

Knowledge Visualization of Performance of Student Based on Attendance

By

Hoe Jun Sheng

A REPORT

SUBMITTED TO

Universiti Tunku Abdul Rahman

in partial fulfillment of the requirements

for the degree of

BACHELOR INFORMATION SYSTEMS (HONOURS)

BUSINESS INFORMATION SYSTEMS

Faculty of Information and Communication Technology
(Kampar Campus)

JAN 2024

REPORT STATUS DECLARATION FORM

Title: Knowledge Visualization of Performance of Student
Based on Attendance

Academic Session: 202401

I HOE JUN SHENG

(CAPITAL LETTER)

declare that I allow this Final Year Project Report to be kept in

Universiti Tunku Abdul Rahman Library subject to the regulations as follows:

1. The dissertation is a property of the Library.
2. The Library is allowed to make copies of this dissertation for academic purposes.

Verified by,

(Author's signature)

(Supervisor's signature)

Address:

33, Hala Bekor 17,

Taman Pertama,

30100, Ipoh, Perak.

Supervisor's name

SHAKIROH BINTI KHAMIS

Date: 23 April 2024

Date: 25/04/2024

Universiti Tunku Abdul Rahman			
Form Title : Sample of Submission Sheet for FYP/Dissertation/Thesis			
Form Number: FM-IAD-004	Rev No.: 0	Effective Date: 21 JUNE 2011	Page No.: 1 of 1

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY

UNIVERSITI TUNKU ABDUL RAHMAN

Date: 11 April 2024

SUBMISSION OF FINAL YEAR PROJECT

It is hereby certified that **Hoe Jun Sheng** (ID No: **19ACB03230**) has completed this final year projec entitled “ **Knowledge Visualization of Performance of Student Based on Attendance** ” under the supervision of **Dr Shakiroh Binti Khamis** (Supervisor) from the Department of **DDET**, Faculty of Information and Communication Technology, and _____ (Co-Supervisor)* from the Department of _____, Faculty/Institute* of _____.

I understand that University will upload softcopy of my final year project in pdf format into UTAR Institutional Repository, which may be made accessible to UTAR community and public.

Yours truly,



Hoe Jun Sheng

(Student Name)

*Delete whichever not applicable

DECLARATION OF ORIGINALITY

I declare that this report entitled “**Knowledge Visualization of Performance of Student Based on Attendance**” is my own work except as cited in the references. The report has not been accepted for any degree and is not being submitted concurrently in candidature for any degree or other award.

Signature : 

Name : Hoe Jun Sheng

Date : 11 April 2024

ACKNOWLEDGEMENTS

I would like to express my sincere thanks and appreciation to my supervisors, Dr Shakiroh Binti Khamis who has given me this bright opportunity to engage in a knowledge visualization project, her expertise, guidance, and unwavering support throughout this endeavor have been invaluable. It is my first step to establish a career in this field. A million thanks to you.

I appreciate the countless researchers and authors whose work served as the basis for our project. Their pioneering efforts in various fields of knowledge have inspired and informed this visualization, and I am grateful for their contributions to the global body of knowledge. I must also say thanks to my friends and family who stood by me during the challenging moments of this project, providing us with emotional support and encouragement.

Finally, I am thankful to our educational institution, UTAR, for providing the resources and environment necessary for the successful completion of this project.

ABSTRACT

The project titled "Knowledge Visualization of Student Performance Based on Attendance" represents a multifaceted exploration that merges research and development endeavors. In the contemporary educational landscape, student performance evaluation stands as a pivotal facet of pedagogical improvement. Recognizing the significance of attendance as a determinant of academic success, this project embarks on a comprehensive investigation to analyze, visualize, and enhance the understanding of the intricate relationship between attendance and student performance. The research component of this project involves an in-depth analysis of attendance data collected from diverse educational institutions, encompassing different grade levels and disciplines. Through rigorous statistical analysis and data mining techniques, we aim to uncover meaningful patterns, correlations, and insights that illuminate the impact of attendance on student outcomes. Simultaneously, leveraging the robust features of Power BI, the project translates these findings into interactive, visually-rich dashboards and reports, facilitating a deeper understanding of the attendance-performance relationship. This holistic approach aims to empower educators, administrators, and students with the means to make informed decisions, optimize attendance strategies, and ultimately enhance academic outcomes in the ever-evolving educational ecosystem. Ultimately, the fusion of research and development in this project aims to empower educational stakeholders with actionable insights and practical solutions to enhance student success rates. The knowledge visualization techniques employed will not only shed light on the nuances of attendance's impact but also serve as a catalyst for data-driven decision-making and continuous improvement in educational institutions. In an era characterized by data-driven decision-making, this project represents a crucial step toward the enhancement of educational practices by providing a holistic view of the interplay between attendance and student performance. As we delve into this dynamic intersection of research and development, we envision a future where educational outcomes are optimized through a deeper understanding of the role attendance plays in shaping student success.

TABLE OF CONTENTS

TITLE PAGE	i
REPORT STATUS DECLARATION FORM	ii
SUBMISSION SHEET FOR FYP	iii
DECLARATION OF ORIGINALITY	iv
ACKNOWLEDGEMENTS	v
ABSTRACT	vi
TABLE OF CONTENTS	vii
LIST OF FIGURES	x
LIST OF TABLES	xi
LIST OF SYMBOLS	xii
LIST OF ABBREVIATIONS	xiii
CHAPTER 1 INTRODUCTION	1
1.1 Problem Statement and Motivation	1
1.2 Project Scope	2
1.3 Project Objectives	2
1.4 Impact, Significance and Contribution	3
1.5 Background Information	4
1.6 Report Organization	5
CHAPTER 2 LITERATURE REVIEW	6
2.0 Previous Works on Knowledge Visualization of Student Performance	6 6
2.1 Heatmap Visualization	6
2.1.1 Heatmap Visualization Advantages and Disadvantages	7
2.2 Scatter Plot and Distribution Visualization	8
2.2.1 Scatter Plot Visualization Advantages and Disadvantages	9
2.3 Logistic Regression Visualization	10
2.3.1 Logistic Regression Visualization Advantages and Disadvantages	11
2.4 Web-Based Visual Analytics System	12

2.4.1 Web-Based Visual Analytics System Advantages and Disadvantages	13
2.4 Learning Analytics Dashboard (LAD)	14
2.4.1 Learning Analytic Dashboard Advantages and Disadvantages	15
2.6 Previous Use of Prescriptive Analysis from Different Industry	16
2.7 Comparison of Previous Works with The Proposed Solution	16
CHAPTER 3 PROPOSED METHOD/APPROACH	18
3.1 Data Science Life Cycle (DSLCL)	18
3.1.1 Business Understanding	19
3.1.2 Data Mining	19
3.1.3 Data Cleaning	19
3.1.4 Data Exploration	20
3.1.5 Feature Engineering	20
3.1.6 Predictive Modeling	20
3.1.7 Data Visualization	20
3.2 Power BI Architecture	21
CHAPTER 4 SYSTEM DESIGN	24
4.1 Data Science Life Cycle (DSLCL) Implementation	24
4.1.1 Business Understanding	24
4.1.2 Data Mining	24
4.1.3 Data Cleaning	32
4.1.4 Data Exploration	38
4.1.5 Feature Engineering	39
4.1.6 Predictive Modeling	40
4.1.7 Data Visualization	41

4.2	System Flowchart	43
4.3	System Block Diagram	44
CHAPTER 5 SYSTEM IMPLEMENTATION		46
5.1	Hardware	46
5.2	Software	47
5.2.1	Power BI	47
5.2.2	Microsoft Excel Spreadsheet	48
5.3	Dashboard Operation	48
5.4	Implementation Issues and Challenges	53
5.5	Timeline	54
CHAPTER 6 FINDINGS		55
6.1	Descriptive Analysis	55
6.2	Descriptive Analysis	56
6.3	Descriptive Analysis	56
CHAPTER 7 CONCLUSION AND FUTURE WORKS		58
7.1	Descriptive Analysis	58
7.2	Descriptive Analysis	59
REFERENCES		60
APPENDIX		
A.1	Dummy Data	A-1
A.2	Weekly Report	A-3
A.3	Poster	A-15
PLAGIARISM CHECK RESULT		P-1
FYP 2 CHECKLIST		C-1

LIST OF FIGURES

Figure Number	Title	Page
Figure 2.1	Greyscale Heatmap of students' performance for three course pairs	6
Figure 2.2	Scatter Plot and Distribution Visualization of Grade Prediction - Computer Science	8
Figure 2.3	Logistic regression visualization	10
Figure 2.4	Web-Based Visualization	12
Figure 2.5	Learning Analytics Dashboard	14
Figure 3.1	Data Science Life Cycle	18
Figure 3.2	Architecture of Power BI	21
Figure 4.1.3.1	Visualization from subset A	39
Figure 4.1.3.2	Visualization from subset B	39
Figure 4.1.7.1	Visualization of Descriptive Page	41
Figure 4.1.7.2	Visualization of Predictive Page	42
Figure 4.1.7.3	Visualization of Prescriptive Page	42
Figure 4.2	Flowchart of Power BI	43
Figure 4.3	Block Diagram	44
Figure 5.2.1	Microsoft Power BI	47
Figure 5.2.2	Microsoft Excel Spreadsheet	48
Figure 5.3.1	Import data from Excel to Power BI	49
Figure 5.3.2	Data in Power Query Editor before editing	49
Figure 5.3.3	Data in Power Query Editor after editing	50
Figure 5.3.4	Table view of data	50
Figure 5.3.5	Descriptive Page	51
Figure 5.3.6	Predictive Page	52
Figure 5.3.7	Prescriptive Page	53
Figure 5.5	Gantt Chart of Project	54

LIST OF TABLES

Table Number	Title	Page
Table 2.6	Previous Use of Prescriptive Analysis from Different Industry	16
Table 3.2.1	Power BI tools and features	22
Table 3.2.2	Types of data source in Power BI	22
Table 3.2.3	Some important visualization	23
Table 4.1.2	Raw Students Data Information	25
Table 4.1.3	Students Data Cleaning Process with Highlight	33
Table 4.1.5.1	Feature selection	40
Table 4.1.5.2	Feature Construction	40
Table 5.1	Specifications of laptop	46
Table 5.3.1	Example of completed data in Excel Sheet	48

LIST OF SYMBOLS

A	Arithmetic mean
n	number of values
a_i	data set values

LIST OF ABBREVIATIONS

<i>BI</i>	Business Intelligence
<i>FICT</i>	Faculty of Information and Communication Technology
<i>LAD</i>	Learning Analytics Dashboard
<i>DSLCL</i>	Data Science Life Cycle
<i>GPA</i>	Grade Point Average

CHAPTER 1

Introduction

In this chapter, we present the background and motivation of our research, our contributions to the field, and the outline of the thesis.

1.1 Problem Statement and Motivation

According to research, one of the most crucial factors influencing students' academic success and performance is their attendance [1]. Most of the education institution uses this traditional methods of attendance tracking and performance analysis which that often lack sophistication and real-time insights. The need arises to comprehensively investigate and illuminate this complex relationship between attendance and student performance. By harnessing the power of knowledge visualization techniques, particularly through the utilization of Power BI, this project seeks to address this critical gap. The challenge lies not only in conducting rigorous research to uncover meaningful patterns but also in developing an accessible, user-friendly software tool capable of translating these insights into actionable strategies for educators and students alike.

The motivation behind this project is to help educators and education institutions that are in need with this information about students' performance and grades to analyze how themselves and the group doing that affect ones performance and grades. Without this, educators are lack of understanding on how the students are performing and it is important to have this visualization for improving education system effectiveness and able to provide learning personalization for students. By doing so, we propose a visual analytic tool to perform some visual analysis of student data from programme courses that combines data visualization and machine learning approaches. We will analyze the result and present them in graphs, charts and/or diagrams for educators to easily understand and make decision based on that.

1.2 Project Scope

Throughout this project development, an graphical and actionable dashboard showing all the relevant information about the relationship between student performance and student attendance should be developed. The project will involve collecting students data such as attendance data and result data from UTAR. An advanced data visualization techniques will be employed to translate the information data into visual representations using Power BI. The project will also involve conducting thorough literature reviews to ensure that the research is built upon a strong theoretical foundation. The dashboard should also undergo rigorous testing and validation to ensure accuracy and usability.

1.3 Project Objectives

The aim of the thesis is to develop an interactive dashboard based on student attendance that able to determine the impact of students attendance on students performance. For instance, with the dashboard visualization, we also aim to evaluate the interactive dashboard based on student attendance by using usability testing.

In generally, this project will try to provide an understandable relationship between student performance and attendance across various trimester, programme and subject areas. The dashboard should also provide data-driven insights to educators to improve their decision making. Ultimately, the project aims to contribute to improved student outcomes in educational institutions.

However, the project will not cover specific educational policies or curricular changes, and it does not aim to replace the existing educational systems but to have an alternative tool to provide data-driven insights. It is not intended to be a comprehensive solution for all educational challenges but rather a valuable tool in the pursuit of improved student performance through attendance analysis.

1.4 Impact, Significance and Contribution

By having this educational tools available for the education institution, they can have a deep insights on how the student will perform regarding to their attendance performance for each subject. This could benefit them to make future prediction and make adjustment on how to improve the student performance. As the problem was the traditional method lacking of sophistication and real-time insight, many educators and institution are looking forward to have a better solution for this. And this is the time where this project may be helpful for them in terms of complexity and timing. Therefore they need to know the tools Power BI.

In the other hand, by using Power BI, educators and institutions can have a more understandable and interact-able dashboard where it visualize better than other types or system of visualization. This project has to be carried out to eliminate the problem faced from the traditional method which is used commonly by many education institution.

1.5 Background Information

Numerous research studies have examined the significance of analytic and predictive methods, and factors that influence students performance in an effort to improve educational goal achievement, provide new contemporary opportunities for maximizing educational system effectiveness, and provide learning personalization. However, aspects in predicting students' performance is a remarkable curiosity in educational perspective. One of the way to predict students' performance is through analyzing students' attendance. Students' attendance is one information that most influence on students' performance, poor attendance and lack of participate in activities, events or classes may bring awful performance of students [2].

Historically, attendance has been regarded as a measure of student engagement and a proxy for academic commitment and research has shown that student achievement and attendance are directly correlated such as low attendance rate means poor academic performance [1]. But is is really directly correlated nowadays? However, the connection between attendance and academic achievement has become more complex in today's dynamic educational environment, which is defined by technology breakthroughs, changing learning spaces, and increasingly diverse student populations [3]. Traditional approaches of gathering and analyzing attendance data, which frequently depend on labor-intensive manual procedures and paper-based systems, are no longer as efficient at capturing the intricacies of contemporary schooling. Therefore, this may show that the relationship is not directly correlated anymore nowadays. And we need to provide a knowledge visualization showing this relationship.

Previously, there is some knowledge visualization of student performance developed in different ways where we will look more into it in Chapter 2 Literature Review. Those development are different than what this project aim to develop. Those development shows more complex visualization and does not provide real-time insights to the educators and educational institutions.

Therefore, in Chapter 3, we will present a proposed method and approach to fix the issue presented in the traditional method and previous development of visualization. We will show how to use the features and tools to do visualization by using Power BI that show the relevant information of student data and show the relationship between

CHAPTER 1

student attendance and performance. We will also provide the system implementation and testing in Chapter 4 that show the process and phases on creating and designing the dashboard.

After that, in Chapter 5 we will discuss about the findings of the dashboard by analyzing each element of analytics. And finally we will discuss about the conclusion of the whole project.

1.6 Report Organization

This report is organized into 5 chapters: Chapter 1 Introduction, Chapter 2 Literature Review, Chapter 3 Proposed Method/Approach, Chapter 4 System Implementation and Testing, Chapter 5 Findings and Chapter 6 Discussion & Conclusion. The first chapter is the introduction of this project which include problem statement and motivation, project scope, project objectives, impact, significance and contribution, background information and report organization. The second chapter is the literature reviews on the previous works of knowledge visualization of student performance develop by previous researcher. The third chapter is the discussion of the proposed method or approach used to develop this project. The fourth chapter is regarding the details on how to implement and design the dashboard. The fifth chapter report about the finding of the analytics, and last chapter discuss about the conclusion of the whole project.

CHAPTER 2

Literature Reviews

2.0 Previous Works on Knowledge Visualization of Student Performance

In this chapter, we perform literature reviews on the previous works of knowledge visualization of student performance develop by previous researcher. We identify its method used and the advantages and disadvantages. Then we compare their work with the proposed solution.

2.1 Heatmap Visualization

DeCotes proposed a Heatmap visualization to show the students' grade for different course taken [4]. Heatmap is a graphical way to visualize data information where the values contained in a matrix through variations in coloring. This is very useful to visualize differences of multiple variables to show the correlations patterns. In his works, he compared 3 different courses and show the results that different course pairs yield different behaviors. The study also show that students who did poorly in level one course will also getting similar grades in the level two course. By using this, They could more easily distinguish the change in grades from one level to another.

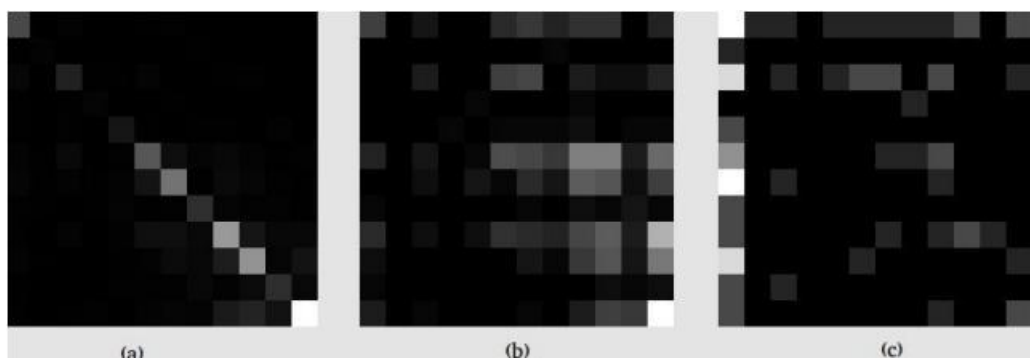


Figure 2.1 Greyscale Heatmap of students' performance for three course pairs.

