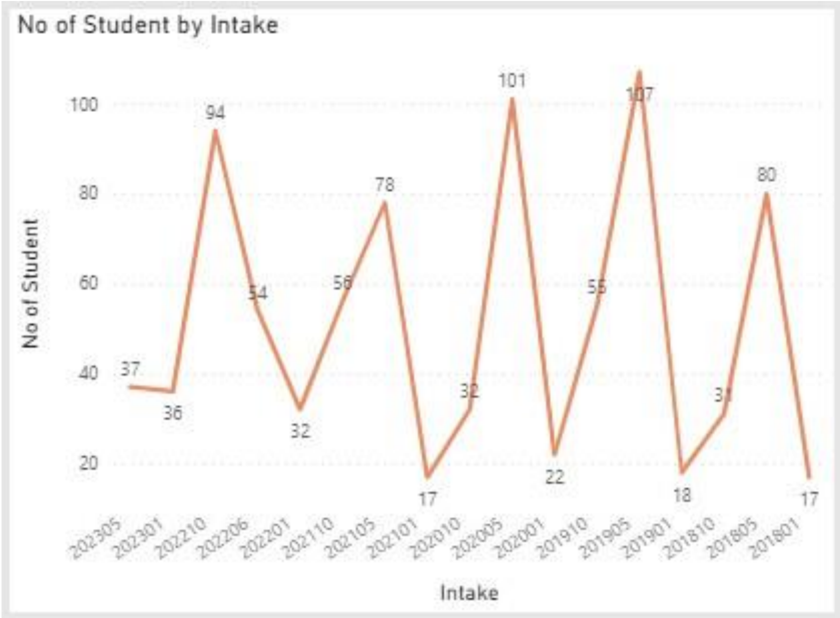

Figure 16 Male new student over the year

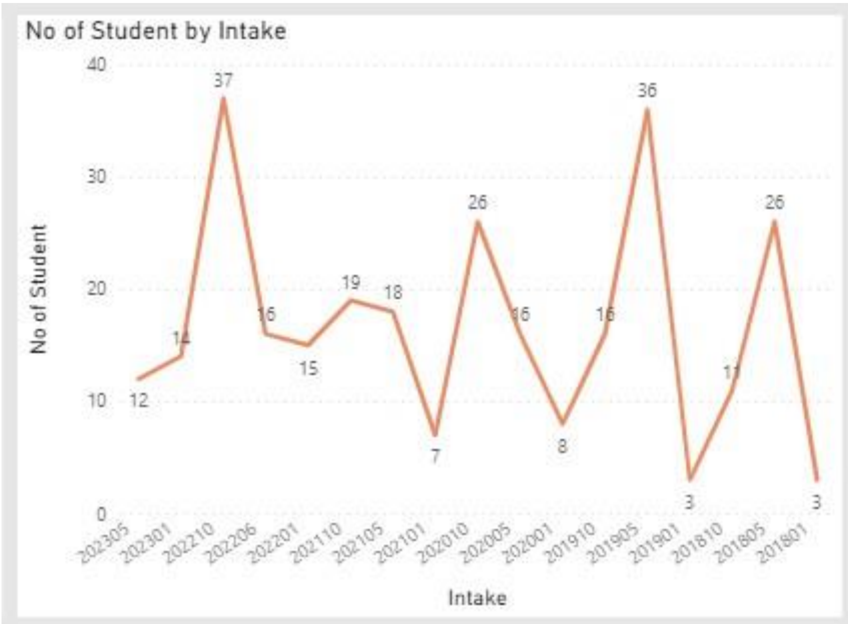

Figure 17 Female new student over the year

In terms of academic achievement, the median CGPA computed for 631 active students is 3.26, where more than 50% of students achieving a CGPA of at least 3.26. From gender perspective, female students had a higher median CGPA of 3.36 than male students, who had a CGPA of 3.19. Majority of students have CGPAs ranging from 3.0000 to 3.6599, followed by a regular degree ranging from 2.0000 to 2.9999. The next highest performance category is distinction, which includes students with a CGPA more than or equal to 3.6700. There are 9 students that are failing which their CGPA is less than 2.0000.

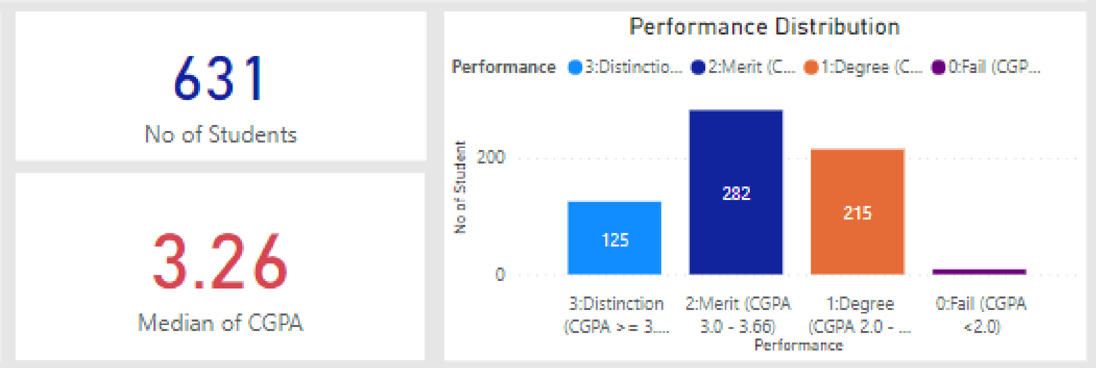

Figure 18 Data cards and bar chart

The nationality of students is clearly visualised via a filled map based on their geographical location. Throughout the year, the majority of students are Malaysians, with international students from China, Indonesia, Brunei, Nepal, Pakistan, Namibia, and other countries. In term of international student, China had the highest number of admissions but also the highest number of withdrawals from study among all countries. The alumni of this programme are all from Malaysia and only one international alumna from Syrian Arab Republic.