2.1.1 Heatmap Visualization Advantages and Disadvantages

Although Heatmap seem to be very difficult to understand what the information about to present to decision maker, it has its own advantages that is why developers use Heatmap to perform analysis. Lets talk about its advantages. Heatmap allows a quicker general understanding of the scenario and make it very convenience with brief introduction provided. Despite the fact that Heatmap do not present much detail, but it allow a quick way to get overall conclusion [5].

For the disadvantages of Heatmap, the value differences are difficult to be precisely determined. Although it is easy to contrast the saturation level of only 2 colors, but it still not a easy work for human eye to look at. It is almost impossible for human eye to compare the shape as it get more complex [5]. For instance, decision maker might acquire incorrect perception when achieve the information in Heatmap too directly which will affect the decision outcome. So it become a very hard task for presenter to present and brief all the components and information in the Heatmap to decision maker and time consuming

2.2 Scatter Plot and Distribution Visualization

A scatter plot is a type of plot or mathematical diagram using Cartesian coordinates to display values for typically two variables for a set of data. The scatter plot and distribution visualization proposed by Rovira et al. predicted the grades of students for Computer Science course with predicted values against real values for second-year grades. The histogram provided show the distribution of the predicted grads in X-axis and real grades in Y-axis [6].

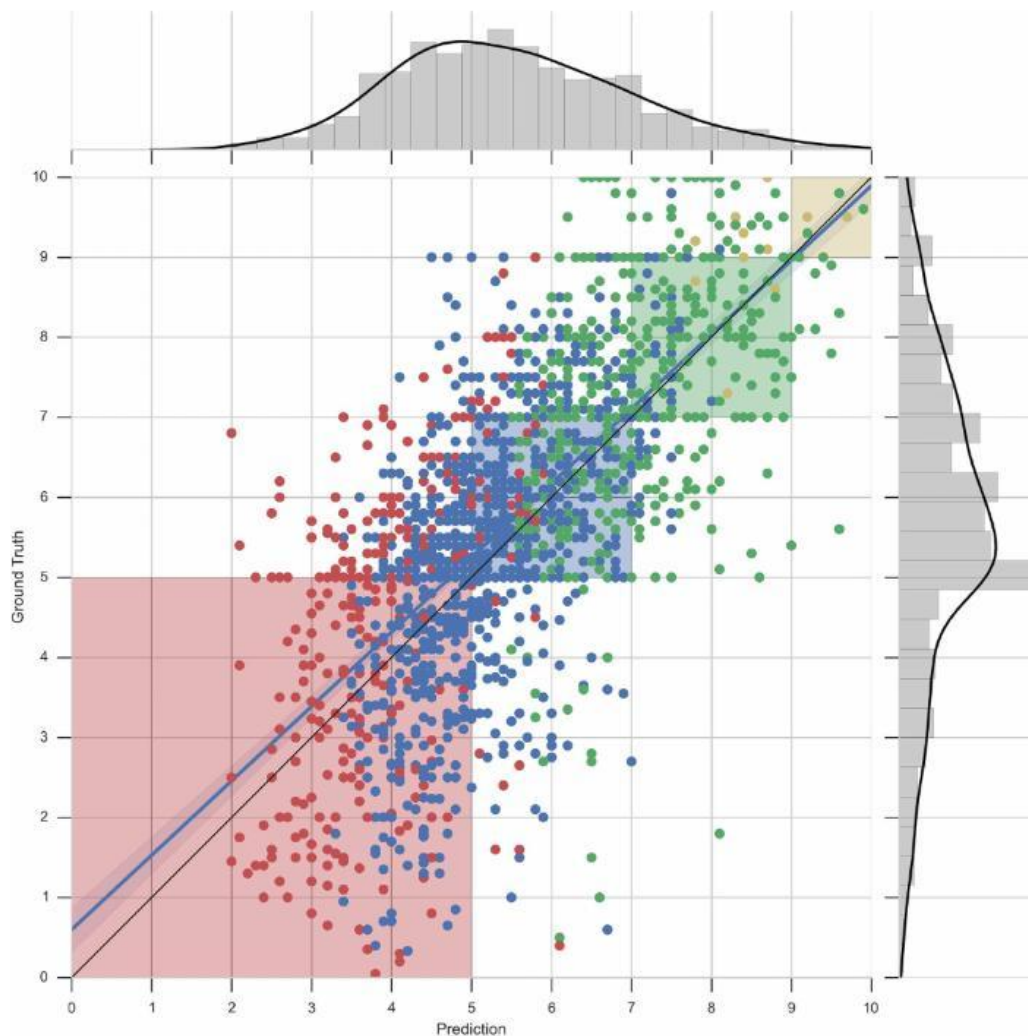


Figure 2.2 Scatter Plot and Distribution Visualization of Grade Prediction - Computer Science

2.2.1 Scatter Plot Visualization Advantages and Disadvantages

Scatter Plot is one of the easiest way to represent bivariate data. So here are some of the advantages of using a scatter plot to present information to decision maker. First and foremost, Scatter Plot are very convenience to use and present by just only locate and mark the points of information and make it easy to understand and clarify by decision maker. Scatter Plot predict how the dependent variable will behave in relation to the independent variable and determined the relationship of two variables. When one variable is easy to assess while the other variable are not, Scatter Plot become a useful tool. For example in this case, Scatter Plot is useful because the real grades from previous years can be easily assess while the predicted grades is very hard to define.

Even though Scatter Plot consists these many advantages to visualize the grades prediction, there exist some problem and disadvantages in using Scatter Plot. One of them is it only describe the interrelationship direction but not the degree of it. And from the previous work from Rovira et al. [6], the grade predictions become more challenging due to the data lack of samples of grades. Therefore, decision maker do not have sufficient information to precisely carry out prediction. The Scatter Plot also cause them to wrongly predict the grades fall in red-shaded area where students have good grades previously due to the colors of the points.

2.3 Logistic Regression Visualization

Logistic regression is a machine learning algorithm that utilized supervised learning used for classification problems trying to forecast the information. The highest probability value will be used as the predicted value. In this paper, Etemadpour et al. proposed the logistic regression visualization to present the prediction of students' grades based on their performance in previous semesters [7].

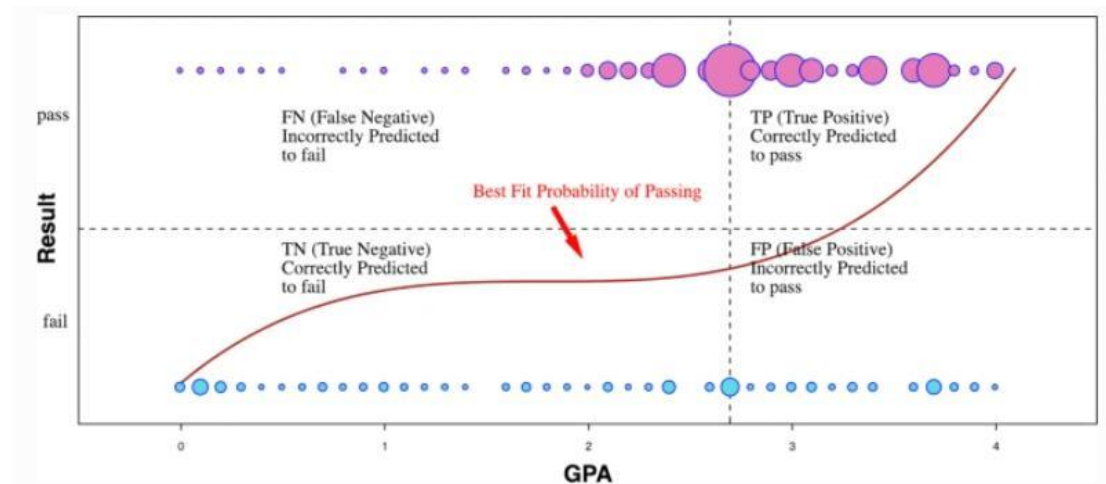


Figure 2.3 Logistic regression visualization

In the above visualization, the X-axis shows the students GPA and Y-axis shows the students' result whether pass or fail. The proposed visualization is segregated into 4 categories which are false negative which means that students were wrongly predicted to fail; true negative indicate that students were correctly predicted to fail; true positive means students were correctly predicted to pass; and false positive where students were wrongly predicted to pass. Then, developers use activation function to convert the outcome into the curved line where it represents the probability of passing.

2.3.1 Logistic Regression Visualization Advantages and Disadvantages

There are multiple advantages to employ logistic regression to find the probability of success and failure. Firstly is its simplicity and transparency. Logistic regression is one of the most straightforward predictive algorithms available. It allows us to observe the entire process and comprehend what is happening at each stage. For instance, logistic regression provides probabilistic output where it delivers a decimal number that can be clarified as a sample possibility. By this, a decision maker can get a more accurate prediction. Furthermore, logistic regression presents feature importance assessment and provides direction for each feature's influence on the value [8].

The disadvantages of logistic regression are the assumption of log linearity, which is hard to hold, hoping there exist any correlations between the predictors and the response's log is typically unfeasible. Uncertainty in feature importance is also one problem of logistic regression. The weight of every feature might mean the amount and how the feature interacts with the response, which can be ambivalent. It should rely on independent and dependent variables' association and also their connection. Logistic regression is also fragile against big-influentials which might damage the model. It also requires sufficient data in order to get a more accurate prediction [8].

2.4 Web-Based Visual Analytics System

[9] developed this system that visualize number of methods for visualizing and analyzing academic data. This system able to demonstrate the connections between grades and attendance, genders and grades, parental education levels and grades, and student populations and genders. In this work, several machine learning models were constructed to predict student performance based on the absence rate, and visualization tool were used to assess the efficacy of the models.

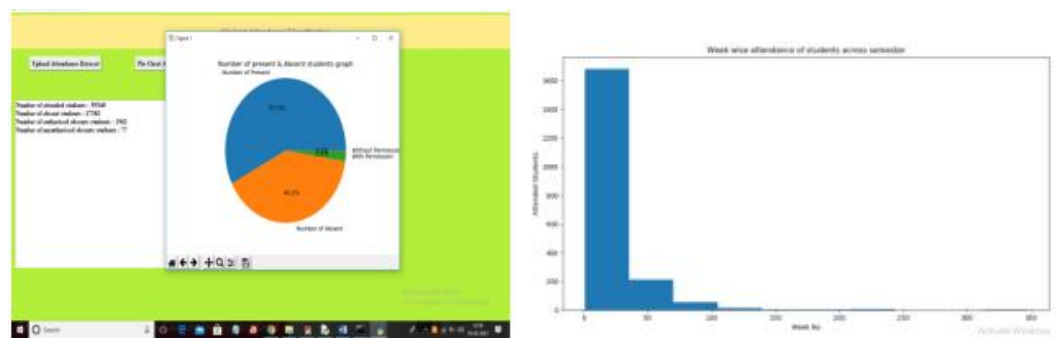


Figure 2.4 Web-Based Visualization

The methodology behind this web-based visual analytics system is using and applied the predictive analytics techniques from previous work development to the system, and performed descriptive statistical analysis to gain insight. This system enable educators to upload dataset into the model, create visualization based on the dataset and filter or sort the data to match the outcome of the graphics.

2.4.1 Web-Based Visual Analytics System Advantages and Disadvantages

Although this system is mostly different from the previous work of visualization, as [9] mentioned that some disadvantages of this system is it required a manual data input to record the attendance, so it required more human effort. Since the records are kept in handwritten registers, getting the information is not as simple. For instance, the appropriate field must get the relevant feed input. If the incorrect inputs are made, the application might not function properly. Thus, educators and educational institution finds it challenging to use. Another disadvantages is that the system does not visualize all the information at once, so user has to often navigate to other features to find what they are looking for.

Despite its disadvantages, as said earlier, this work visualize differently from the previous work of visualization. It provide an easy understandable visualization for the educators and educational institution. It does not contain complex visualization and the procedures to use it is very simple. It provide an actionable interface in the system to help user gain insight more easily.

2.5 Learning Analytics Dashboard (LAD)

This dashboard proposed by [10] showed all 3 components in 1 dashboard which include the descriptive, predictive and prescriptive. This dashboard has a great level of openness and tells the educators how trustworthy the prediction models are, and the important aspects that influence the predictions to let them have more clarity on how the process go through from descriptive to predictive.

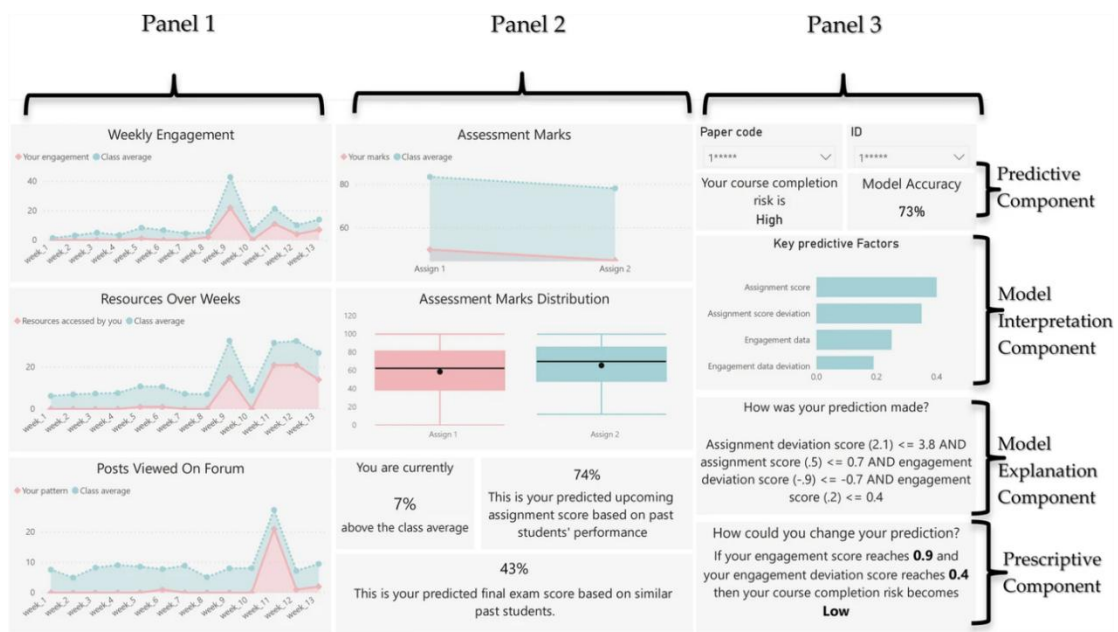


Figure 2.5 Learning Analytics Dashboard

Above figure show that the 3 components of descriptive, predictive and prescriptive were assign to the 3 panels. Panel 1 shows how well the students participation are. This panel can only show the descriptive components. Panel 2 shows the data of students' performance. This panel involve 2 analytics components which are: Descriptive that show the students assessment marks and distribution. Predictive that give students forecasts of assignment and final exam scores according to previous students identical learning type. Panel 3 contains all 3 analytics components. This panel show a comprehensive forecast to students based on their descriptive components. This panel also provide explanation for students regarding how they got the forecast value based on their learning type and give recommendation to adjust how students should perform in order to get good performance.

2.5 Learning Analytics Dashboard Advantages and Disadvantages

The advantages of this LAD is the dashboard able to provide all 3 components of descriptive, predictive and prescriptive analytics in one single dashboard page. This allow users and educational institution to have a agile look and make easier decisions quicker. Furthermore, this dashboard is so transparent to provide users with predictive reliability, and also provide interpretable dashboard element to go through the process of predictive to prescriptive [10].

About the disadvantages, as [10] mentioned in their report, there were insufficient of studies to identify the impacts on users outcome based on the visualizations method and features on the dashboard. Moreover, the author were aware of the ethical use of personal data of students where it may demotivate students to learn due to wrong advise on their studies.

2.6 Previous Use of Prescriptive Analysis from Different Industry

Industry	Use of Prescriptive Analysis	Prescriptive Components
Healthcare	Managing patient	- Anomaly detection [11]
	Managing medical decision	- Patient's data [12]
Manufacturing	Automate decision making	- Optimisation techniques [13] [14]
	Improve sustainable manufacturing	- Constraint satisfaction modeling [14]
Business	Business process optimization	- Real time data[15]
	Manage sales [16]	- Page views - Email interactions - Site search - Content Engagement

Table 2.6 Previous Use of Prescriptive Analysis from Different Industry

2.7 Comparison of Previous Works with The Proposed Solution

In the previous work, most of the visualization utilize the machine learning model to create the visualization such as heatmap matrix, scatter plot, logistic regression and others model. Although these models provide a quick insights on the data, but not all people can understand the relationship, meaning or information that the visualization trying to present because the complexity is very high and it is very sensitive to data accuracy. Previous work on visualization also do not present the information at once, user had to apply different algorithms in order to get the ideal visualization for them to study the relationship. Compared with this proposed solution in this project, with

CHAPTER 2

using Power BI, the model aim to provide all relevant information that educators and educational institution needed. It enable to provide a all in one visualization for user to study its relationship. The graphic is less complex comparing to machine learning model and it provide an actionable dashboard for user to filter and sort.

CHAPTER 3

Proposed Method/Approach

In this part we will cover the methodology used, which will be carried out in stages, including the data science life cycle, features and tools selection in Power BI, dashboard creation, data analysis, and findings.

3.1 Data Science Life Cycle

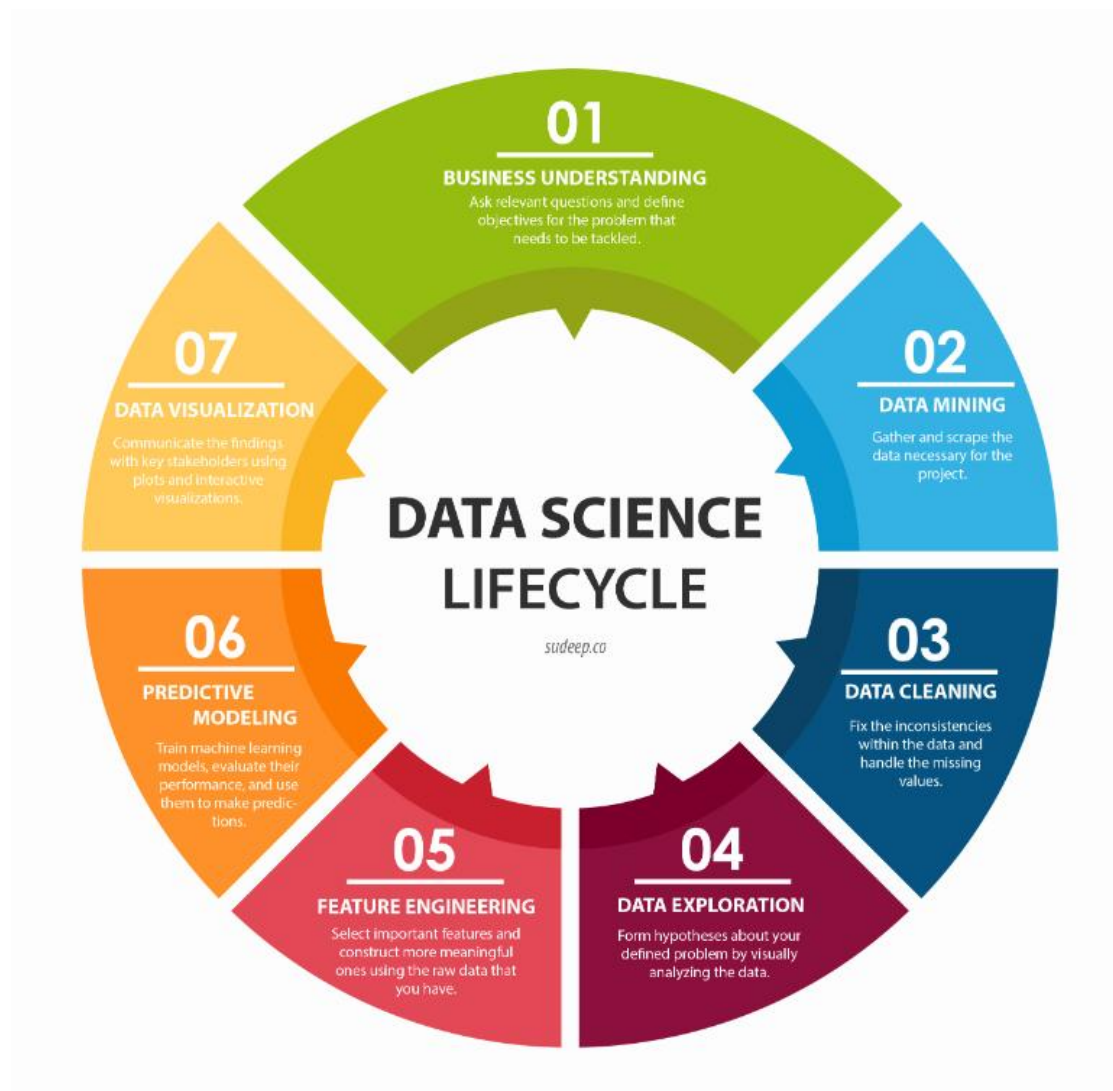


Figure 3.1 Data Science Life Cycle

According to [17], data science is becoming one of the most popular fields in the technology business quite quickly. It allow us to analyze large datasets thanks to the speedy improvements in computer efficiency, which have made it possible to find

previously unheard-of patterns and insights regarding user behavior and global trends. Below are the overview stages of data science life cycle provided by [17], which include business understanding, data mining, data cleaning, data exploration, feature engineering, predictive modeling and data visualization.

3.1.1 Business Understanding

Business understanding also known as identifying the objectives, is one important stage prior to perform a data science work where it is crucial to identify the problem to be solve. In general, the 5 types of question to be solve using data science are regression, classification, clustering, anomaly detection and recommendation types of questions. Being able to ask the appropriate questions to leverage the power of data to generate outcomes for your organization is more of an art than a science, and mastering this takes a great deal of expertise.

3.1.2 Data Mining

After identifying the problems and objectives, we can proceed to the process of compiling the data from different sources. Locating the appropriate data requires work and time. We have to look at several questions in this stage such as what, where and how to get the right data.

3.1.3 Data Cleaning

Data cleaning is one stage that needed the most time to work with after we have all the appropriate data. This stage include cleaning up the data and preparing the data. As your project data get larger, the time to perform this stage takes longer. The main problem in this stage is the inconsistency of data and is crucial to the project if it does not handle well. For instance, people tend to forgot to handle the missing value in the data that can cause a big error while creating the model, these missing value should be deleted or replace with average value.

3.1.4 Data Exploration

This stage is to analyse and understand the pattern of the data after we got a clean set of data. We use a random subset of data to perform visualization analysis and study the pattern and make assumption of how this results will compare and perform differently with the rest of data in the dataset.

3.1.5 Feature Engineering

A feature in machine learning is a quantifiable characteristic or quality of an observed phenomena. This stage shows the technique of turning your raw data into informative characteristics that describe the business challenge you're trying to address by applying domain expertise. The accuracy of the predictive model you build in the following step will be directly impacted by this stage. There is 2 types of feature engineering to be perform which are feature selection and feature construction. Feature selection is to eliminate the features that do not bring much information to the dataset that cause the curse of dimensionality. Feature construction is to generate new features from existing one.

3.1.6 Predictive Modeling

In this stage, we decide the model to fix the problem mentioned in business understanding stage. The number, types, and quality of your data, your willingness to commit time and computational resources, and the kind of output you hope to achieve will all affect the model you ultimately train. We perform testing with the right model to achieve a outcome that makes the most sense to the project.

3.1.7 Data Visualization

The reason data visualization is a challenging field is that, although it may look straightforward, mastering it may be one of the most difficult tasks. It includes the combination of communication, psychology, statistics and art to visualize efficient and effective data.

3.2 Power BI Architecture

Power BI Architecture

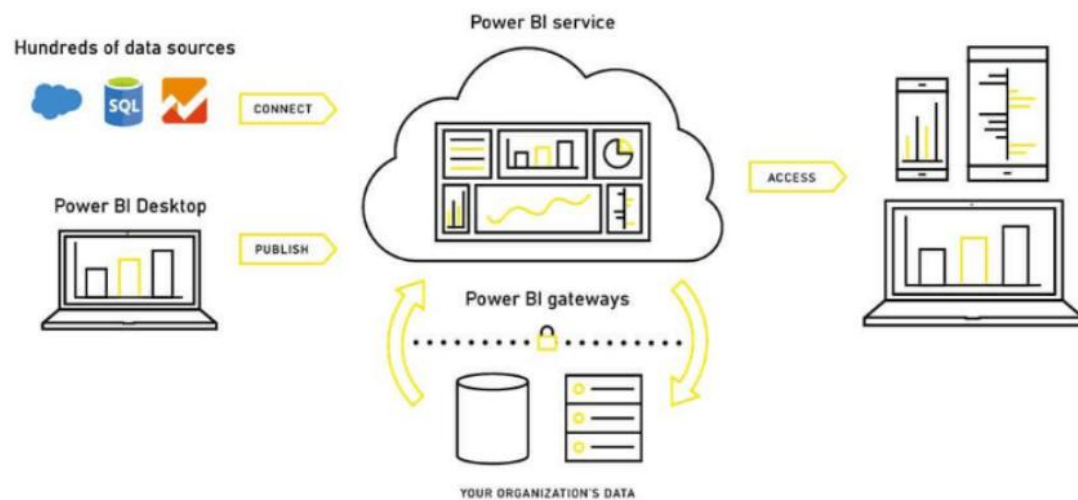


Figure 3.2 Architecture of Power BI [18]

According to Figure 3.2, it shows the architecture of Power BI which is a methodology to generate types of visualization using Power BI. During the first stage we will collect the sample data from data source for analysis preparation. We first identified the research domain which is the overall student attendance and grades from FICT. In the domain, there consist of different courses that students are undertaking.

We used Microsoft Excel to store the dataset. There consists of course code, course name, grades and attendance. We will look into the overall attendance and grades of courses from different semester. These information will be imported to Power BI to transform into interactive visualization and showing the relationship of variables, then perform analytics based on the visualization. This allowed educators and institutions to gain a more meaningful and clearer insights to make future prediction and adjustment for students.

In the second stage, a Power BI Desktop version is required. Comparing to the web version, since the visualization is still in developing stage, desktop version is much suitable to design work and create report while the web version is used after the full development is published into the Service and allow collaboration [19]. With using the Power BI Desktop, we can fully access its features and tools to develop the suitable visualization. Refer to Table 3.2, Table 3.3 and Table 3.4 [20].

Tools	Features
Import/Export Datasource	Allow user to import or export dataset (refer to table 3.3 for types of datasource)
Query Editor	Allow user to change and format the data before being loaded into Power BI Model
Views	Report View: create, edit and adjust visualization Data View: inspect, explore and analyze data Model View: show model relationship
Modelling Ribbon	A panel to upload datasets, make calculations, edit data types, format data category for a column, insert visualizations.
Field List	Contain elements of uploaded dataset.
Navigation Pane	A panel to control the Views.
Visualization Pane	A panel contains types of visualization (refer to Table 3.4)

Table 3.2.1 Power BI tools and features

Data source	Description
CSV	A list of data is contained in a plain text file. One or more values may appear in each row, each of which is delimited by a comma.
Excel	Manually enter data into a worksheet or queried and loaded from outside data sources.
SQL	Directly connect to model databases of SQL Server.
Cloud	Live connection to Cloud databases such as Azure.
Power BI	Allow query and loading of data from external data source, or manually enter data.

Table 3.2.2 Types of datasource in Power BI

CHAPTER 3

Stacked Bar/Column Charts	Tree Maps
Clustered Bar/Column Charts	Tables and Matrix
Line/Area Charts	Pie Charts
Ribbon Charts	Python-Scripts
Waterfall Charts	R-Scripts

Table 3.2.3 Some important visualization

CHAPTER 4

System Design

4.1 Data Science Life Cycle (DSLCL) Implementation

In this part of the chapter, we show the process of how the DSLCL implement into the project. We will cover all the seven stages of DSLCL and provide explanation on the implementation.

4.1.1 Business Understanding

In this stage, we will identify our objectives and find the problems we tend to solve with using DSLCL. The objectives of this project is to propose a new actionable dashboard with visual representations that show the relationship of student performance and their attendance and to use this tool to analyze and study the correlation between this relationship to make prediction and make future adjustment on how student attendance could improve the student performance. As Chapter 2 Literature Review show that there is not much previous work implementing the knowledge visualization that show the relationship between students' grades and attendance, and provide recommendation to improve students' performance. Therefore, the work on this project can resolve this issue to help educators and decision makers to have insights on this relationship.

4.1.2 Data Mining

In this stage, we will collect the students information based on the question of what, where and how to get the data. For the data we need to collect, since the project title is Knowledge Visualization of Student Performance Based on Attendance, we will only collect the students' information in terms of grades and attendance rate for the respective courses. Next we will identify the location to get these data, we will collect these data from the students from UTAR FICT. Finally, the way to achieve these data is that we will request the data from teachers in FICT, as students we have no permission to get data directly from the database of FICT. We will not include the

CHAPTER 4

personal information of students such as name and ID to protect the privacy of students.

UCCD1203	DATABASE DEVELOPMENT AND APPLICATIONS	C	44/48	91.7
UCCT1104	APPLICATION PROGRAMMING	B-	60/65	92.3
MPU32143	ENGLISH FOR INFORMATION TECHNOLOGY	B-	25/28	89.3
UCCN1004	DATA COMMUNICATIONS AND NETWORKING	B-	49/55	89.1
UBMM1011	SUN ZI'S ART OF WAR AND BUSINESS STRATEGIES	PS	14/14	100
UCCB1013	FUNDAMENTALS OF BUSINESS SYSTEMS	B+	37/43	86
UCCT1113	INFORMATION TECHNOLOGY INFRASTRUCTURE FOR BUSINESS	C+	38/50	76
UCCM2233	STATISTICS	B-	46/52	88.5
UCCD3253	ENTERPRISE SYSTEMS	B	47/50	94
UAMG1043	INTERPERSONAL COMMUNICATION	C+	31/33	93.9
UCCD1013	ANALYSIS AND DESIGN OF INFORMATION SYSTEMS	C	45.5/53.5	85
UCCD2043	INFORMATION TECHNOLOGY PROJECT MANAGEMENT	B	40/50	80
UCCD2502	INTRODUCTION TO INVENTIVE PROBLEM SOLVING AND PROPOSAL WRITING	B	26/32	81.2
UCCD2323	FRONT-END WEB DEVELOPMENT	C	34/40	85
UCCT2513	MINI PROJECT	B+	33/33	100
UCCB2333	SOCIAL MEDIA STRATEGIES FOR BUSINESS	B+	33/40	82.5
UCCB3583	PROJECT I	B-	0/0	0
UCCD3053	INFORMATION TECHNOLOGY PROFESSIONAL ETHICS	B-	45/47	95.7
MPU3123	TAMADUN ISLAM DAN TAMADUN ASIA (TITAS)	PS	33/42	78.6
UCCA3053	KNOWLEDGE MANAGEMENT	C+	43/52.5	81.9
UCCD3023	DIGITAL ENTREPRENEURSHIP	A-	42/45	93.3
UCCB3144	ADVANCED DIGITAL MEDIA TECHNOLOGY	B-	56/62	90.3
UCCB3133	BUSINESS ANALYTICS	C	42/47	89.4
UCCD3013	ECOMMERCE PRACTICES	B+	47/48	97.9
UCCD2213	SOFTWARE ENGINEERING PRINCIPLES	C+	42/46	91.3
MPU33013	MALAYSIAN ECONOMY	B	36/37	97.3
UCCB1104	INTRODUCTION TO BUSINESS PROGRAMMING	B	48/56	85.7
UCCD1004	PROGRAMMING CONCEPTS AND PRACTICES	C	48/58	82.8
UCCD1203	DATABASE DEVELOPMENT AND APPLICATIONS	C	34/44	77.3

CHAPTER 4

UCCD1013	ANALYSIS AND DESIGN OF INFORMATION SYSTEMS	C	39/48	81.2
UCCD2213	SOFTWARE ENGINEERING PRINCIPLES	C	37/46	80.4
UCCN1004	DATA COMMUNICATIONS AND NETWORKING	C+	32.5/39.5	82.3
UCCM2233	STATISTICS	C	36/43	83.7
UCCN1223	CYBERSECURITY	C	25/38	65.8
UCCD2043	INFORMATION TECHNOLOGY PROJECT MANAGEMENT	C	20/34	58.8
UCCB2333	SOCIAL MEDIA STRATEGIES FOR BUSINESS	C+	15/21	71.4
UCCD2003	OBJECT-ORIENTED SYSTEMS ANALYSIS AND DESIGN	A	34/38	89.5
UCCD2323	FRONT-END WEB DEVELOPMENT	A	46/46	100
UCCD2043	INFORMATION TECHNOLOGY PROJECT MANAGEMENT	A+	42/48	87.5
UCCD2044	OBJECT-ORIENTED PROGRAMMING PRACTICES	A	54/58	93.1
UBAF1043	BUSINESS ACCOUNTING	A+	38.5/45.5	84.6
UCCD2303	DATABASE TECHNOLOGY	A	44/48	91.7
UCCD2502	INTRODUCTION TO INVENTIVE PROBLEM SOLVING AND PROPOSAL WRITING	A-	30/32	93.8
UCCD2123	USER EXPERIENCE DESIGN	A-	34/42	81
MPU34032	COMMUNITY PROJECT	PS	22/22	100
UCCB2513	MINI PROJECT	A-	39/42	92.9
UCCB3133	BUSINESS ANALYTICS	A	48/49	98
UCCB3583	PROJECT I	B+	0/0	0
UCCD3013	ECOMMERCE PRACTICES	A	41/48	85.4
UCCB3224	DATA MINING TECHNIQUES	A+	50/58	86.2
UCCD3023	DIGITAL ENTREPRENEURSHIP	A	30/30	100
UCCT2103	FINANCIAL TECHNOLOGY	A	40/44	90.9
UCCD3053	INFORMATION TECHNOLOGY PROFESSIONAL ETHICS	A	44/46	95.7
UBAF1043	BUSINESS ACCOUNTING	B+	33.5/40	83.8
UCCD1213	FUNDAMENTALS OF DIGITAL MEDIA TECHNOLOGY	B	41/48	85.4
MPU32143	ENGLISH FOR INFORMATION TECHNOLOGY	B+	28/30	93.3
UCCM2233	STATISTICS	B	45/53	84.9
UBMM1011	SUN ZI'S ART OF WAR AND BUSINESS STRATEGIES	PS	10-Dec	83.3
UAMG1043	INTERPERSONAL COMMUNICATION	B+	32/35	91.4
UCCB1013	FUNDAMENTALS OF BUSINESS SYSTEMS	B	42/43	97.7
UCCB1104	INTRODUCTION TO BUSINESS PROGRAMMING	B-	50/56	89.3

CHAPTER 4

UCCD1203	DATABASE DEVELOPMENT AND APPLICATIONS	B	40/44	90.9
UCCD1013	ANALYSIS AND DESIGN OF INFORMATION SYSTEMS	B	46/48	95.8
UCCD1004	PROGRAMMING CONCEPTS AND PRACTICES	B	51/58	87.9
MPU3153	PENGHAYATAN ETIKA DAN PERADABAN (FOR LOCAL STUDENTS)	PS	39/42	92.9
UCCN1004	DATA COMMUNICATIONS AND NETWORKING	A	41.5/45.5	91.2
UCCD3053	INFORMATION TECHNOLOGY PROFESSIONAL ETHICS	B	43/46	93.5
UCCN1223	CYBERSECURITY	B+	39/45	86.7
UCCD1013	ANALYSIS AND DESIGN OF INFORMATION SYSTEMS	C	53.5/53.5	100
UCCN1004	DATA COMMUNICATIONS AND NETWORKING	F	36/38	94.7
UCCM2233	STATISTICS	C	53/53	100
UCCD1203	DATABASE DEVELOPMENT AND APPLICATIONS	B-	47/47	100
UCCB1013	FUNDAMENTALS OF BUSINESS SYSTEMS	C+	50/50	100
UBMM1011	SUN ZI'S ART OF WAR AND BUSINESS STRATEGIES	PS	14/14	100
UCCD3053	INFORMATION TECHNOLOGY PROFESSIONAL ETHICS	C	40/44	90.9
MPU33013	MALAYSIAN ECONOMY	B+	42/42	100
UBMM2013	OPERATIONS MANAGEMENT	B-	37/37	100
UALJ2013	INTRODUCTION TO JAPANESE	PS	40/42	95.2
UCCN1004	DATA COMMUNICATIONS AND NETWORKING	A-	46/49	93.9
UBAF1043	BUSINESS ACCOUNTING	B-	46/46	100
MPU3183	FALSAFAH DAN ISU SEMASA	PS	39/39	100
UCCM1363	DISCRETE MATHEMATICS	B-	44/49	89.8
UCCD1004	PROGRAMMING CONCEPTS AND PRACTICES	B+	57/57	100
UCCD1013	ANALYSIS AND DESIGN OF INFORMATION SYSTEMS	C+	54.5/54.5	100
UCCB1013	FUNDAMENTALS OF BUSINESS SYSTEMS	C+	41/43	95.3
UCCD1213	FUNDAMENTALS OF DIGITAL MEDIA TECHNOLOGY	C+	52/52	100
UCCD2303	DATABASE TECHNOLOGY	C+	44/48	91.7
UCCD2044	OBJECT-ORIENTED PROGRAMMING PRACTICES	B-	58/58	100
UCCD2043	INFORMATION TECHNOLOGY PROJECT MANAGEMENT	C+	46/48	95.8
UCCB2333	SOCIAL MEDIA STRATEGIES FOR BUSINESS	C+	42/47	89.4