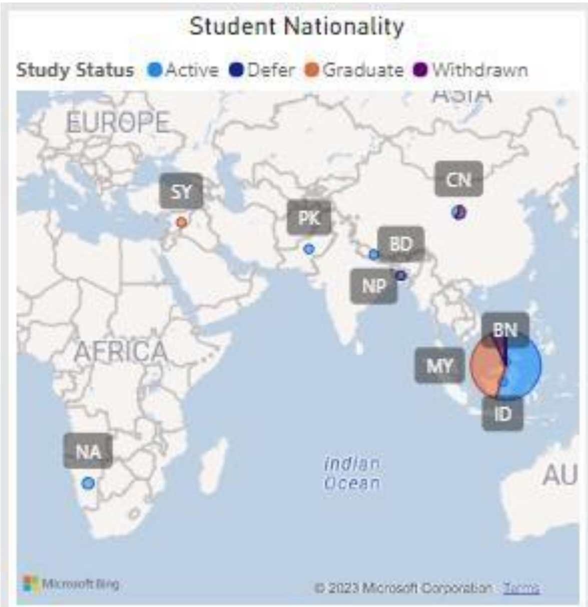

Figure 19 Filled Map for Student Nationality

A line graph is used to show the trend of new students for each intake. From 2018 to 2021, the May intake always had the greatest number of new students in a year, followed by the January and October intakes. Over this four years, May 2021 intake recorded the lowest number of new students as compared to May of 2018 to 2020 which suggested the potential influence of Covid-19 pandemic. This trend, however, did not continue in 2022, where October intake recorded the highest number of new students in that year. This could be due to changes of the Sijil Pelajaran Malaysia (SPM) examination timeline for Malaysian students, which has shifted from the end of the year to the beginning of the new year and the result only released after June.

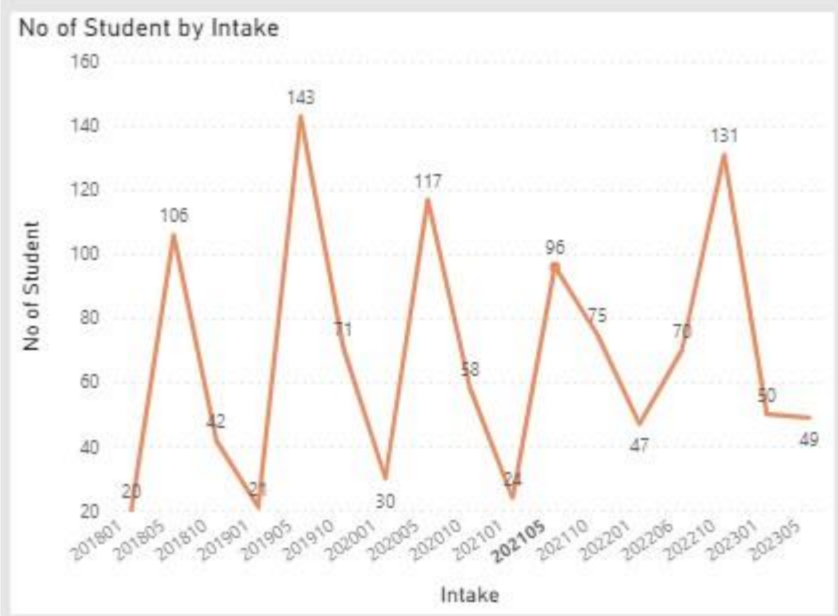

Figure 20 New students for each intake

In term of entry qualification, most students enter with a foundation in science, followed by a foundation in arts, Malaysian Higher School Certificate (STPM), Unified Examination Certificate (UEC), A-Level, and other equivalent qualifications. While the number of new students for other qualifications continues to fluctuate, the number of new students from foundation in arts has shown an increasing trend over the year after the university accepted it as one of the entry requirements.

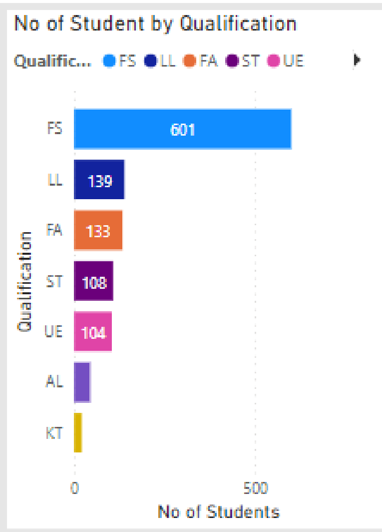

Figure 21 No of students for each qualification

When a noticeable trend is identified on the dashboard, stakeholders can further investigate the situation and develop a strategy to improve performance. For example, stakeholders may work with their respective departments to promote the programme to females in order to balance the gender ratio. Furthermore, stakeholders may assist and support international students, particularly those from China, as an initiative to improve their performance and reduce their withdrawal from study rate.

4.1.2 Second Dashboard – Student Performance

The median CGPA is the value obtained by half of the students. Among active students, students with external qualifications from the 2022 May intake had the highest median CGPA of 3.6. Excluding foundation in science and arts, students from the 2022 October intake with A-Level, TARUMT, STPM, UEC, and other qualifications had the lowest median CGPA when compared to other intakes with the same qualification. Overall, the median CGPA of all intakes' students from foundation in arts ranges from 2.5 to 3.1, while those from foundation in science range from 2.4 to 3.2.