CHAPTER 4

UCCD2003	OBJECT-ORIENTED SYSTEMS ANALYSIS AND DESIGN	B-	36/38	94.7
UCCD2323	FRONT-END WEB DEVELOPMENT	C+	46/46	100
MPU34032	COMMUNITY PROJECT	PS	22/22	100
UCCD2502	INTRODUCTION TO INVENTIVE PROBLEM SOLVING AND PROPOSAL WRITING	B+	30/30	100
UCCB2513	MINI PROJECT	B	36/39	92.3
UCCB3133	BUSINESS ANALYTICS	B-	39/39	100
UCCB1104	INTRODUCTION TO BUSINESS PROGRAMMING	A-	53/56	94.6
UCCD3223	MOBILE APPLICATIONS DEVELOPMENT	C+	42.5/46.5	91.4
UCCD3013	ECOMMERCE PRACTICES	B+	39/45	86.7
MPU3123	TAMADUN ISLAM DAN TAMADUN ASIA (TITAS)	PS	39/42	92.9
UCCA3053	KNOWLEDGE MANAGEMENT	B+	46/52.5	87.6
UCCD3053	INFORMATION TECHNOLOGY PROFESSIONAL ETHICS	B	42/47	89.4
UCCB3224	DATA MINING TECHNIQUES	B	51/58	87.9
UCCD3023	DIGITAL ENTREPRENEURSHIP	A-	31/31	100
UBMH1013	ORGANISATION AND HUMAN RESOURCE	B-	39.5/44	89.8
UCCB3583	PROJECT I	B	0/0	0
UCCB3113	INFORMATION TECHNOLOGY CONTROL AND AUDIT	B	45/46	97.8
MPU32143	ENGLISH FOR INFORMATION TECHNOLOGY	B-	33/33	100
UCCB1104	INTRODUCTION TO BUSINESS PROGRAMMING	A	0/0	0
UCCD1203	DATABASE DEVELOPMENT AND APPLICATIONS	A-	0/0	0
UCCM2233	STATISTICS	A-	51/52	98.1
UCCN1004	DATA COMMUNICATIONS AND NETWORKING	B+	0/0	0
UCCD1213	FUNDAMENTALS OF DIGITAL MEDIA TECHNOLOGY	B	46/52	88.5
UCCM1363	DISCRETE MATHEMATICS	B+	43/49	87.8
UBMM1011	SUN ZI'S ART OF WAR AND BUSINESS STRATEGIES	PS	14/14	100
UCCB1013	FUNDAMENTALS OF BUSINESS SYSTEMS	B+	42/43	97.7
UCCD1004	PROGRAMMING CONCEPTS AND PRACTICES	A	57/57	100
UCCD1013	ANALYSIS AND DESIGN OF INFORMATION SYSTEMS	A-	51.5/54.5	94.5
UCCD2003	OBJECT-ORIENTED SYSTEMS ANALYSIS AND DESIGN	A-	33/37	89.2
UCCD2044	OBJECT-ORIENTED PROGRAMMING PRACTICES	A	54/58	93.1

CHAPTER 4

UCCD2043	INFORMATION TECHNOLOGY PROJECT MANAGEMENT	B+	45/48	93.8
UCCB2113	BUSINESS INFORMATION MANAGEMENT	B+	34/36	94.4
UCCD2323	FRONT-END WEB DEVELOPMENT	A-	42/46	91.3
UCCD2303	DATABASE TECHNOLOGY	B+	44/48	91.7
UCCB1223	RAPID APPLICATION DEVELOPMENT	A	38/46	82.6
UCCB2513	MINI PROJECT	A-	39/42	92.9
MPU34032	COMMUNITY PROJECT	PS	20/22	90.9
UCCD2502	INTRODUCTION TO INVENTIVE PROBLEM SOLVING AND PROPOSAL WRITING	A-	28/30	93.3
UCCB3224	DATA MINING TECHNIQUES	A	58/58	100
UALF1003	INTRODUCTION TO FRENCH	PS	32/41	78
UCCD3023	DIGITAL ENTREPRENEURSHIP	B+	30/30	100
UCCD3013	ECOMMERCE PRACTICES	A	45/48	93.8
UCCD3223	MOBILE APPLICATIONS DEVELOPMENT	B+	42/48	87.5
UCCD3053	INFORMATION TECHNOLOGY PROFESSIONAL ETHICS	A	44/46	95.7
UBAF1043	BUSINESS ACCOUNTING	A	40.5/46	88
UCCM1153	INTRODUCTION TO CALCULUS AND APPLICATIONS	B	54/54	100
MPU32143	ENGLISH FOR INFORMATION TECHNOLOGY	B-	34.5/35.5	97.2
UCCE1024	CIRCUIT THEORY AND BASIC ELECTRONICS	A	0/0	0
UCCD2123	USER EXPERIENCE DESIGN	B	0/0	0
MPU3183	FALSAFAH DAN ISU SEMASA	PS	36/36	100
UCCN1004	DATA COMMUNICATIONS AND NETWORKING	A	38/38	100
UCCD1203	DATABASE DEVELOPMENT AND APPLICATIONS	A	48/48	100
UCCB1104	INTRODUCTION TO BUSINESS PROGRAMMING	A	50.5/50.5	100
UCCM2233	STATISTICS	A-	48/48	100
UBMM2023	ORGANISATIONAL BEHAVIOUR	B-	43/43	100
UBAF1043	BUSINESS ACCOUNTING	A	41.5/41.5	100
UCCN1223	CYBERSECURITY	A	42/42	100
UCCB1013	FUNDAMENTALS OF BUSINESS SYSTEMS	B	40/43	93
UCCD2044	OBJECT-ORIENTED PROGRAMMING PRACTICES	A+	52/55	94.5
UCCM1363	DISCRETE MATHEMATICS	A	44/45	97.8
UCCD1213	FUNDAMENTALS OF DIGITAL MEDIA TECHNOLOGY	A-	44/48	91.7
UCCD1013	ANALYSIS AND DESIGN OF INFORMATION SYSTEMS	A	45/47	95.7
UCCB2333	SOCIAL MEDIA STRATEGIES FOR BUSINESS	B+	39/41	95.1

CHAPTER 4

UCCD2003	OBJECT-ORIENTED SYSTEMS ANALYSIS AND DESIGN	A	46/46	100
UCCD2043	INFORMATION TECHNOLOGY PROJECT MANAGEMENT	A-	50/50	100
UCCD2303	DATABASE TECHNOLOGY	A-	46/46	100
UCCB3224	DATA MINING TECHNIQUES	A	57/58	98.3
UCCD2323	FRONT-END WEB DEVELOPMENT	B	44/44	100
UCCB2513	MINI PROJECT	B+	39/42	92.9
UCCD2502	INTRODUCTION TO INVENTIVE PROBLEM SOLVING AND PROPOSAL WRITING	A-	30/34	88.2
UAMG1043	INTERPERSONAL COMMUNICATION	B	36/40	90
MPU32143	ENGLISH FOR INFORMATION TECHNOLOGY	B	33/33	100
MPU33013	MALAYSIAN ECONOMY	A-	41/41	100
UCCD3053	INFORMATION TECHNOLOGY PROFESSIONAL ETHICS	B-	37/39	94.9
UCCB1013	FUNDAMENTALS OF BUSINESS SYSTEMS	A-	42/43	97.7
UCCM2233	STATISTICS	C	50/54	92.6
UCCN1004	DATA COMMUNICATIONS AND NETWORKING	B-	50/51	98
UCCD1203	DATABASE DEVELOPMENT AND APPLICATIONS	C	42/44	95.5
UCCN1223	CYBERSECURITY	C+	40/45	88.9
UCCB3133	BUSINESS ANALYTICS	W	0/0	0
UCCT1104	APPLICATION PROGRAMMING	B-	56/56	100
UCCD3253	ENTERPRISE SYSTEMS	C+	45/47	95.7
UCCT1113	INFORMATION TECHNOLOGY INFRASTRUCTURE FOR BUSINESS	B+	44/48	91.7
UCCB2333	SOCIAL MEDIA STRATEGIES FOR BUSINESS	C	19/21	90.5
MPU3183	FALSAFAH DAN ISU SEMASA	PS	33/36	91.7
UBMM2023	ORGANISATIONAL BEHAVIOUR	B-	47.5/47.5	100
UCCD1013	ANALYSIS AND DESIGN OF INFORMATION SYSTEMS	B+	54.5/54.5	100
UALB1003	INTRODUCTION TO GERMAN LANGUAGE	PS	36/36	100
UCCD1213	FUNDAMENTALS OF DIGITAL MEDIA TECHNOLOGY	A-	52/52	100
UCCD1004	PROGRAMMING CONCEPTS AND PRACTICES	A-	57/57	100
UCCM1363	DISCRETE MATHEMATICS	B	44/49	89.8
UBMM1011	SUN ZI'S ART OF WAR AND BUSINESS STRATEGIES	PS	14/14	100
UCCB1013	FUNDAMENTALS OF BUSINESS SYSTEMS	B+	43/43	100
UCCD2323	FRONT-END WEB DEVELOPMENT	B+	46/46	100
UCCD2003	OBJECT-ORIENTED SYSTEMS ANALYSIS AND DESIGN	A-	35/38	92.1

CHAPTER 4

UCCD2044	OBJECT-ORIENTED PROGRAMMING PRACTICES	A-	60/58	103.4
UCCD2303	DATABASE TECHNOLOGY	A-	48/48	100
UCCB2333	SOCIAL MEDIA STRATEGIES FOR BUSINESS	A-	46/47	97.9
UCCD2043	INFORMATION TECHNOLOGY PROJECT MANAGEMENT	A	48/48	100
UCCD2502	INTRODUCTION TO INVENTIVE PROBLEM SOLVING AND PROPOSAL WRITING	A-	30/32	93.8
UCCD2123	USER EXPERIENCE DESIGN	B+	40/42	95.2
MPU34032	COMMUNITY PROJECT	PS	22/22	100
UCCB2513	MINI PROJECT	B	42/42	100
UCCD3013	ECOMMERCE PRACTICES	B+	45/48	93.8
UCCD3023	DIGITAL ENTREPRENEURSHIP	B	30/30	100
UCCB3224	DATA MINING TECHNIQUES	A-	58/58	100
UCCB3133	BUSINESS ANALYTICS	A-	47/49	95.9
UBAF1043	BUSINESS ACCOUNTING	B	46/46	100
UCCD3053	INFORMATION TECHNOLOGY PROFESSIONAL ETHICS	B-	46/46	100
UCCD1213	FUNDAMENTALS OF DIGITAL MEDIA TECHNOLOGY	C+	37/37	100
UBAF1043	BUSINESS ACCOUNTING	B	35/35	100
MPU32143	ENGLISH FOR INFORMATION TECHNOLOGY	B+	34/34	100
UCCM2233	STATISTICS	B+	43/45	95.6
UCCB1104	INTRODUCTION TO BUSINESS PROGRAMMING	B	55/57	96.5
UBMM1011	SUN ZI'S ART OF WAR AND BUSINESS STRATEGIES	PS	14/14	100
UCCM1363	DISCRETE MATHEMATICS	B-	46/47	97.9
UCCB1013	FUNDAMENTALS OF BUSINESS SYSTEMS	B	25/25	100
UCCD1203	DATABASE DEVELOPMENT AND APPLICATIONS	B-	16/16	100
UCCD2043	INFORMATION TECHNOLOGY PROJECT MANAGEMENT	C+	42/47	89.4
UCCD2003	OBJECT-ORIENTED SYSTEMS ANALYSIS AND DESIGN	F	39/46	84.8
UCCB2333	SOCIAL MEDIA STRATEGIES FOR BUSINESS	B+	42/55	76.4
UCCD2303	DATABASE TECHNOLOGY	F	38/48	79.2
UCCD2323	FRONT-END WEB DEVELOPMENT	C+	39/49	79.6
UCCD2044	OBJECT-ORIENTED PROGRAMMING PRACTICES	F	47/56	83.9
UCCN1223	CYBERSECURITY	B+	32/36	88.9
UCCB2513	MINI PROJECT	B	42/42	100
UCCD2502	INTRODUCTION TO INVENTIVE PROBLEM SOLVING AND PROPOSAL WRITING	A-	32/36	88.9

CHAPTER 4

UCCA3053	KNOWLEDGE MANAGEMENT	B+	41/52.5	78.1
MPU34152	LEADERSHIP AND TEAMBUILDING	PS	20/20	100
UCCD2003	OBJECT-ORIENTED SYSTEMS ANALYSIS AND DESIGN	C+	42/46	91.3
UBMM1013	MANAGEMENT PRINCIPLES	B-	39/42	92.9
MPU33013	MALAYSIAN ECONOMY	A-	30/34	88.2
UCCD2303	DATABASE TECHNOLOGY	A-	42/46	91.3
UCCB3583	PROJECT I	B-	0/0	0
UCCD2044	OBJECT-ORIENTED PROGRAMMING PRACTICES	C+	49/55	89.1
UCCB3224	DATA MINING TECHNIQUES	B+	45/58	77.6
UCCB3133	BUSINESS ANALYTICS	F	40/49	81.6
UCCB1223	RAPID APPLICATION DEVELOPMENT	B-	46/46	100
UCCB1223	RAPID APPLICATION DEVELOPMENT	A-	38/40	95
UCCB1223	RAPID APPLICATION DEVELOPMENT	A	40/40	100
UCCB1223	RAPID APPLICATION DEVELOPMENT	B+	44/46	95.7
UCCB1223	RAPID APPLICATION DEVELOPMENT	A-	34/40	85
UCCB1223	RAPID APPLICATION DEVELOPMENT	A-	34/40	85
UCCB1223	RAPID APPLICATION DEVELOPMENT	B+	46/46	100
UCCB1223	RAPID APPLICATION DEVELOPMENT	A+	40/46	87
UCCB1223	RAPID APPLICATION DEVELOPMENT	B+	44/46	95.7
UCCB3133	BUSINESS ANALYTICS	C	41/47	87.2
UCCB3133	BUSINESS ANALYTICS	C+	41/47	87.2
UCCB3133	BUSINESS ANALYTICS	A+	47/47	100
UCCB3133	BUSINESS ANALYTICS	B-	38/47	80.9
UCCB3133	BUSINESS ANALYTICS	B+	44/47	93.6
UCCB3133	BUSINESS ANALYTICS	B+	35/41	85.4
UCCB3133	BUSINESS ANALYTICS	B-	39/47	83
UCCB3133	BUSINESS ANALYTICS	B-	34/40	85

Table 4.1.2 Raw Students Data Information

4.1.3 Data Cleaning

In this stage, we will clean up the data from the raw data we have collected previously. The purpose of this stage is to make the data look tidy by rearranging and grouping and eliminating the duplicates value, null value, inappropriate value and outliers. Below is the data after cleanup that highlight with different color.