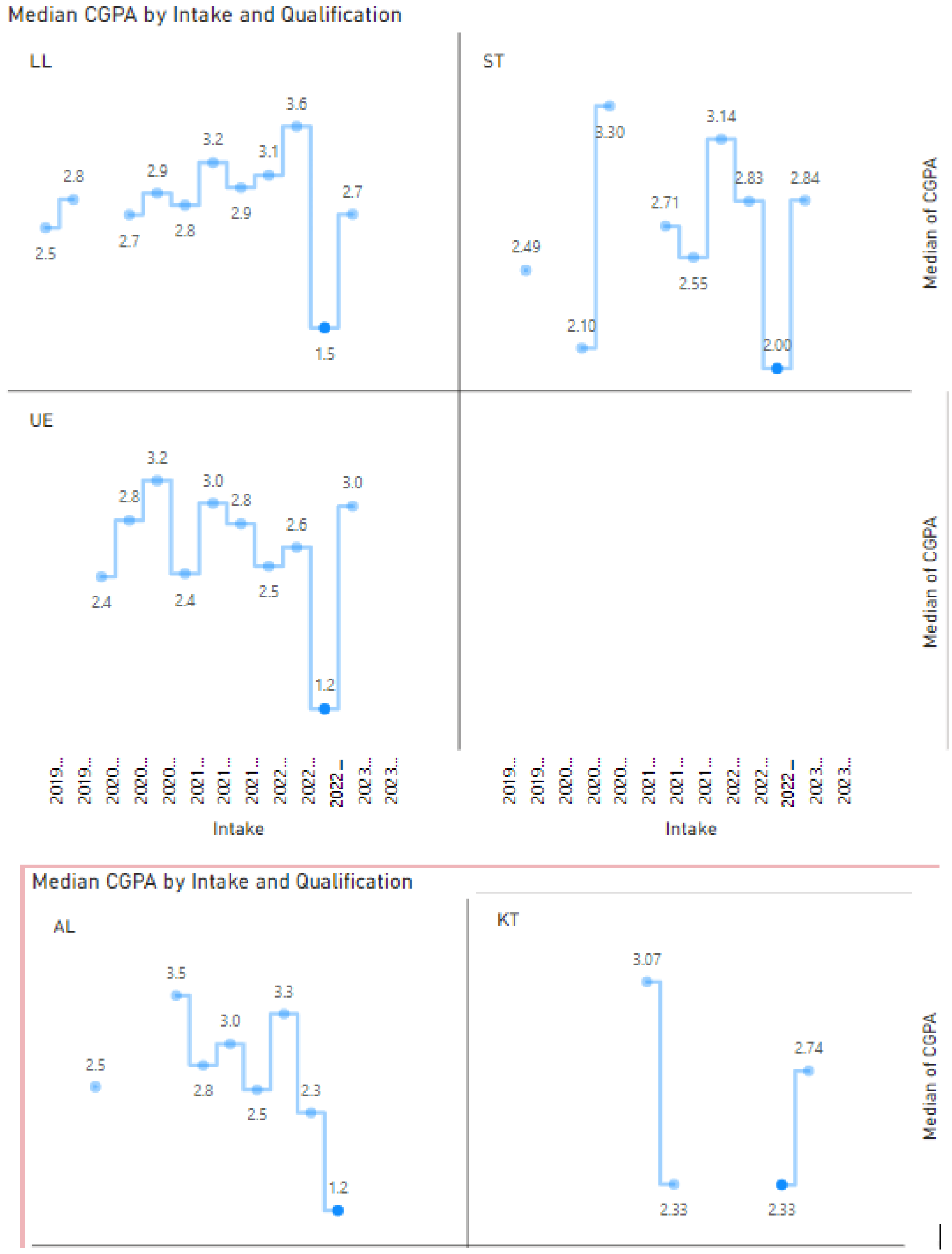

Figure 22 Median CGPA by Qualification and Intake

In terms of graduated students, the May intake of foundation in science students had a higher median CGPA than the other intakes within the same year. UEC students had a lower median CGPA than other qualifications, with the lowest being 2.48. When drill into the semester dimension, it was discovered that the median CGPA for all students increased in the first two semesters and fluctuated slightly thereafter.

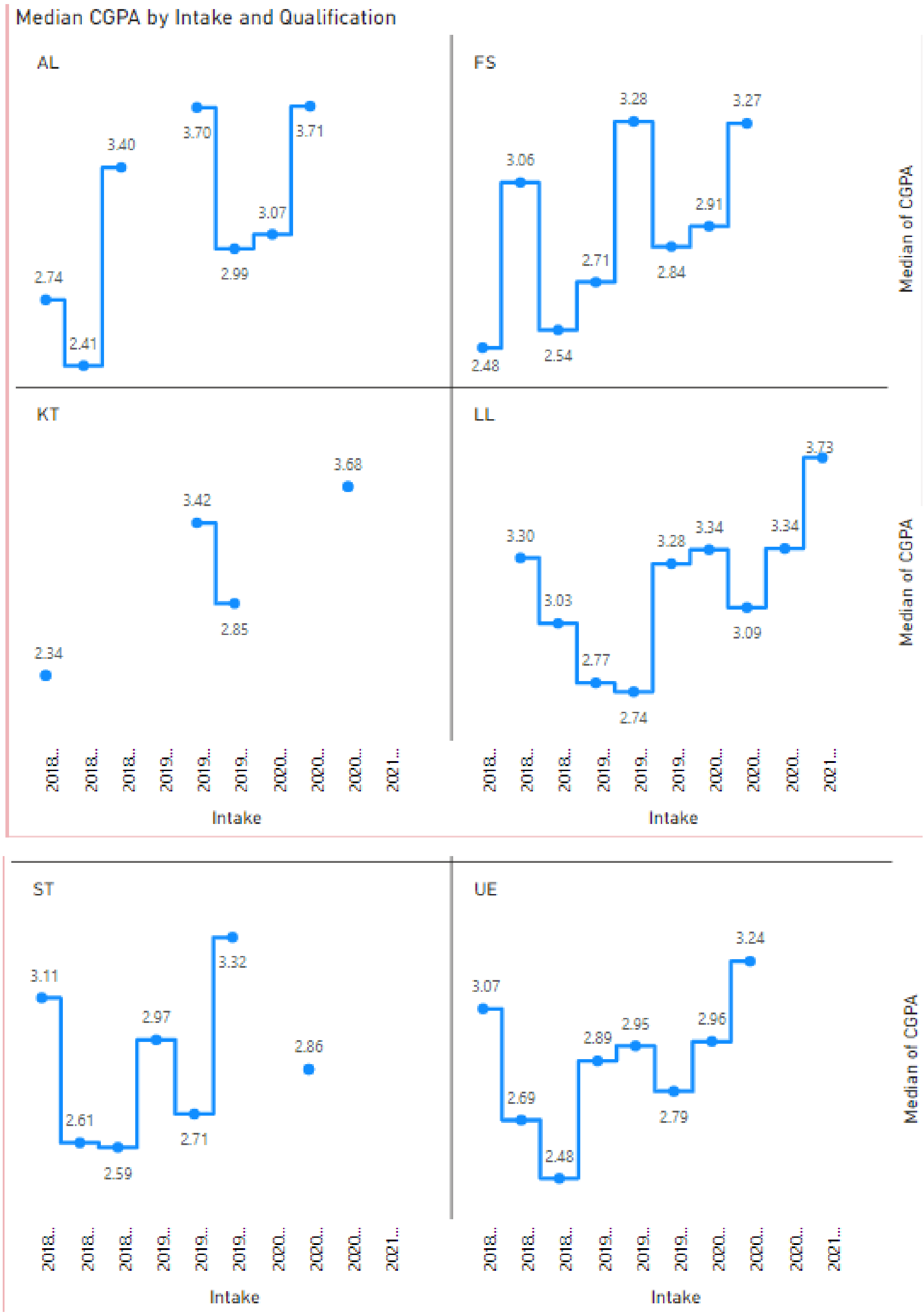

Figure 23 Median CGPA by Qualification and Intake

The line graph and scatter plot allow student with poor performance or outliers to be identified efficiently so that aids and support can be provided. For instance, targeted support or consultation session could be provided for group of students with poor performance.

4.1.3 Third Dashboard – Graduate on Time

Since the 2018 intakes, 436 students have graduated, with 380 graduating on time and the remaining 56 students does not. The majority of students graduate with merit, followed by distinction, and the fewest with a regular degree. For those who do not graduate on time, the majority of their degrees are only normal degrees with CGPA less than 3.00. In terms of qualification, UEC had the highest percentage of students who did not graduate on time when compared to the others.

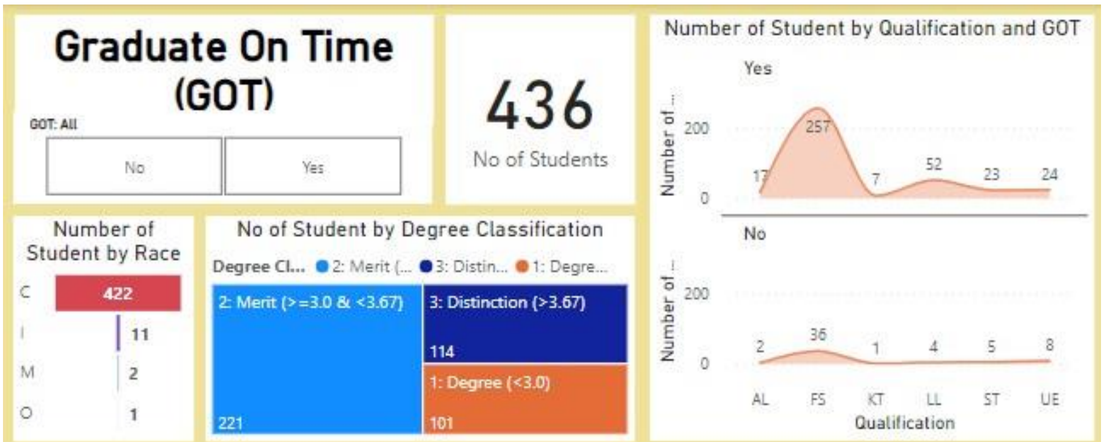

Figure 24 GOT Distribution

The May 2019 intake had the highest number of students that graduated on time. There are 98 more students who graduate on time compared to those who do not. However, it has been shown that the January to October 2018 intake featured an unusually high proportion of students who did not graduate on time. This condition could be caused by pandemic covid-19, which disrupts the students' study schedule.