Yellow Highlight: Inappropriate data

Blue Highlight: Insufficient data in groups (Min. 5 data per group)

CHAPTER 4

Red Highlight: Missing value or null value

UCCD1203	DATABASE DEVELOPMENT AND APPLICATIONS	C	44/48	91.7
		C	34/44	77.3
		B	40/44	90.9
		B-	47/47	100
		A	48/48	100
		C	42/44	95.5
		B-	16/16	100
MPU32143	ENGLISH FOR INFORMATION TECHNOLOGY	B-	25/28	89.3
		B+	28/30	93.3
		B-	33/33	100
		B-	34.5/35.5	97.2
		B	33/33	100
		B+	34/34	100
		B-	49/55	89.1
UCCN1004	DATA COMMUNICATIONS AND NETWORKING	C+	32.5/39.5	82.3
		A	41.5/45.5	91.2
		F	36/38	94.7
		A-	46/49	93.9
		A	38/38	100
		B-	50/51	98
		PS	14/14	100
UBMM1011	SUN ZI'S ART OF WAR AND BUSINESS STRATEGIES	PS	10/12	83.3
		PS	14/14	100
		PS	14/14	100
		PS	14/14	100
		PS	14/14	100
		PS	14/14	100
UCCB1013	FUNDAMENTALS OF BUSINESS SYSTEMS	B+	37/43	86
		B	42/43	97.7
		C+	50/50	100
		C+	41/43	95.3
		B+	42/43	97.7
		B	40/43	93
		A-	42/43	97.7
		B+	43/43	100
		B	25/25	100
UCCM2233	STATISTICS	B-	46/52	88.5
		C	36/43	83.7
		B	45/53	84.9
		C	53/53	100
		A-	51/52	98.1
		A-	48/48	100

CHAPTER 4

		C	50/54	92.6
		B+	43/45	95.6
UCCD1013	ANALYSIS AND DESIGN OF INFORMATION SYSTEMS	C	45.5/53.5	85
		C	39/48	81.2
		B	46/48	95.8
		C	53.5/53.5	100
		C+	54.5/54.5	100
		A-	51.5/54.5	94.5
		A	45/47	95.7
		B+	54.5/54.5	100
		B	40/50	80
UCCD2043	INFORMATION TECHNOLOGY PROJECT MANAGEMENT	C	20/34	58.8
		A+	42/48	87.5
		C+	46/48	95.8
		B+	45/48	93.8
		A-	50/50	100
		A	48/48	100
		C+	42/47	89.4
		B	26/32	81.2
UCCD2502	INTRODUCTION TO INVENTIVE PROBLEM SOLVING AND PROPOSAL WRITING	A-	30/32	93.8
		B+	30/30	100
		A-	28/30	93.3
		A-	30/34	88.2
		A-	30/32	93.8
		A-	32/36	88.9
		C	34/40	85
UCCD2323	FRONT-END WEB DEVELOPMENT	A	46/46	100
		C+	46/46	100
		A-	42/46	91.3
		B	44/44	100
		B+	46/46	100
		C+	39/49	79.6
		B+	33/33	100
UCCT2513/UCCB2513	MINI PROJECT	A-	39/42	92.9
		B	36/39	92.3
		A-	39/42	92.9
		B+	39/42	92.9
		B	42/42	100
		B	42/42	100
		B+	33/40	82.5
UCCB2333	SOCIAL MEDIA STRATEGIES FOR BUSINESS	C+	15/21	71.4
		C+	42/47	89.4
		B+	39/41	95.1

CHAPTER 4

		C	19/21	90.5
		A-	46/47	97.9
		B+	42/55	76.4
UCCD3053	INFORMATION TECHNOLOGY PROFESSIONAL ETHICS	B-	45/47	95.7
		A	44/46	95.7
		B	43/46	93.5
		C	40/44	90.9
		B	42/47	89.4
		A	44/46	95.7
		B-	37/39	94.9
		B-	46/46	100
UCCD3023	DIGITAL ENTREPRENEURSHIP	A-	42/45	93.3
		A	30/30	100
		A-	31/31	100
		B+	30/30	100
		B	30/30	100
UCCB3133	BUSINESS ANALYTICS	C	42/47	89.4
		A	48/49	98
		B-	39/39	100
		A-	47/49	95.9
		F	40/49	81.6
		C	41/47	87.2
		C+	41/47	87.2
		A+	47/47	100
		B-	38/47	80.9
		B+	44/47	93.6
		B+	35/41	85.4
		B-	39/47	83
		B-	34/40	85
UCCD3013	ECOMMERCE PRACTICES	B+	47/48	97.9
		A	41/48	85.4
		B+	39/45	86.7
		A	45/48	93.8
		B+	45/48	93.8
UCCB1104	INTRODUCTION TO BUSINESS PROGRAMMING	B	48/56	85.7
		B-	50/56	89.3
		A-	53/56	94.6
		A	50.5/50.5	100
		B	55/57	96.5
UCCD1004	PROGRAMMING CONCEPTS AND PRACTICES	C	48/58	82.8
		B	51/58	87.9
		B+	57/57	100
		A	57/57	100

CHAPTER 4

		A-	57/57	100
UCCN1223	CYBERSECURITY	C	25/38	65.8
		B+	39/45	86.7
		A	42/42	100
		C+	40/45	88.9
		B+	32/36	88.9
UCCD2003	OBJECT-ORIENTED SYSTEMS ANALYSIS AND DESIGN	A	34/38	89.5
		B-	36/38	94.7
		A-	33/37	89.2
		A	46/46	100
		A-	35/38	92.1
		F	39/46	84.8
		C+	42/46	91.3
UCCD2044	OBJECT-ORIENTED PROGRAMMING PRACTICES	A	54/58	93.1
		B-	58/58	100
		A	54/58	93.1
		A+	52/55	94.5
		A-	60/58	103.4
		F	47/56	83.9
		C+	49/55	89.1
UBAF1043	BUSINESS ACCOUNTING	A+	38.5/45.5	84.6
		B+	33.5/40	83.8
		B-	46/46	100
		A	40.5/46	88
		A	41.5/41.5	100
		B	46/46	100
		B	35/35	100
UCCD2303	DATABASE TECHNOLOGY	A	44/48	91.7
		C+	44/48	91.7
		B+	44/48	91.7
		A-	46/46	100
		A-	48/48	100
		F	38/48	79.2
		A-	42/46	91.3
UCCB3224	DATA MINING TECHNIQUES	A+	50/58	86.2
		B	51/58	87.9
		A	58/58	100
		A	57/58	98.3
		A-	58/58	100
		B+	45/58	77.6
UCCD1213	FUNDAMENTALS OF DIGITAL MEDIA TECHNOLOGY	B	41/48	85.4
		C+	52/52	100
		B	46/52	88.5

CHAPTER 4

		A-	44/48	91.7
		A-	52/52	100
		C+	37/37	100
UCCM1363	DISCRETE MATHEMATICS	B-	44/49	89.8
		B+	43/49	87.8
		A	44/45	97.8
		B	44/49	89.8
		B-	46/47	97.9
UCCB1223	RAPID APPLICATION DEVELOPMENT	A	38/46	82.6
		B-	46/46	100
		A-	38/40	95
		A	40/40	100
		B+	44/46	95.7
		A-	34/40	85
		A-	34/40	85
		B+	46/46	100
		A+	40/46	87
		B+	44/46	95.7
UCCT1113	INFORMATION TECHNOLOGY INFRASTRUCTURE FOR BUSINESS	C+	38/50	76
		B+	44/48	91.7
UCCD3253	ENTERPRISE SYSTEMS	B	47/50	94
		C+	45/47	95.7
UAMG1043	INTERPERSONAL COMMUNICATION	C+	31/33	93.9
		B+	32/35	91.4
		B	36/40	90
MPU3123	TAMADUN ISLAM DAN TAMADUN ASIA (TITAS)	PS	33/42	78.6
		PS	39/42	92.9
UCCA3053	KNOWLEDGE MANAGEMENT	C+	43/52.5	81.9
		B+	46/52.5	87.6
		B+	41/52.5	78.1
UCCB3144	ADVANCED DIGITAL MEDIA TECHNOLOGY	B-	56/62	90.3
UCCD2213	SOFTWARE ENGINEERING PRINCIPLES	C+	42/46	91.3
		C	37/46	80.4
MPU33013	MALAYSIAN ECONOMY	B	36/37	97.3
		B+	42/42	100
		A-	41/41	100
		A-	30/34	88.2
UCCD2123	USER EXPERIENCE DESIGN	A-	34/42	81
		B+	40/42	95.2
MPU34032	COMMUNITY PROJECT	PS	22/22	100
		PS	22/22	100
		PS	20/22	90.9
UCCT2103	FINANCIAL TECHNOLOGY	A	40/44	90.9

CHAPTER 4

MPU3153	PENGHAYATAN ETIKA DAN PERADABAN (FOR LOCAL STUDENTS)	PS	39/42	92.9
UBMM2013	OPERATIONS MANAGEMENT	B-	37/37	100
UALJ2013	INTRODUCTION TO JAPANESE	PS	40/42	95.2
MPU3183	FALSAFAH DAN ISU SEMASA	PS	39/39	100
		PS	36/36	100
		PS	33/36	91.7
UCCD3223	MOBILE APPLICATIONS DEVELOPMENT	C+	42.5/46.5	91.4
		B+	42/48	87.5
UBMH1013	ORGANISATION AND HUMAN RESOURCE	B-	39.5/44	89.8
UCCB3113	INFORMATION TECHNOLOGY CONTROL AND AUDIT	B	45/46	97.8
UCCB2113	BUSINESS INFORMATION MANAGEMENT	B+	34/36	94.4
UCCT1104	APPLICATION PROGRAMMING	B-	60/65	92.3
		B-	56/56	100
UCCM1153	INTRODUCTION TO CALCULUS AND APPLICATIONS	B	54/54	100
UCCE1024	CIRCUIT THEORY AND BASIC ELECTRONICS	A	0/0	0
UBMM2023	ORGANISATIONAL BEHAVIOUR	B-	43/43	100
		B-	47.5/47.5	100
UALB1003	INTRODUCTION TO GERMAN LANGUAGE	PS	36/36	100
MPU34152	LEADERSHIP AND TEAMBUILDING	PS	20/20	100
UBMM1013	MANAGEMENT PRINCIPLES	B-	39/42	92.9
UCCB3583	PROJECT I	B-	0/0	0
UCCB3133	BUSINESS ANALYTICS	W	0/0	0
UCCD2123	USER EXPERIENCE DESIGN	B	0/0	0
UCCB1104	INTRODUCTION TO BUSINESS PROGRAMMING	A	0/0	0
UCCD1203	DATABASE DEVELOPMENT AND APPLICATIONS	A-	0/0	0
UCCN1004	DATA COMMUNICATIONS AND NETWORKING	B+	0/0	0
UCCB3583	PROJECT I	B	0/0	0
UCCB3583	PROJECT I	B+	0/0	0
UCCB3583	PROJECT I	B-	0/0	0

Table 4.1.3 Students Data Cleaning Process with Highlight

4.1.4 Data Exploration

In this stage, we will use a subset from the data from the data to perform visualization to do testing, and we will analyse, study the pattern and make assumption for others data reaction on the same visualization.

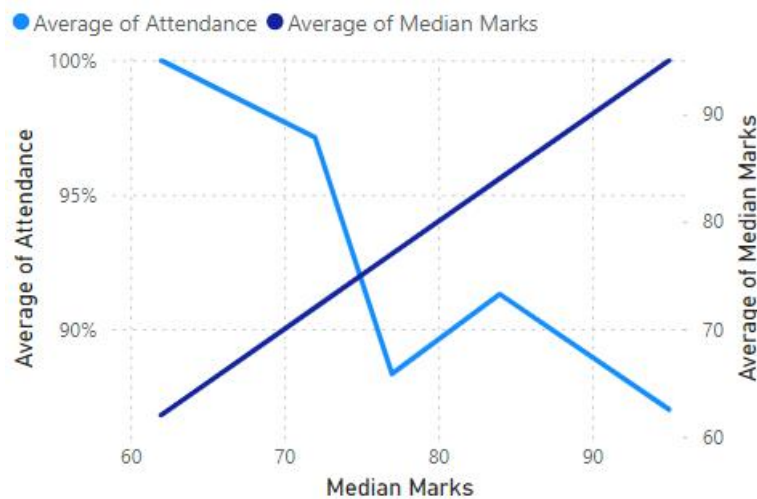


Figure 4.1.4.1 Visualization from subset A

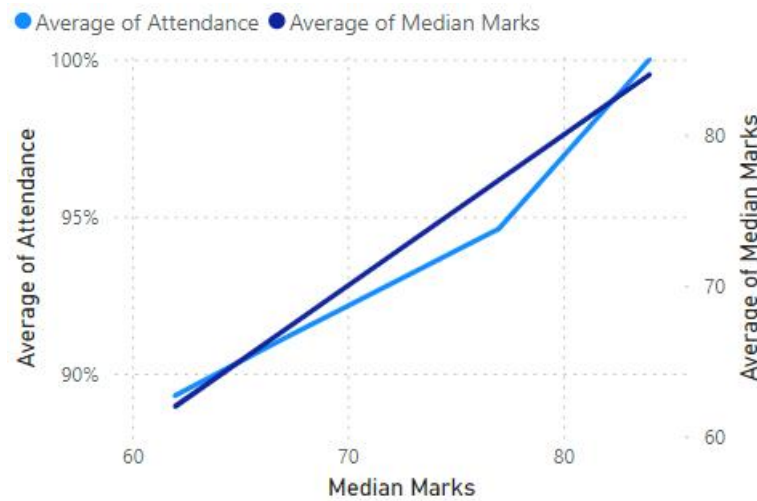


Figure 4.1.4.2 Visualization from subset B

From the 2 visualization of above figure, we can conclude that some of the courses may show no correlation between students grades and attendance such as the pattern show in subset A, and some of the courses may show positive correlation between students grades and attendance such as pattern show in subset B.

4.1.5 Feature Engineering

In this stage, we will turn the raw data into some informative feature. We will go through 2 types of feature engineering which are feature selection where we eliminate

CHAPTER 4

features that do not bring any meaning to the visualization results, and feature construction where we create a new feature from existing one.

UCCD1203	DATABASE DEVELOPMENT AND APPLICATIONS	C	44/48	91.7
		C	34/44	77.3
		B	40/44	90.9
		B-	47/47	100
		A	48/48	100
		C	42/44	95.5
		B-	16/16	100

Table 4.1.5.1 Feature selection

Course Code	Course Name	Grade	Median Marks	Average Marks	Average Grade	Average GPA	Attendance	Average Attendance
UCCD1203	DATABASE DEVELOPMENT AND APPLICATIONS	C	52	63	B-	2.6700	91.70%	96.35%
		B	67				90.90%	
		B-	62				100.00%	
		A	85				100.00%	
		C	52				95.50%	
		B-	62				100.00%	
Course Grade Prediction based on Attendance Rate			Recommendation on Grades					
<p>If $x \geq 96.35\%$, your possible course grade will be B-, B, B+, A-, A, or A+</p> <p>If $x \leq 96.35\%$, your possible course grade will be B-, C+, C, or F</p> <p>If $x < 80.00\%$, your course grade will be F</p>			<p>There is no correlation between the course marks and attendance.</p> <p>If students wish to achieve a grade above average, students should:</p> <ul style="list-style-type: none"> - focus on self practice since attendance will not directly affect the grades. - understand all the database vocabulary and concepts. - be able to develop an ERD. - be able to construct SQL statements. 					

Table 4.1.5.2 Feature Construction

4.1.6 Predictive Modeling

In this stage, we identify the model to perform prediction on the student performance based on attendance. In this project, we use the average of students attendance of a course and average of students marks of a course to perform prediction.

$$A = \frac{1}{n} \sum_{i=1}^n a_i$$

where

A = arithmetic mean

n = number of values

CHAPTER 4

a_i = data set values

We insert this average formula into Excel by typing “=AVERAGE(number1,...)” and the Excel will automatically calculate the average of the range of numbers.

4.1.7 Data Visualization

In this final stage, we will import all the data into Power BI to perform data visualization in dashboard form. We will create 3 page of dashboard that represent descriptive page, predictive page and prescriptive page respectively. The explanation of the visualization dashboard will be explain in Chapter 5 System Implementation.