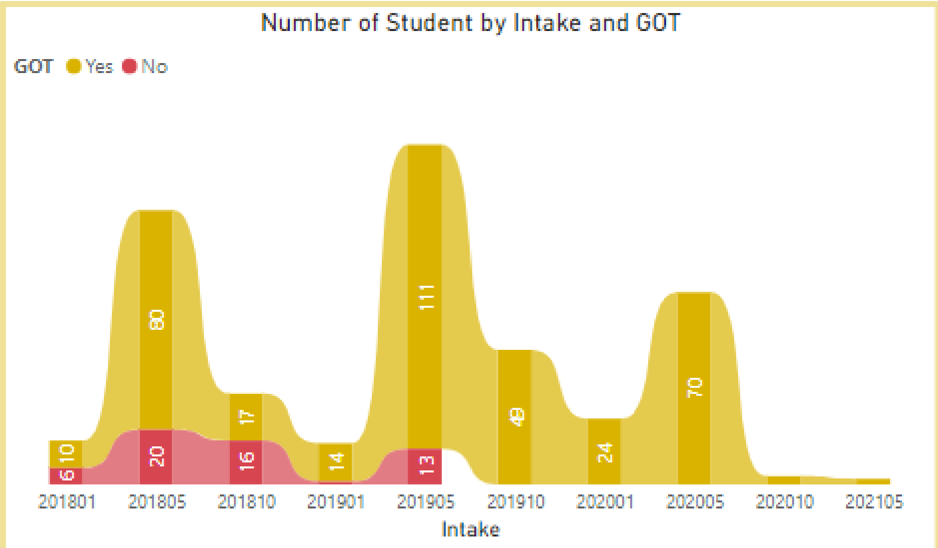

Figure 25 No of Student by Intake and GOT

The chance of not graduating on time is highly related to the degree classification and CGPA, according to key influencer indicators. It has been demonstrated that people with normal degree performance are 5.13 times more likely to not graduate on time. Furthermore, the likelihood of not graduating on time increases as the CGPA decreases. With segmentation, it was found that 75% of the students with CGPA less than or equal to 2.6571 are not graduating on time.

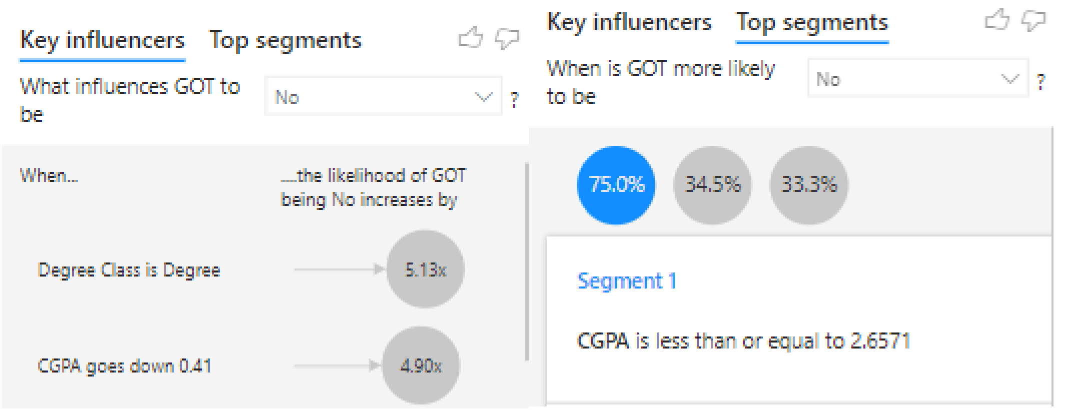

Figure 26 Key Influencer and Segment for GOT

According to the forecast graph, few students from the October 2020 to January 2021 intakes are anticipated to graduate on time. When drill down on the data point, the forecast list displays the student's details that they are unlikely to graduate on time, revealing that the students have low credit hour earned and a poor CGPA. By using the forecast graph, stakeholders may identify students who are unlikely to graduate on time and targeted solutions to assist them.

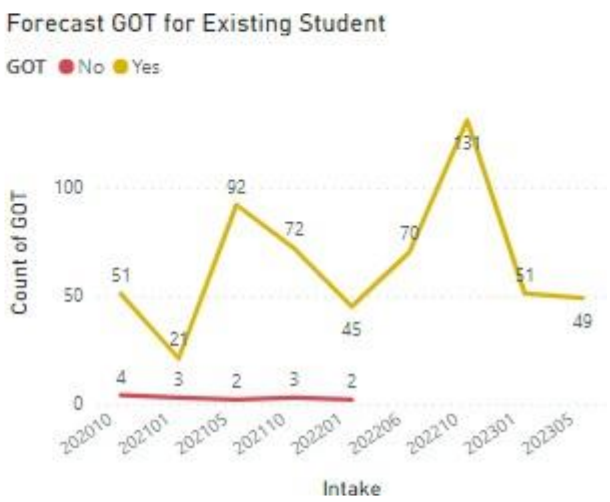

Figure 27 Forecast of GOT for existing student

4.1.4 Fourth Dashboard – Student Study Status Distribution

A line graph is used to plot the number of students in each study status, such as active, deferred, graduated, and withdrawn. Active refers to current students who are continuing their studies, defer refers to students who have postponed their studies, graduate refers to students who have finished their studies, and withdrawn refers to students who have stopped their studies and left the university.

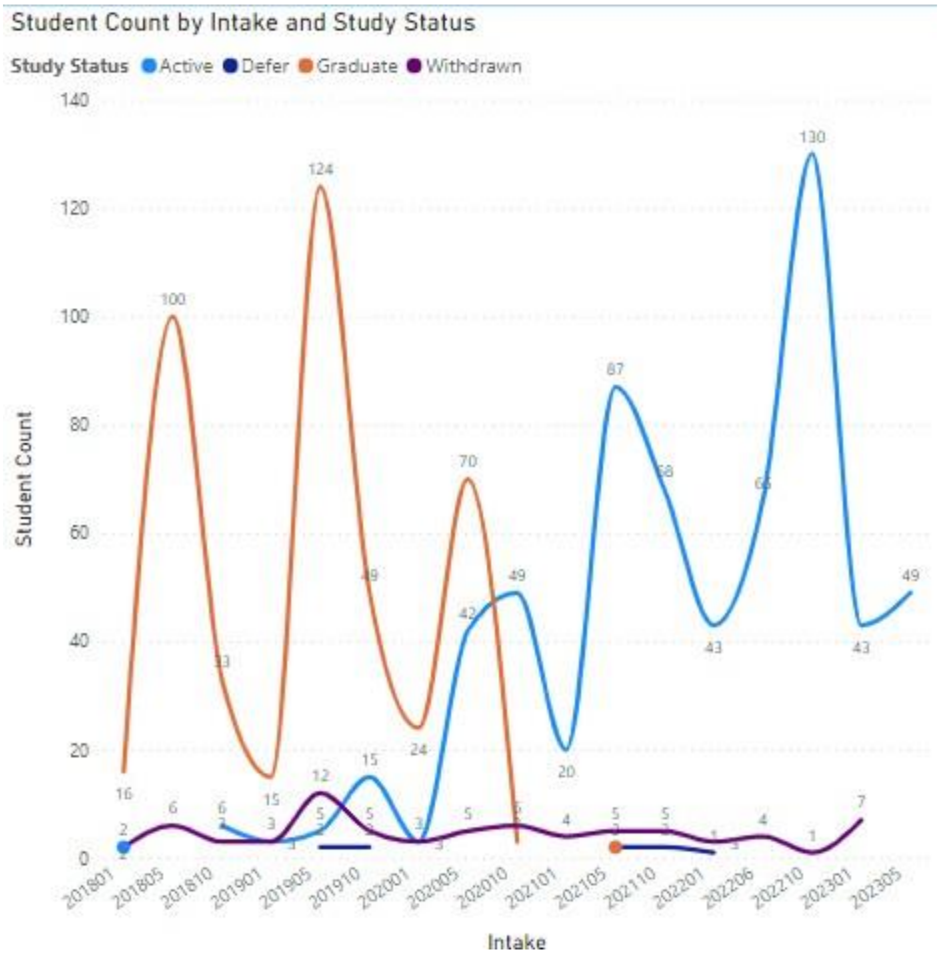

Figure 28 Student Count by Study Status and Intake

As expected, the number of active students decreased across different intakes as their status was changed to graduate. However, the graph had revealed that there are still some active students from the previous intake who did not graduate on time.

When looking at withdrawal cases, it was discovered that the May 2019 Intake had the most students with withdrawn status. Drilling down into the data point revealed that 7 out of the 12 withdrawn students had withdrawn after earning 56 credit hours and being in their second year. Their withdrawal could be due to a pandemic of Covid-19 outbreak in early 2020 which is at the start of their second year.

According to the key influencer analysis, the likelihood of withdrawal increases significantly when a student is in their first trimester or has a CGPA less than 2.00. Furthermore, male students withdraw from university at a higher rate than female students. According to segmentation, 46.4% of students with 26 or fewer credit hours withdrawn. With this insight, stakeholders can examine the syllabus for the first trimester of the programme structure to ensure that students are interested and capable of coping with it.

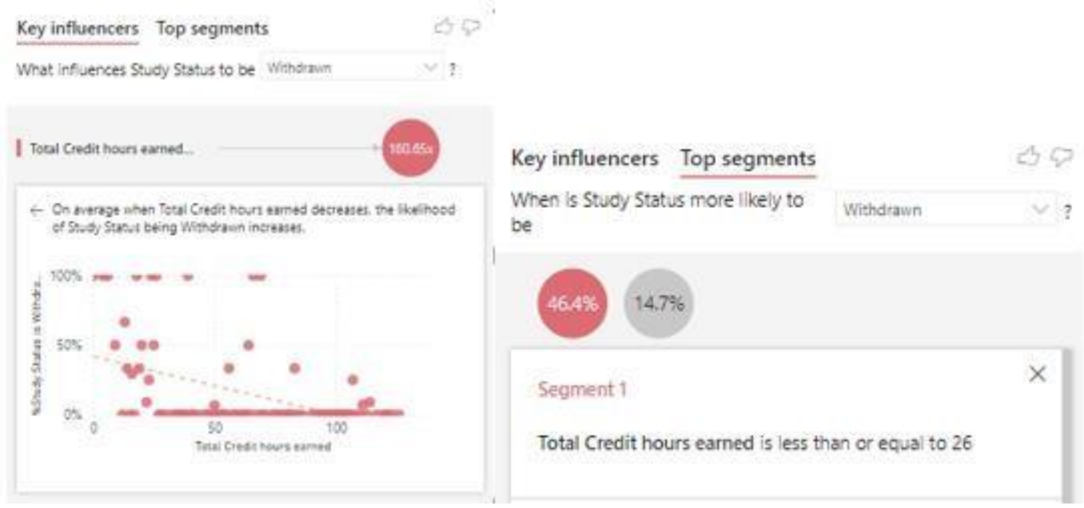

Figure 29 Key Influencer and Segment for Study Status

4.2 System Usability Survey Evaluation

According to usability expert Nielsen (2000), five is the optimal user size for usability testing where most of the usability problem can be discovered. Due to time constraints, only five users that are actively involved in decision making such as head of department, head of programme, and lecturers from the faculty are invited to evaluate the dashboard. During the testing, the users are given time to explore and interact with the dashboard developed on Power BI. The users are expected to get meaningful analysis and insight regarding students' performance from the dashboard. The participating users had filled up a google form after testing out the dashboard. The responses are then analysed and SUS score computed as followed.

Table 10 SUS Survey Question

No	Question
1	I think that I would like to use this dashboard frequently.
2	I found the dashboard unnecessarily complex.
3	I thought the dashboard was easy to use.
4	I think that I would need the support of a technical person to be able to use this dashboard.
5	I found the various functions in this dashboard were well integrated.
6	I thought there was too much inconsistency in this dashboard.
7	I would imagine that most people would learn to use this dashboard very quickly.
8	I found the dashboard very troublesome to use.
9	I felt very confident using the dashboard.
10	I needed to learn a lot of things before I could get going with this dashboard.