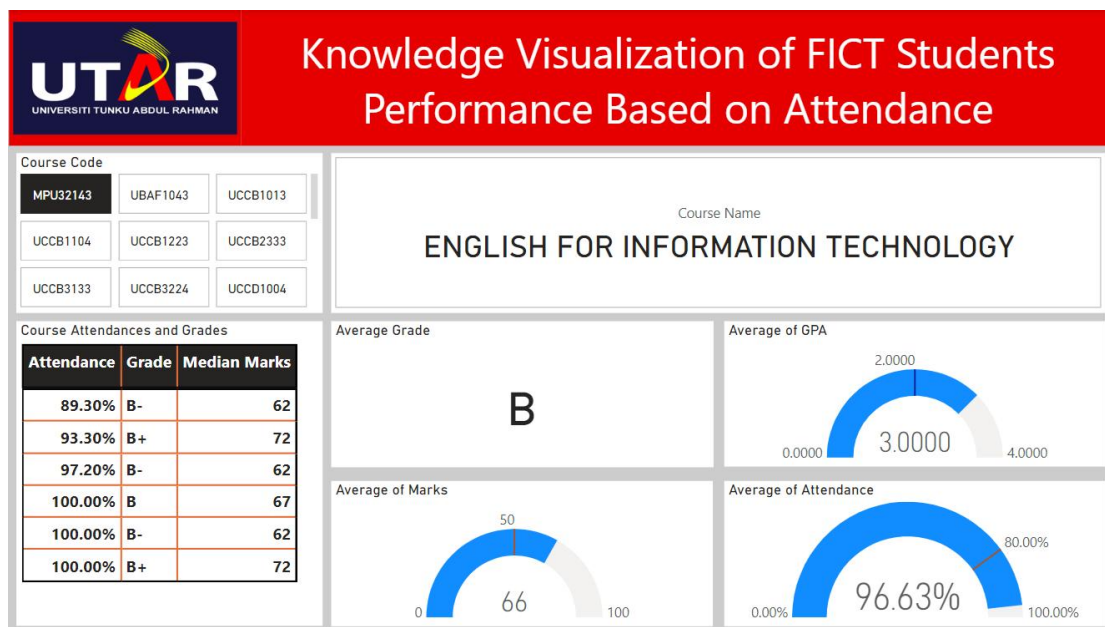


Figure 4.1.7.1 Visualization of Descriptive Page

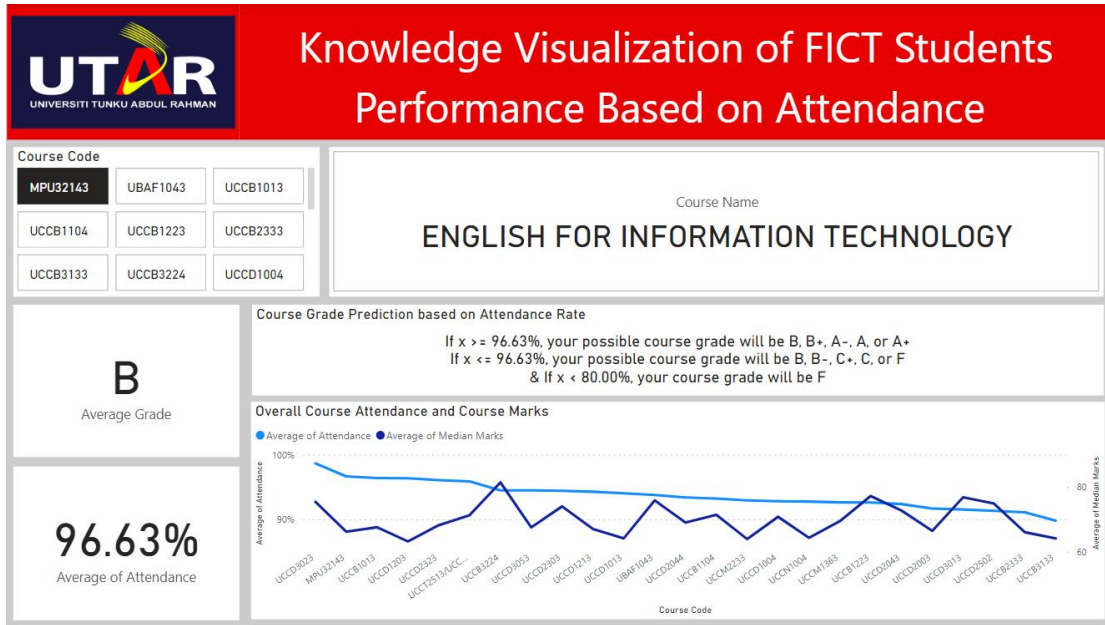


Figure 4.1.7.2 Visualization of Predictive Page

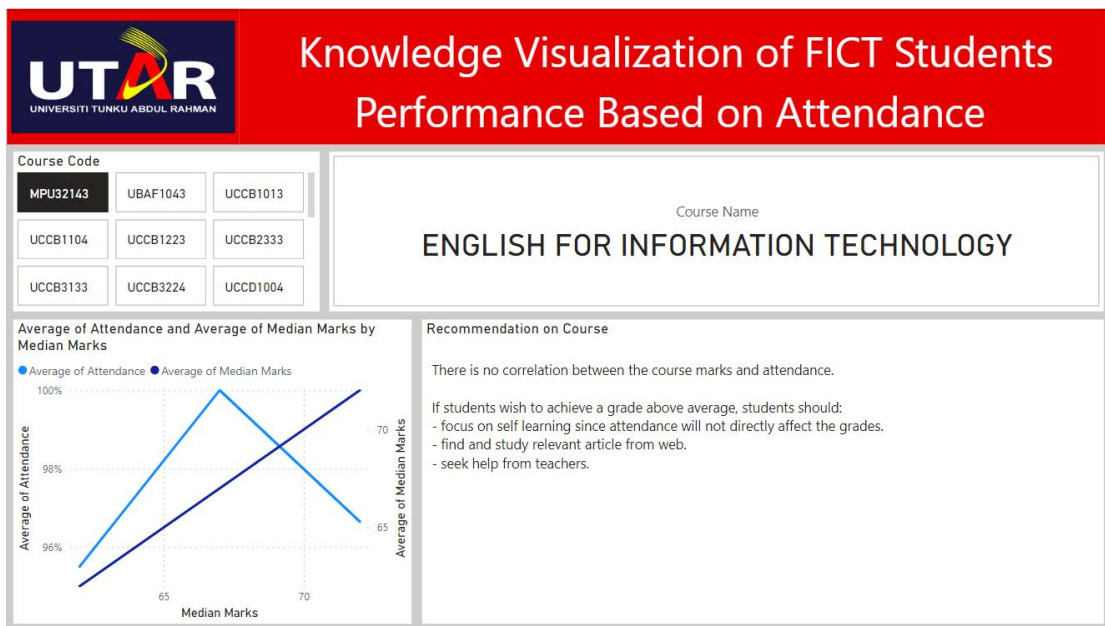


Figure 4.1.7.3 Visualization of Prescriptive Page

4.2 System Flowchart

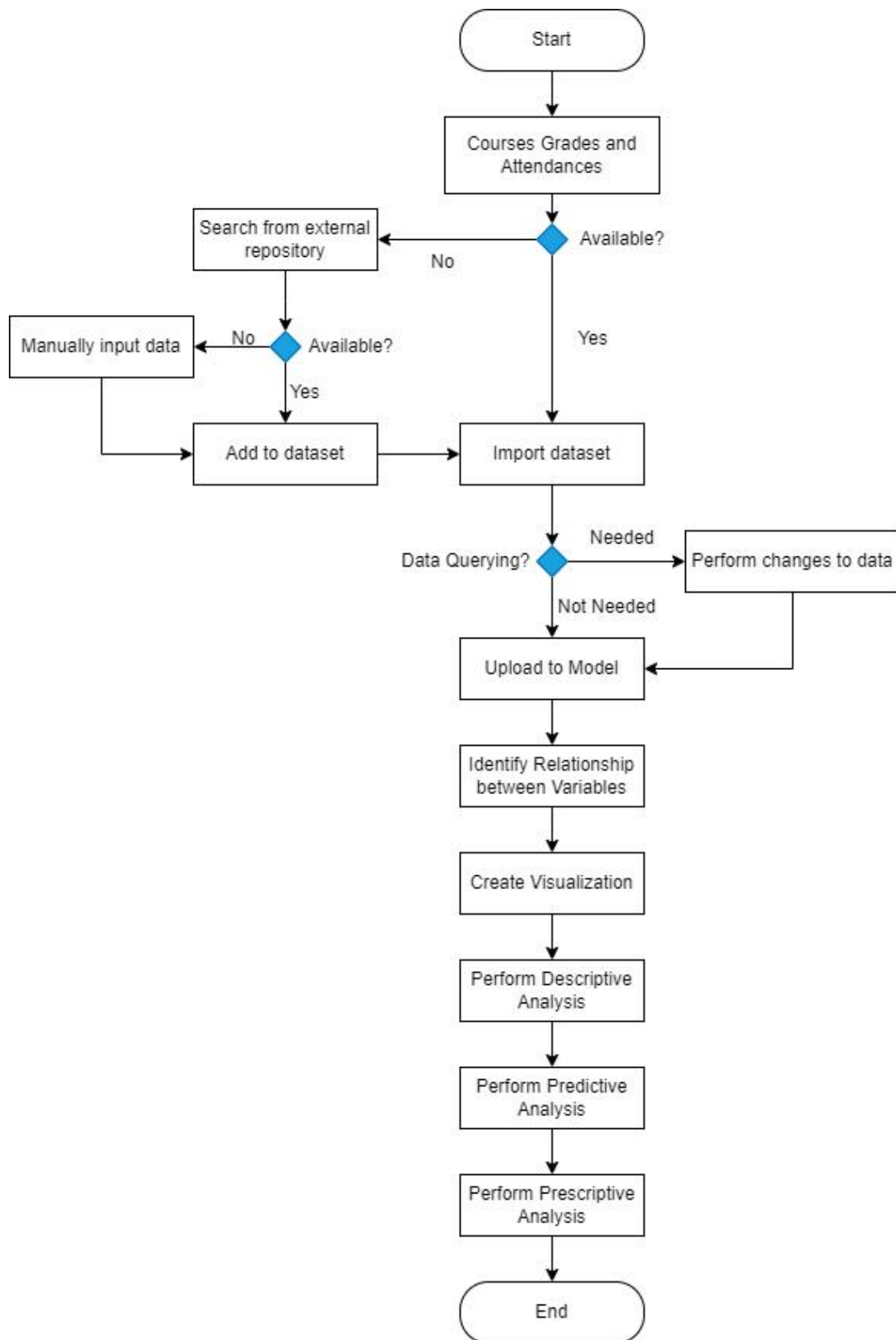


Figure 4.2 Flowchart of Power BI

4.3 System Block Diagram

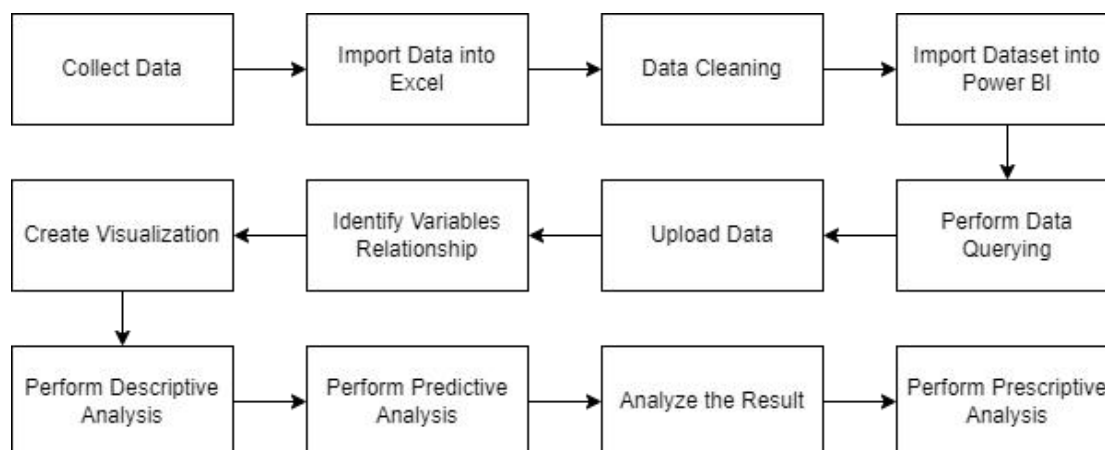


Figure 4.3 Block Diagram

According to Figure 4.5, in order to perform knowledge visualization of student performance based on attendance, firstly we need to collect data from UTAR students. The target data is from FICT department. The data includes to be collected are the course code, course name, overall attendance rate and overall grades for the respective course. Once we collected enough data, we import the data into Excel Spreadsheets to have a more easier view to the data. Next, we proceed to data cleaning which include the process of identifying and fixing those missing values or eliminate inappropriate data. This process is critical that it ensure the accuracy of data. The next step is importing the dataset into Power BI. This can be done in Power BI desktop application. Before the dataset upload into the model, we need identify whether the dataset need to perform data querying. This process will use the Query Editor in Power BI to change and format the data then to be upload into the model. Then we need to identify the relationship between each variables and make only relevant and useful information based on the relationship. Once the relationship between variables has been clarify, we can start create visualization by dragging the variables into the types of visualization. We need to ensure the relationship is show in the correct representation of visualization in this process. After all relevant visualization has been created, we can arrange it in the dashboard to have a more tidy look. This enable the ease of understanding in the next step which is analyzing the visualization. Then we can start to perform the descriptive analysis based on the visualization created that describe the visualization. Based on the descriptive analysis, we can perform the next

CHAPTER 4

step which is predictive analysis to do prediction on future results based on previous grades and attendances pattern. Furthermore, we analyse the pattern of the prediction that can allow us to proceed to the last step which is prescriptive analysis that provide recommendation to the students accordingly to the courses.

CHAPTER 5

System Implementation

5.1 Hardware

The hardware involved in this project is a laptop. The laptop is used to develop the knowledge visualization of students' performance based on attendance and perform data analysis with Power BI and Excel. The specification of the laptop to develop the project is as shown as below:

Description	Specifications
Model	Asus TUF Gaming FX505GD
Processor	Intel(R) Core(TM) i5-8300H CPU @ 2.30GHz
Operating System	Windows 11 Home Single Language
Graphic	NVIDIA GeForce GTX 1050
Memory	12 GB RAM
Storage	477 GB SSD

Table 5.1 Specifications of laptop

5.2 Software

5.2.1 Power BI



Figure 5.2.1 Microsoft Power BI

The main software to develop this project is Microsoft Power BI. It allow us to develop the dashboard of knowledge visualization of students' performance based on their attendance. Power BI is very suitable to create data visualization due to it able to streamline publication and distribution capabilities. Dashboard in Power Bi is chosen as the visualization to present the results and prediction of students' performance to educators and decision maker. It can display real-time update of data and visuals to allow us to analyze the information changes [21]. About Power BI dashboard, it is a single-page documents where it contain types of graphs and charts. We will present our results and prediction of students' performance with graphs and charts. The purpose we use Power BI dashboard is it allow us, educators and decision maker to analyze the data and observe each important metrics in a same page. It also allow us to develop visualization using different datasets and reports and present in graphs and charts [22].

5.2.2 Microsoft Excel Spreadsheet



Figure 5.2.2 Microsoft Excel Spreadsheet

Microsoft Excel is another software included to help developing this project. Microsoft Excel is compatible with Power BI [23] and is one of the its datasources. It allows us to store the dataset needed for Power BI to be imported to perform visualization using the data in the dataset.

5.3 Dashboard Operation

In this sub-topic, we will show the operation of the knowledge visualization dashboard of students performance based on attendance step by step. We will cover from the data cleaning in Excel Sheet and Power Query Editor in Power BI, until the design of the dashboard.

First of all, we will import the raw data into the Excel Sheet to perform data cleaning and feature engineering where we eliminate null values and inappropriate values, rearrange and group the data to make it more tidy, and add new features for later dashboard design purposes.

UNIVERSITY	DATABASE DEVELOPMENT AND APPLICATIONS	C	50-54	52	03	B	2,000	91.25%	96.25%	If $x > 96.25%$, your possible course grade will be S, B+, A, A-, or A. If $x < 96.25%$, your possible course grade will be B, C+, C, or F.	There is no correlation between the course marks and attendance. If students wish to achieve a grade above average, students should focus on all practice since attendance will not directly affect the grade achieved at the database vocabulary and concepts. It is able to transfer an SQL. It is able to construct SQL statements.
		B	55-59	57				90.00%			
		B-	50-54	52				100.00%			
		A	50-59	55				100.00%			
		C	50-54	52				92.50%			
		B-	50-54	52				100.00%			
UNIVERSITY	ENGINEERING INFORMATION TECHNOLOGY	B	50-54	52	06	B	3,000	89.33%	96.67%	If $x > 96.67%$, your possible course grade will be S, B+, A, A-, or A. If $x < 96.67%$, your possible course grade will be B, C+, C, or F.	There is no correlation between the course marks and attendance. If students wish to achieve a grade above average, students should focus on all learning since attendance will not directly affect the grade. It is not able to transfer an SQL. It is not able to help from transfer.
		B+	70-74	72				93.33%			
		B-	50-54	52				100.00%			
		B	50-54	52				97.22%			
		B	55-59	57				100.00%			
		B+	70-74	72				100.00%			

Table 5.3.1 Example of completed data in Excel Sheet

CHAPTER 5

After we get the completed version of data, we can proceed to the next step where we import the Excel data into Power BI.

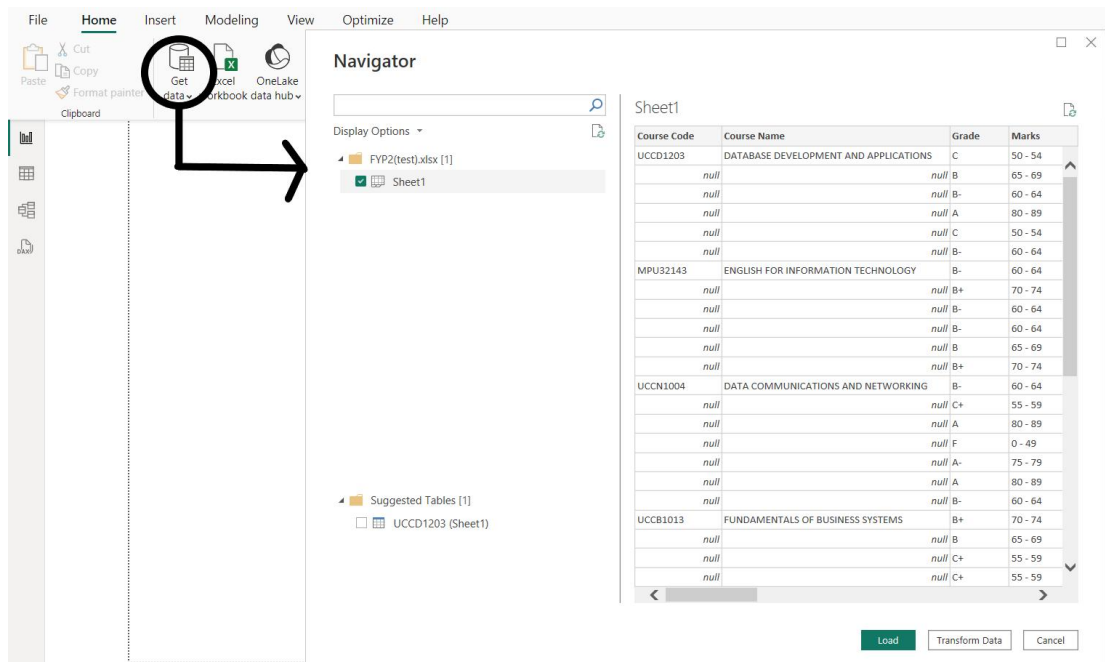


Figure 5.3.1 Import data from Excel to Power BI

Next, as show in Figure 5.3.1, there appear to be some null value in the data as Power BI could not directly import merged data into the project from Excel. In order to fix this, we will click the “Transform Data” button, and Power BI will take us to Power Query Editor to replace those null value and change the data types.

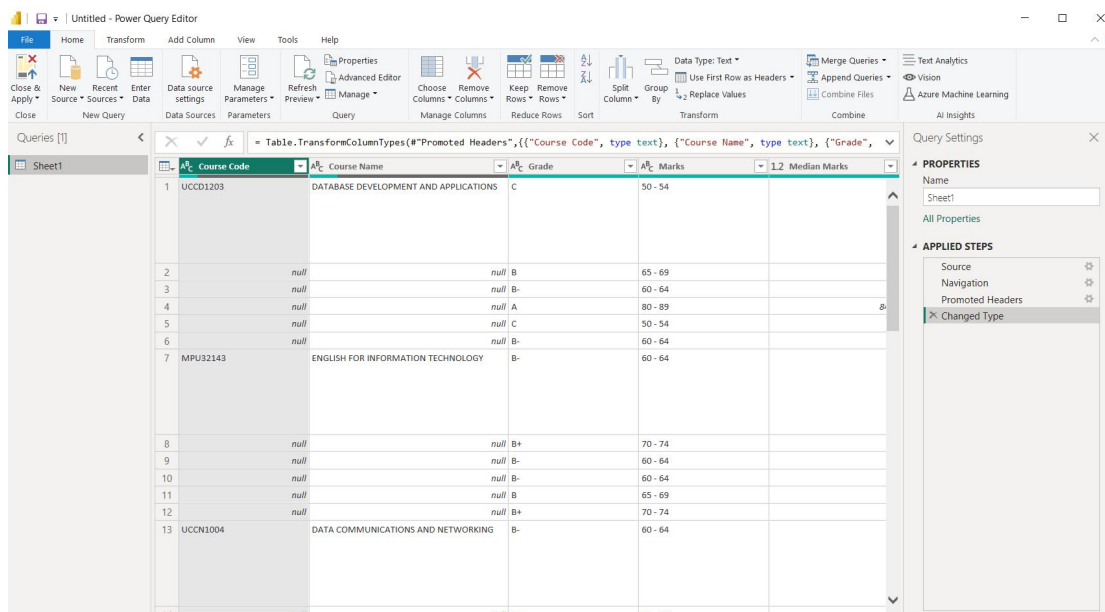


Figure 5.3.2 Data in Power Query Editor before editing

CHAPTER 5

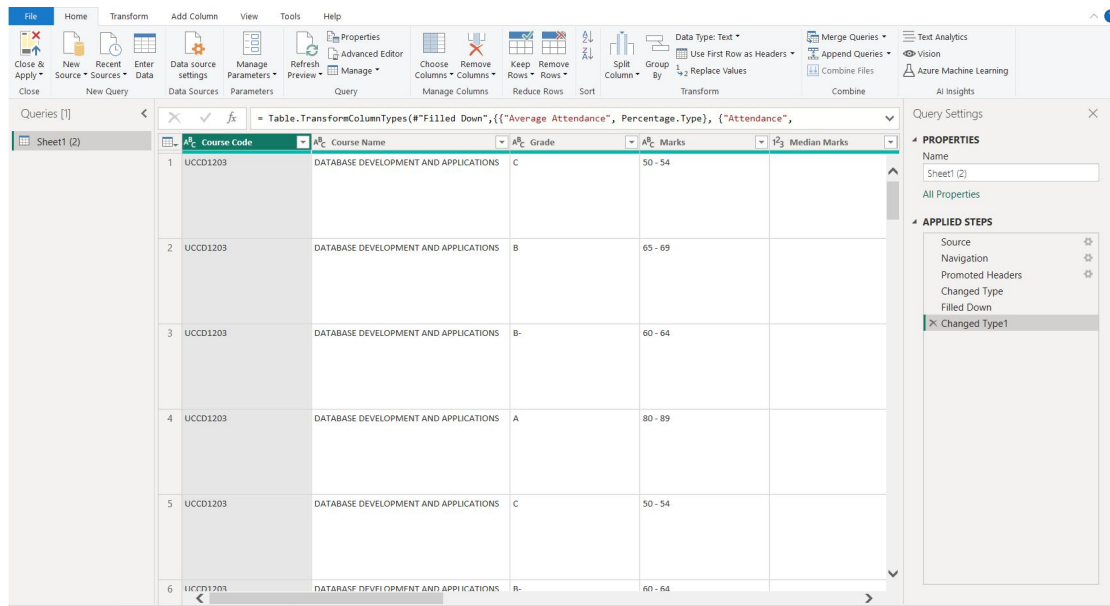


Figure 5.3.3 Data in Power Query Editor after editing

After we have edit the data using Power Query Editor, we will apply this data onto the project. The data will be show in the Table view in the project.