Table 11 SUS Responses and Scoring

Participant	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	SUS Score
P1	5	1	4	2	4	1	5	1	4	5	**80.0**
P2	5	2	4	1	4	2	4	2	4	1	**82.5**
P3	4	2	4	2	4	2	4	2	3	1	**75.0**
P4	4	3	4	2	4	1	4	2	3	2	**72.5**
P5	4	2	4	2	4	3	4	2	5	2	**75.0**

SUS Score	Grade	Adjective Rating
> 80.3	A	Excellent
68 – 80.3	B	Good
68	C	Okay
51 – 68	D	Poor
< 51	F	Awful

Figure 30 Guideline of SUS Score

The average SUS score obtained by the five participants is 77. According to the general guideline provided by Sauro (2011), 77 considered as good indicating that the dashboard developed is deemed satisfied and capable for users to interact with minimal stress and errors. In addition to the SUS scoring question, two other questions are asked to assess the clarity and usefulness of the dashboard in monitoring student performance. From the responses, all 5 users agreed that the data visualisation is clear and easy to understand, and the dashboard is able to provide insight for decision making based on student performance. Besides, some user had commented that the interface design is simple and clear, and each data segment is responsive and link to other relevant data. User also feedback that improvement can be made on the position of filter to prevent confusion.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

Dashboarding is an effective way for analysing the trend of student performance in different aspect and act as support in decision making. In this project, all of the objectives of this project have been successfully achieved to implement the comprehensive dashboard for monitoring of student performance.

The first objective achieved involved data pre-processing on the dataset, ensuring that the data used for subsequent visualisation was clean, accurate, and ready for exploration.

The second objective which focused on feature engineering is achieved by discovering significant features with Scikit-learn library and predict performance with significant features. This step had provided valuable insights for the development of dashboard and showcase factors that influencing academic outcomes.

The third objective of developing for student performance analysis is achieved by using Power BI. The Power BI dashboard serves as a comprehensive tool for educators and managers to easily visualise and interpret student performance in a glance of view.

Lastly, the fourth objective achieved focus on the evaluation of the student performance dashboard through the System Usability Scale (SUS). The positive feedback from SUS further indicates the dashboard is user friendly and meets users' satisfaction.

In a nutshell, the successful completion of these objectives had made contribution to the area of educational data analysis. Execution of this project serve as a valuable resource for institutions that interested to enhance their monitoring of student performance.

5.2 Recommendations for future work

5.2.1 User-oriented evaluation

The evaluation method adopted in this project focuses on the usability of the dashboard. An iterative process such as prototyping methodology can be used to obtain continuous feedback from users and perform subsequent dashboard enhancements to better evaluate and improve the development of performance dashboards.

5.2.2 Collect more data attributes

The attributes used in this project is mostly limited to the demographic and only consist of aggregated GPA and CGPA. Student performance such as academic result and graduate on time might be affected by other factors such as attendance, student survey and etc. Increase in data attribute can improve the analysis result and performance in predicting student performance.

5.2.3 Broader area of analysis

In this project, the dashboard developed focus on analysing student performance in term of graduate on time, performance and study status. However, there is room for further exploration into additional areas such as the achievement of programme learning outcome and course learning outcome which could be analysed by visualising students’ coursework and assessment result.

5.2.4 Real time dashboard

The dashboard developed rely on the dataset provided by the faculty in excel format and require further processing with Python. If there are APIs that allow connection to the database containing the most recent data, the dashboard can be improved to provide real-time analysis. It could also integrate with Power Automate to run any python script needed to pre-process the data before visualising it in Power BI.

REFERENCES

Abu Saa, A., Al-Emran, M. and Shaalan, K., 2019. Factors Affecting Students' Performance in Higher Education: A Systematic Review of Predictive Data Mining Techniques. *Technology, Knowledge and Learning*, [online] 24(4), pp.567–598. https://doi.org/10.1007/S10758-019-09408-7/TABLES/14.

Chanlekha, H. and Niramitranon, J., 2018. Student performance prediction model for early-identification of at-risk students in traditional classroom settings. *MEDES 2018 - 10th International Conference on Management of Digital EcoSystems*, [online] pp.239–245. https://doi.org/10.1145/3281375.3281403.

Denwattana, N. and Saengsai, A., 2017. A framework of Thailand higher education dashboard system. *20th International Computer Science and Engineering Conference: Smart Ubiquitos Computing and Knowledge, ICSEC 2016*. https://doi.org/10.1109/ICSEC.2016.7859883.

Dewan, P. and Librarian, R., 2015. *Words Versus Pictures: Leveraging the Research on Visual Communication*.

Eckerson, W., 2010. *Performance Dashboards: Measuring, Monitoring, and Managing Your Business*. Wiley.

Gordanier, J., Hauk, W. and Sankaran, C., 2019. Early intervention in college classes and improved student outcomes. *Economics of Education Review*, 72, pp.23–29. https://doi.org/10.1016/j.econedurev.2019.05.003.

Hastie, T., Tibshirani, R. and Friedman, J., 2008. *Springer Series in Statistics The Elements of Statistical Learning Data Mining, Inference, and Prediction*.

Islam, M. and Jin, S., 2019. An Overview of Data Visualisation. *International Conference on Information Science and Communications Technologies: Applications, Trends and Opportunities, ICISCT 2019*. https://doi.org/10.1109/ICISCT47635.2019.9012031.

Khan, A. and Ghosh, S.K., 2021. Student performance analysis and prediction in classroom learning: A review of educational data mining studies. *Education and Information Technologies*, [online] 26(1), pp.205–240. https://doi.org/10.1007/S10639-020-10230-3/METRICS.

Khatibi, V., Keramati, A. and Shirazi, F., 2020. Deployment of a business intelligence model to evaluate Iranian national higher education. *Social Sciences & Humanities Open*, 2(1), p.100056. https://doi.org/10.1016/J.SSAHO.2020.100056.

Lara, J.A., Lizcano, D., Martínez, M.A., Pazos, J. and Riera, T., 2014. A system for knowledge discovery in e-learning environments within the European Higher Education Area - Application to student data from Open University of Madrid, UDIMA. *Computers and Education*, 72, pp.23–36. https://doi.org/10.1016/j.compedu.2013.10.009.

Meyliana, Widjaja, H.A.E. and Santoso, S.W., 2014. University dashboard: An implementation of executive dashboard to university. In: *2014 2nd International*

Conference on Information and Communication Technology (ICoICT). pp.282–287. https://doi.org/10.1109/ICoICT.2014.6914080.

Moss, L.T. and Atre, S., 2003. *Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications*. [online] *undefined*. Available at: <https://www.semanticscholar.org/paper/Business-Intelligence-Roadmap%3A-The-Complete-Project-Moss-Atre/4e9078005edbb68b0a89d0b2c0c493e09c7f2b87> [Accessed 17 November 2022].

Muntean, M., Sabau, G., Bologa, A.-R., Surcel, T. and Florea, A., n.d. Performance Dashboards for Universities. [online] Available at: <http://www.ase.ro> [Accessed 15 March 2023].

Nafiisah, N., Noor Azuar, N., Lufti, S.L., Darweena, S., Ahmad, S. and Feisal, A., 2021. INTERACTIVE DASHBOARD FOR TRACKING SYSTEM DASHBOARD USING POWER BI. *International Alumni Convention and Innovation Exhibition (INACIX)* .

Nielsen, J., 2000. Why you only need to test with 5 users. [online] Nielsen Norman Group. Available at: <https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/> [Accessed 15 December 2023].

Olszak, C.M., 2014. Business Intelligence in cloud. *Polish Journal of Management Studies*, [online] 10(Vol. 10, No. 2), pp.115--125. Available at: <https://www.infona.pl//resource/bwmeta1.element.baztech-fe7a1044-ff3d-4ca9-aa57-ccdca425eea1> [Accessed 10 November 2023].

Romero, C. and Ventura, S., 2010. *Educational data mining: A review of the state of the art*. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, https://doi.org/10.1109/TSMCC.2010.2053532.

Santoso, L.W. and Yulia, 2017. Data Warehouse with Big Data Technology for Higher Education. *Procedia Computer Science*, 124, pp.93–99. https://doi.org/10.1016/J.PROCS.2017.12.134.

Sauro, J., 2011. *Measuring Usability with the System Usability Scale (SUS) – MeasuringU*. [online] Available at: <https://measuringu.com/sus/> [Accessed 5 November 2023].

Thanh, T.B., TD, D., Hong Le, D. and Aquino Jr., P.G., 2021. Simple Student Grades Analytics Dashboards In Higher Education In Vietnam. *Journal of Contemporary Issues in Business and Government*, [online] 27(1), pp.445–453. Available at: <https://cibgp.com/article_7424.html>.

APPENDICES

APPENDIX A: SUS Survey Google Form

Student Performance Dashboard Evaluation

This questionnaires is targeted to evaluate and collect feedback regarding the usage of Power BI dashboard to monitor student performance.

You are required to answer the following questionnaire after interacting with the dashboard.

* Indicates required question

1. Email *

2. I think that I would like to use this dashboard frequently. *

   Mark only one oval.

	1	2	3	4	5	
Stro	◯	◯	◯	◯	◯	Strongly agree

3. I found the dashboard unnecessarily complex. *

   Mark only one oval.

	1	2	3	4	5	
Stro	◯	◯	◯	◯	◯	Strongly agree

4. I thought the dashboard was easy to use. *

Mark only one oval.

	1	2	3	4	5	
Stro	◯	◯	◯	◯	◯	Strongly agree

5. I think that I would need the support of a technical person to be able to use this dashboard. *

Mark only one oval.

	1	2	3	4	5	
Stro	◯	◯	◯	◯	◯	Strongly agree

6. I found the various functions in this dashboard were well integrated. *

Mark only one oval.

	1	2	3	4	5	
Stro	◯	◯	◯	◯	◯	Strongly agree

7. I thought there was too much inconsistency in this dashboard. *

Mark only one oval.

	1	2	3	4	5	
Stro	◯	◯	◯	◯	◯	Strongly agree

8. I would imagine that most people would learn to use this dashboard very quickly. *

Mark only one oval.

	1	2	3	4	5	
Stro	◯	◯	◯	◯	◯	Strongly agree

9. I found the dashboard very troublesome to use. *

Mark only one oval.

	1	2	3	4	5	
Stro	◯	◯	◯	◯	◯	Strongly agree

10. I felt very confident using the dashboard. *

Mark only one oval.

	1	2	3	4	5	
Stro	◯	◯	◯	◯	◯	Strongly agree

11. I needed to learn a lot of things before I could get going with this dashboard. *

Mark only one oval.

	1	2	3	4	5	
Stro	◯	◯	◯	◯	◯	Strongly agree

12. The data visualization is clear and easy to understand. *

Mark only one oval.

	1	2	3	4	5	
Stro	◯	◯	◯	◯	◯	Strongly agree

13. The dashboard able to provide insight for decision making based on student performance. *

Mark only one oval.

	1	2	3	4	5	
Stro	◯	◯	◯	◯	◯	Strongly agree

14. Any other feedback on the dashboard?

____
____
____
____
____