Course Code	Course Name	Grade	Marks	Median Marks	Average Marks	Average Grade	Average GPA	Attendance	Average
UCDD1203	DATABASE DEVELOPMENT AND APPLICATIONS	C	50 - 54	52	63 B-	2.6700	91.70%		
UCDD1203	DATABASE DEVELOPMENT AND APPLICATIONS	B	65 - 69	67	63 B-	2.6700	90.90%		
UCDD1203	DATABASE DEVELOPMENT AND APPLICATIONS	B-	60 - 64	62	63 B-	2.6700	100.00%		
UCDD1203	DATABASE DEVELOPMENT AND APPLICATIONS	A	80 - 89	84	63 B-	2.6700	100.00%		
UCDD1203	DATABASE DEVELOPMENT AND APPLICATIONS	C	50 - 54	52	63 B-	2.6700	95.50%		
UCDD1203	DATABASE DEVELOPMENT AND APPLICATIONS	B-	60 - 64	62	63 B-	2.6700	100.00%		
MPU32143	ENGLISH FOR INFORMATION TECHNOLOGY	B-	60 - 64	62	66 B	3.0000	89.30%		
MPU32143	ENGLISH FOR INFORMATION TECHNOLOGY	B+	70 - 74	72	66 B	3.0000	92.30%		
MPU32143	ENGLISH FOR INFORMATION TECHNOLOGY	B-	60 - 64	62	66 B	3.0000	100.00%		
MPU32143	ENGLISH FOR INFORMATION TECHNOLOGY	B-	60 - 64	62	66 B	3.0000	97.20%		
MPU32143	ENGLISH FOR INFORMATION TECHNOLOGY	B	65 - 69	67	66 B	3.0000	100.00%		
MPU32143	ENGLISH FOR INFORMATION TECHNOLOGY	B+	70 - 74	72	66 B	3.0000	100.00%		
UCCN1004	DATA COMMUNICATIONS AND NETWORKING	B-	60 - 64	62	62 B-	2.6700	89.10%		
UCCN1004	DATA COMMUNICATIONS AND NETWORKING	C+	55 - 59	57	62 B-	2.6700	82.30%		
UCCN1004	DATA COMMUNICATIONS AND NETWORKING	A	80 - 89	84	62 B-	2.6700	91.20%		
UCCN1004	DATA COMMUNICATIONS AND NETWORKING	F	0 - 49	24	62 B-	2.6700	94.70%		
UCCN1004	DATA COMMUNICATIONS AND NETWORKING	A-	75 - 79	77	62 B-	2.6700	93.90%		
UCCN1004	DATA COMMUNICATIONS AND NETWORKING	A	80 - 89	84	62 B-	2.6700	100.00%		
UCCN1004	DATA COMMUNICATIONS AND NETWORKING	B-	60 - 64	62	62 B-	2.6700	98.00%		
UCCB1013	FUNDAMENTALS OF BUSINESS SYSTEMS	B+	70 - 74	72	66 B	3.0000	86.00%		
UCCB1013	FUNDAMENTALS OF BUSINESS SYSTEMS	B	65 - 69	67	66 B	3.0000	97.70%		
UCCB1013	FUNDAMENTALS OF BUSINESS SYSTEMS	C+	55 - 59	57	66 B	3.0000	100.00%		
UCCB1013	FUNDAMENTALS OF BUSINESS SYSTEMS	C+	55 - 59	57	66 B	3.0000	95.30%		
UCCB1013	FUNDAMENTALS OF BUSINESS SYSTEMS	B+	70 - 74	72	66 B	3.0000	97.70%		
UCCB1013	FUNDAMENTALS OF BUSINESS SYSTEMS	B	65 - 69	67	66 B	3.0000	93.00%		
UCCB1013	FUNDAMENTALS OF BUSINESS SYSTEMS	A-	75 - 79	77	66 B	3.0000	97.70%		
UCCB1013	FUNDAMENTALS OF BUSINESS SYSTEMS	B+	70 - 74	72	66 B	3.0000	100.00%		

Figure 5.3.4 Table view of data

Above the Table view option is the Report view section, this is where we design the visualization using the data from the table. In the visualization, we will create 3 pages where it show the descriptive page, predictive page and prescriptive page respectively. We will break into each page with detail information on the types of visualization selection.

A) Descriptive Page

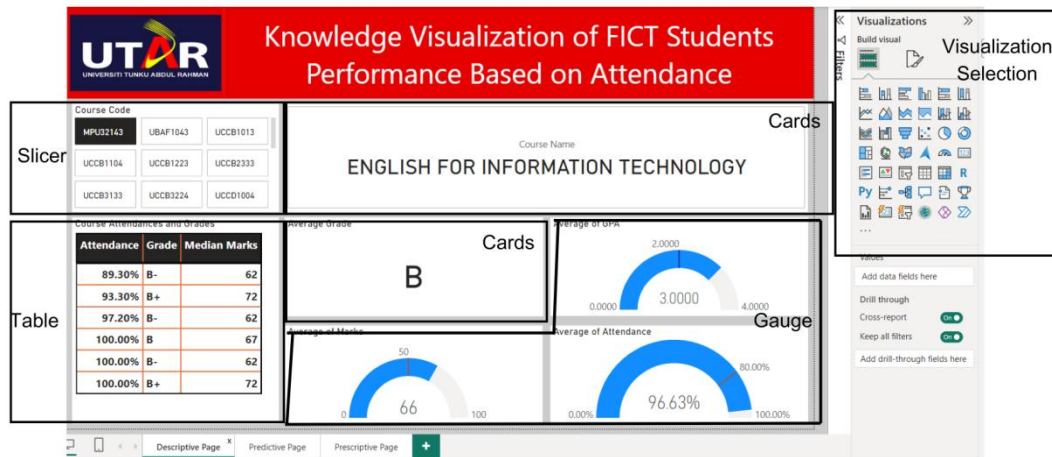


Figure 5.3.5 Descriptive Page

In this page, we use multiple visualization to describe the data. First is the Slicer used for Course Code, this visualization allow us to select a specific course to view its data, when a course is selected, all the relevant data will be show accordingly to the course. Next is the Table that show the Attendance, Grade and Median Marks of a specific course. For the Course Name and Average Grade, we use Cards as the visualization as it only show letters information. Lastly, we use the Gauge visualization to show the Average of GPA, Average of Marks and Average of Attendance. In the gauge, we able to set the minimum value, maximum value and target value which is the red line in the gauge that indicate the minimum requirement to pass the course. On the right side show the Visualization Selection that provides all kind of visualization option for different types of data.

B) Predictive Page

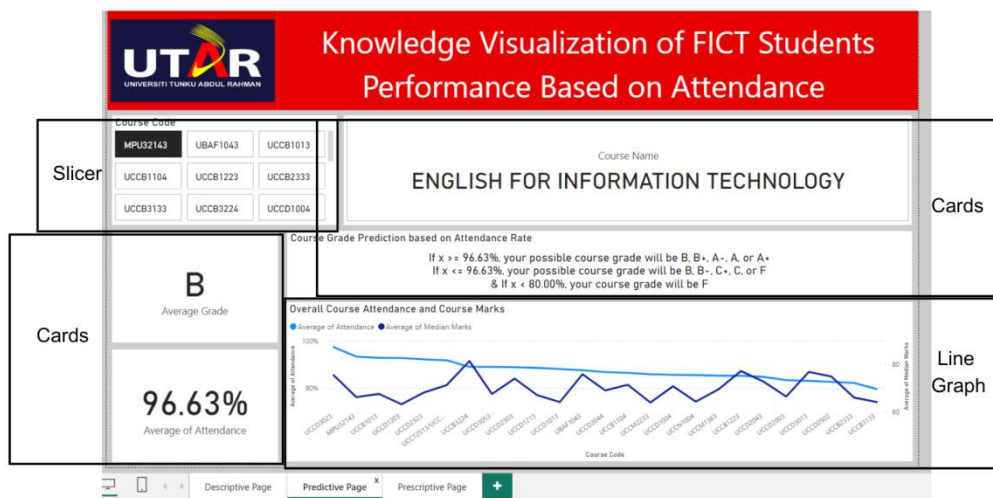


Figure 5.3.6 Predictive Page

For Predictive Page, we remain some of the visualization from the Descriptive Page which are the Slicer that show Course Code and Cards that show Course Name to let user easy navigate through each courses. We add the Average Grade and Average Attendance using Cards to show the relationship of students attendance and grade for each courses. We also use Cards to present the Course Grade Prediction based on Attendance Rate that provide estimated course grade for a specific course based on students current attendance rate. Below that, we use Line Graph to show the pattern of overall Course Attendance and Course Marks of all courses that allow us to forecast future students' grade based on attendance.

C) Prescriptive Page

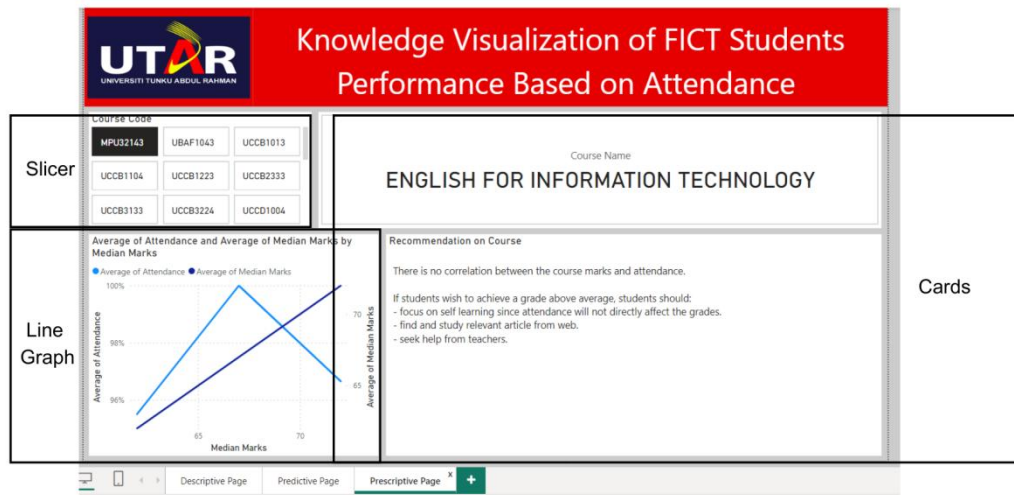


Figure 5.3.7 Prescriptive Page

In this Prescriptive Page, we also remain the Slicer showing Course Code and Cards showing Course Name. We also use the Line Graph to show the relationship and pattern of respective course of students' attendance and grades to find whether there is correlation in this course. Based on the Line Graph, we able to use Cards to provide the Recommendation to decision maker and educational institution regarding students current studies of different courses to achieve good performance and avoid failing the course.

5.4 Implementation Issues and Challenges

Throughout the implementation of this project, multiple challenges have been hindering the process of development. Firstly is the data quantity is insufficient to perform prediction on future grades based on attendance. This will affect the accuracy of prediction. Data quality and integration is one of the issues because different format and sources of data is collected during the data collection process and caused data quality issues. This need to be fix in data cleaning process to ensure the dashboard accuracy. Another issues is the process of data transformation is very time consuming since the data need to be in the proper format, so it had to go through several checking. For instance, in Power BI, some features is very complex to use such as the Data Query Editor. Some visualization also not always work like it suppose to present. In the process of developing the visualization, most time were

CHAPTER 5

consumed in trying to understand how a tools or features work properly especially for new users.

5.5 Timeline

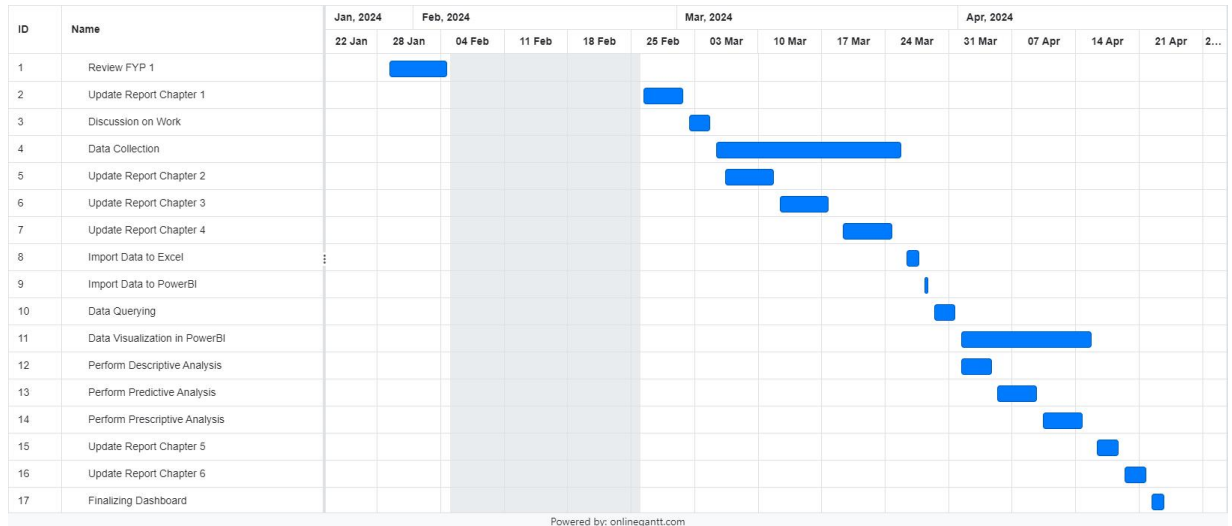


Figure 5.5 Gantt Chart of Project

Chapter 6

Findings

In this chapter, we will look into the findings of the knowledge visualization performed on Power BI dashboard. We will cover all 3 types of analysis which are descriptive, predictive and prescriptive analysis.

6.1 Descriptive Analysis

In the descriptive analysis, the dashboard contain some interactive visualization such as Course Code selection using a Slicer, Course Name display using Cards, Course Attendance, Grade and Marks in a form of Table, Average Grade display using Cards, and use of Gauge to show the Average of GPA, Average of Marks and Average of Attendance. All these visualization will changed accordingly to the selection of Course Code.

For the key findings in descriptive analysis, the dashboard uses many types of visualization that show the information of attendance, grades, marks and GPA for each courses. This allow user to quickly achieve the relevant information that best describe the courses' data such as user able to see how much the students performance are affected by their attendance rate, whether the course show good performance with high attendance, good performance with low attendance, poor performance with low attendance or poor performance with high attendance. The use of gauge is also provide a quick insights that show the range between the average GPA, Marks and Attendance with the minimum requirement to pass the courses. This allow educational institution to make decision on which course to improve that able to increase the students grades and attendances.

The main purpose of this descriptive analysis is to provide the overall view of data of each courses to user, decision maker or educational institution that allow them to have a quick insights of how the students performing based on different courses. By gaining advantages with these insights, educational institutions can adapt different stratagies to each courses to improve students engagement and performance.

6.2 Predictive Analysis

In the predictive analysis, the dashboard utilize the Slicer for Course Code selection, Cards that show the Course Name, Average Grade, Average Attendance and Course Grade Prediction based on Average Attendance Rate and Average Grade of each course, and using a Line Graph to show the Overall Course Attendance and Course Marks of all courses.

For the key findings, the dashboard focus more on the relationship between the grade and attendance where this is the main factor for forecast the students performance. However, due to the insufficient of data for each courses, we were unable to construct a more accurate prediction on students performance based on attendance. So we can only provide the estimated grade range based on previous average attendance and average grade. To improve the accuracy, we need to collect more data from more previous records in order to find the accurate pattern of students performance based on attendance for each courses. Furthermore, in the Line Graph, we able to found that the average attendance rate of each courses is around 80% to 100%, this is a good sign that most of the courses able to keep the students engagement above 80% which is the minimum rate to avoid failing the courses. However, the grades pattern on the Line Graph is inconsistent due to multiple factors such as how is the students assignment grades? Are students paying attention in class? Are the courses content to hard for students to study at the level? Therefore, to forecast the students performance only based on attendance is not enough to identify the outcome.

6.3 Prescriptive Analysis

In this prescriptive analysis, the dashbaord provide actionable recommendation on course using Cards, where this recommendation is based on the correlation show on the Line Graph of Average Attendance and Average Marks, and the Course Name using Slicer, these visualization will change accordingly to the selection of Course Code in the Slicer.

The key findings of the prescriptive analysis is that we can find that most of the course does not have correlation between the course attendance and grades. This mean that the students attendance will not directly affect the performance of students

CHAPTER 6

of the course. Despite all that, there is still some courses show positive correlation between students attendance and grades, which mean the students performance on the course may have impact from their engagement towards the course. Therefore, based on this correlation, we able to provide the recommendation to students whether if the course show no correlation, students should focus more on self learning while if the course show positive correlation, students should focus more on engaging in the class. The recommendation for each courses is based on the course learning outcome that guide the students to achieve it in order to improve their performance.

This prescriptive analysis show a very simple visualization because the main goal of this is to only provide recommendation for course improvement for decision maker and educational institutions. By this, they can enhance students performance by adopting the recommendation given from the prescriptive analysis.

CHAPTER 7

Conclusion and Future Works

7.1 Conclusion

As a conclusion, the purpose of this initiative is to address the important role that student attendance plays in academic success. Many educational institutions use antiquated attendance tracking techniques and don't have access to real-time data. This study aims to close this gap by thoroughly examining the connection between attendance and student achievement using knowledge visualization approaches in Power BI. With this tools, higher educational institutions are able to use the data collected and visualized into a graphical representation to gain useful insights into solving the problem by conducting data analysis utilizing Power BI tools. And by having this data retrieved, educators and educational institutions able effectively make decision, provide recommendation and do future adjustment. The study findings given in this able to assists educators and institutions in understanding students' performance and attendance pattern and to improve the effectiveness of the education system and provide personalized learning experiences. The proposed visualization uses descriptive analytics, predictive analytics and prescriptive analytics to analyze student data. The results will be presented through graphs, charts, and diagrams to facilitate educators' decision-making based on visualized insights.

We utilize the use of Microsoft Excel Spreadsheet and Power BI to develop the knowledge visualization in a more effective way while providing real-time data insights to users. We adapt the 3 types of analysis which are descriptive analysis that best describe the information and provide quick insights to user, predictive analysis that forecast the outcome of the relationship between students attendance and performance, and prescriptive analysis that provide the recommendation to improve the students performance. Therefore in this project, we able to provide a more simple and interact-able solution dashboard for user comparing to the complex visualization proposed by previous work.

7.2 Future works

While this project only shows the relationship between students performance and attendance, it is not enough to provide prediction based on only attendance metrics. Therefore, we can improve this project in the future by adding in more relevant data and variables such as assignment score, test score, and others. In addition, number of data is one important factors that can affect the accuracy of the prediction, we can spend more time to collect relevant data from more different sources. Furthermore, this visualization can be improve by using advanced analytics techniques such as machine learning algorithms to gain more accurate insights of students performance and attendance. Real-time monitoring is also one idea to improve this project that can track students attendance in real time and address the issue more instantly.

REFERENCE

- [1] D. J. Jones, "The impact of student attendance, socio-economic status and mobility on student achievement of third grade students in title i schools," PhD thesis, Virginia Tech, Blacksburg, 2006. [Online]. Available: <https://vtechworks.lib.vt.edu/bitstream/handle/10919/27055/jonesapproveddissertatio nsapr7.pdf?sequence=1&isAllowed=y> [Accessed: 13-July-2023].
- [2] M. R. Edwards and M. E. Clinton, "A study exploring the impact of lecture capture availability and lecture capture usage on student attendance and attainment.," Higher Education, vol. 77, pp. 403–421, Mar. 2019.
- [3] J. A. Gray and M. DiLoreto, "The Effects of Student Engagement, Student Satisfaction, and Perceived Learning in Online Learning Environments" NCPEA International Journal of Educational Leadership Preparation, Vol. 11, No. 1– May, 2016
- [4] M. B. DeCotes, "Data Analytics of University Student Records," Master's Thesis, University of Tennessee, Aug. 2014.
- [5] A. Mężyńska, "Heat map - what exactly is it?," NewDataLabs, 22-Apr-2022. [Online]. Available: <https://newdatalabs.com/en/heat-map-what-exactly-is-it/>. [Accessed: 19-July-2023].
- [6] S. Rovira, E. Puertas, and L. Igual, "Data-driven system to predict academic grades and dropout," PLOS ONE, vol. 12, no. 2, Feb. 2017.
- [7] R. Etemadpour, Y. Zhu, Q. Zhao, Y. Hu, B. Chen, M. A. Sharier, S. Zheng, and J. G. S. Paiva, "Role of absence in academic success: An analysis using visualization tools," Smart Learning Environments, vol. 7, no. 1, Jan. 2020.
- [8] T. M. Phung, "Logistic regression: Advantages and disadvantages," Tung M Phung's Blog, 28-Dec-2019. [Online]. Available: <https://tungmphung.com/logistic-regression-advantages-and-disadvantages/>. [Accessed: 27-July-2023].
- [9] S. Boggavarapu and K. Manikanta, "Students Attendance Visualization", Journal of Engineering Sciences, Vol 13, Aug, 2022.

APPENDIX

- [10] T.Susnjak, G.S.Ramaswami and A.Mathrani. “Learning analytics dashboard: a tool for providing actionable insights to learners”. *Int J Educ Technol High Educ* 19, 12 (2022). <https://doi.org/10.1186/s41239-021-00313-7> [Accessed: 13-March-2024]
- [11] J.Kaur and K.S.Mann. “AI based HealthCare Platform for Real Time, Predictive and Prescriptive Analytics using Reactive Programming”. *Journal of Physics: Conference, Ser. 933 012010, 2017* [Accessed: 11-April-2024]
- [12] N.S.Mosavi and M.F.Santos. “How Prescriptive Analytics Influences Decision Making in Precision Medicine”. *Procedia Computer Science* 177 (2020) 528–533. [Accessed” 11-April-2024]
- [13] B.C.Menezes, J.D.Kelly, A.D.Leal and G.C.Le Roux. “Predictive, Prescriptive and Detective Analytics for Smart Manufacturing in the Information Age”. *IFAC PapersOnLine* 52-1 (2019) 568–573 [Accessed: 11-April-2024]
- [14] M.Relich. “Predictive and Prescriptive Analytics in Identifying Opportunities for Improving Sustainable Manufacturing”. *Sustainability* 2023, 15(9), 7667; <https://doi.org/10.3390/su15097667> [Accessed: 11-April-2024]
- [15] C. Gröger, H. Schwarz and B. Mitschang. “Prescriptive analytics for recommendation-based business process optimization”. *Business Information Systems*, Springer International Publishing (2014, May), pp. 25-37 [Accessed: 11-April-2024]
- [16] C.Cole. “WHAT IS PRESCRIPTIVE ANALYTICS? 6 EXAMPLES”. *Business Insight*. Harvard Business School Online, Nov 2021. <https://online.hbs.edu/blog/post/prescriptive-analytics> [Accessed: 11-April-2024]
- [17] S.Agarwal. “Understanding the Data Science Lifecycle”. *sudeep.co*. 9-Feb-2018 [Online]. Available: <https://www.sudeep.co/data-science/2018/02/09/Understanding-the-Data-Science-Lifecycle.html> [Accessed: 18-April-2024]
- [18] D.Taylor. “Power BI Tutorial: What is Power BI? Why Use? DAX Examples”. *GURU99*. <https://www.guru99.com/power-bi-tutorial.html> [Accessed: 10-August-2023]

APPENDIX

- [19] J. Ochipa. “Microsoft Power BI: Desktop App vs Web App”. Tensoft.
<https://tensoft.com/microsoft-power-bi-desktop-app-vs-web-app/> [Accessed: 11-August-2023]
- [20] P. Harroy, “Power BI - tools and functionalities,” GeeksforGeeks,
<https://www.geeksforgeeks.org/power-bi-tools-and-functionalities/> [Accessed 12-August-2023]
- [21] Stitch, “7 reasons to use Microsoft power bi,” Stitch: A Talend Product, 2022
[Online]. Available: <https://www.stitchdata.com/resources/7-reasons-power-bi/>.
[Accessed: 17-August-2023].
- [22] A. Biswal, “An introduction to power bi dashboard,” Simplilearn, 17-Oct-2022.
[Online]. Available: <https://www.simplilearn.com/tutorials/power-bi-tutorial/power-bi-dashboard#:~:text=A%20Power%20BI%20dashboard%20enables,the%20requirements%20of%20any%20enterprise.> [Accessed: 18-August-2023].
- [23] A. S. Gillis, “What is Excel? everything you need to know - definition by whatis.com,” Enterprise Desktop,
<https://www.techtarget.com/searchenterprisedesktop/definition/Excel> [Accessed: 20-August-2023].

APPENDIX

UC00303	E-COMMERCE PRACTICES	B+	70-74	72	77	A	3.4700	97.90%	91.52%	<p>$F \geq 91.52\%$, your possible course grade will be A, A- or A+</p> <p>$F < 91.52\%$, your possible course grade will be B+, B, B-, C+, C, or F</p> <p>$F = 90.00\%$, your course grade will be F</p>	<p>There is no correlation between the course marks and attendance.</p> <p>If students wish to achieve a grade above average, students should:</p> <ul style="list-style-type: none"> focus on self-learning since attendance will not directly affect the grades. understand the types of business models. be able to prepare appropriate e-commerce marketing strategies on an innovation platform.
		A	60-69	65				85.40%			
		B+	70-74	72				86.70%			
		A	60-69	65				93.00%			
		B+	70-74	72				93.00%			
UC00304	INTRODUCTION TO BUSINESS PROGRAMMING	B	65-69	67	70	B+	3.0000	85.70%	93.22%	<p>$F \geq 93.22\%$, your possible course grade will be A, A- or A+</p> <p>$F < 93.22\%$, your possible course grade will be B+, B, B-, C+, C, or F</p> <p>$F = 90.00\%$, your course grade will be F</p>	<p>There is a positive correlation between course marks and attendance.</p> <p>If students wish to achieve a grade above average, students should:</p> <ul style="list-style-type: none"> achieve attendance above 95.00%, since the attendance is one of the factor of grade marks. pay more attention on class and follow the approach by teachers. able to identify errors and debug programs. <p>If students' attendance is under average and wish to avoid failing the course, students should:</p> <ul style="list-style-type: none"> attend all the class. study and have basic understanding on important topics.
		B-	60-64	62				89.50%			
		A-	75-79	77				94.50%			
		A	60-69	65				100.00%			
		B	65-69	67				96.50%			
UC00304	PROGRAMMING CONCEPTS AND PRACTICES	C	50-54	52	70	B+	3.0000	82.00%	92.80%	<p>$F \geq 92.80\%$, your possible course grade will be A, A- or A+</p> <p>$F < 92.80\%$, your possible course grade will be B+, B, B-, C+, C, or F</p> <p>$F = 90.00\%$, your course grade will be F</p>	<p>There is a positive correlation between course marks and attendance.</p> <p>If students wish to achieve a grade above average, students should:</p> <ul style="list-style-type: none"> achieve attendance above 95.00%, since the attendance is one of the factor of grade marks. pay more attention on class and follow the approach by teachers. able to identify errors and debug programs. <p>If students' attendance is under average and wish to avoid failing the course, students should:</p> <ul style="list-style-type: none"> attend all the class. study and have basic understanding on the topic covered under short class. ask help from lecturers. do more practice.
		B	65-69	67				87.90%			
		B+	70-74	72				100.00%			
		A	60-69	65				100.00%			
		A-	75-79	77				100.00%			
UC00303	OBJECT-ORIENTED SYSTEMS ANALYSIS AND DESIGN	A	60-69	65	63	B-	2.6700	89.50%	91.66%	<p>$F \geq 91.66\%$, your possible course grade will be B+, B, B-, A, A- or A+</p> <p>$F < 91.66\%$, your possible course grade will be B-, C+, C, or F</p> <p>$F = 90.00\%$, your course grade will be F</p>	<p>There is a positive correlation between course marks and attendance.</p> <p>If students wish to achieve a grade above average, students should:</p> <ul style="list-style-type: none"> achieve attendance above 95.00%, since the attendance is one of the factor of grade marks. pay more attention on class and follow the approach by teachers. be able to apply the concepts in the process of system analysis and design. be able to create UML. <p>If students' attendance is under average and wish to avoid failing the course, students should:</p> <ul style="list-style-type: none"> attend all the class. study and practice on the topic covered under short class. ask help from lecturers. search for online website for practice.
		B-	60-64	62				92.10%			
		A	60-69	65				94.00%			
		F	0-49	25				96.00%			
		C+	55-59	57				96.00%			
UC00304	OBJECT-ORIENTED PROGRAMMING PRACTICES	A	60-69	65	68	B	3.0000	93.10%	93.39%	<p>$F \geq 93.39\%$, your possible course grade will be B+, B, B-, A, A- or A+</p> <p>$F < 93.39\%$, your possible course grade will be B-, C+, C, or F</p> <p>$F = 90.00\%$, your course grade will be F</p>	<p>There is no correlation between the course marks and attendance.</p> <p>If students wish to achieve a grade above average, students should:</p> <ul style="list-style-type: none"> focus on self-learning since attendance will not directly affect the grades. be able to write programs with object-oriented concepts and concepts OOPs that extend learning. be able to use IDE. do step by step practices from online courses.
		B-	60-64	62				100.00%			
		A	60-69	65				93.10%			
		A+	90-100	95				94.50%			
		A-	75-79	77				100.00%			
UB01040	BUSINESS ACCOUNTING	C+	55-59	57	76	A	3.6700	84.00%	93.77%	<p>$F \geq 93.77\%$, your possible course grade will be A, A- or A+</p> <p>$F < 93.77\%$, your possible course grade will be B+, B, B-, C+, C, or F</p> <p>$F = 90.00\%$, your course grade will be F</p>	<p>There is no correlation between the course marks and attendance.</p> <p>If students wish to achieve a grade above average, students should:</p> <ul style="list-style-type: none"> focus on self-learning since attendance will not directly affect the grades. search the real life examples that teach step by step. ask help from lecturers. do more practice.
		B	65-69	67				86.20%			
		A	60-69	65				100.00%			
		A	60-69	65				100.00%			
		B	65-69	67				100.00%			
UC00203	DATABASE TECHNOLOGY	C+	55-59	57	71	B+	3.0000	93.70%	94.40%	<p>$F \geq 94.40\%$, your possible course grade will be A, A- or A+</p> <p>$F < 94.40\%$, your possible course grade will be B+, B, B-, C+, C, or F</p> <p>$F = 90.00\%$, your course grade will be F</p>	<p>There is no correlation between the course marks and attendance.</p> <p>If students wish to achieve a grade above average, students should:</p> <ul style="list-style-type: none"> focus on self-learning since attendance will not directly affect the grades. understand the types of relational models, database engines. focus on ERDB. ask help from lecturers.
		B+	70-74	72				93.70%			
		A+	90-100	95				96.20%			
		A	60-69	65				97.90%			
		B	65-69	67				86.20%			
UC00304	DEFINING TECHNIQUES	A	60-69	65	63	A	4.0000	86.20%	94.60%	<p>$F \geq 94.60\%$, your possible course grade will be A, A- or A+</p> <p>$F < 94.60\%$, your possible course grade will be B+, B, B-, C+, C, or F</p> <p>$F = 90.00\%$, your course grade will be F</p>	<p>There is no correlation between the course marks and attendance.</p> <p>If students wish to achieve a grade above average, students should:</p> <ul style="list-style-type: none"> focus on self-learning since attendance will not directly affect the grades. be able to write programs with object-oriented concepts and method and issues to solve the problem. ask help from lecturers.
		B	65-69	67				89.50%			
		A	60-69	65				90.50%			
		B	65-69	67				92.00%			
		C+	55-59	57				100.00%			
UC00213	FUNDAMENTALS OF DIGITAL MEDIA TECHNOLOGY	B	65-69	67	62	B-	2.6700	85.40%	94.27%	<p>$F \geq 94.27\%$, your possible course grade will be B+, B, B-, A, A- or A+</p> <p>$F < 94.27\%$, your possible course grade will be B-, C+, C, or F</p> <p>$F = 90.00\%$, your course grade will be F</p>	<p>There is no correlation between the course marks and attendance.</p> <p>If students wish to achieve a grade above average, students should:</p> <ul style="list-style-type: none"> focus on self-learning and practice since attendance will not directly affect the grades. understand the use of different types of digital media components. be able to use software to create and edit graphics, animation and movies. find and practice to solve various.
		A-	75-79	77				89.50%			
		A	60-69	65				93.70%			
		A	60-69	65				100.00%			
		C+	55-59	57				100.00%			
UC00363	GEOMETRIC METHADRICS	B+	70-74	72	70	B+	3.0000	87.90%	92.62%	<p>$F \geq 92.62\%$, your possible course grade will be A, A- or A+</p> <p>$F < 92.62\%$, your possible course grade will be B+, B, B-, C+, C, or F</p> <p>$F = 90.00\%$, your course grade will be F</p>	<p>There is no correlation between the course marks and attendance.</p> <p>If students wish to achieve a grade above average, students should:</p> <ul style="list-style-type: none"> focus on self-learning since attendance will not directly affect the grades. understand and be able to write logical reasoning and mathematical problem in solving problems. understand and be able to apply algorithms and flowcharts. do more practice.
		A	60-69	65				92.00%			
		B	65-69	67				99.00%			
		B-	60-64	62				97.90%			
		A	60-69	65				97.90%			
UC00322	RARE ANALYSES DEVELOPMENT	A	60-69	65	77	A	3.4700	82.00%	92.60%	<p>$F \geq 92.60\%$, your possible course grade will be A, A- or A+</p> <p>$F < 92.60\%$, your possible course grade will be B+, B, B-, C+, C, or F</p> <p>$F = 90.00\%$, your course grade will be F</p>	<p>There is no correlation between the course marks and attendance.</p> <p>If students wish to achieve a grade above average, students should:</p> <ul style="list-style-type: none"> focus on self-learning since attendance will not directly affect the grades. be able to write programs with object-oriented concepts and method and issues to solve the problem. be able to use IDE. be able to apply the types of data structure into the application.
		B-	60-64	62				86.20%			
		A	60-69	65				92.00%			
		B+	70-74	72				95.70%			
		A	75-79	77				95.00%			

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Trimester 1, Year 4	Study week no.: 1
Student Name & ID: Hoe Jun Sheng 19ACB03230	
Supervisor: Puan Shakiroh Binti Khamis	
Project Title: Knowledge Visualization of Student Performance Based on Attendance	

1. WORK DONE

[Please write the details of the work done in the last fortnight.]

Review the report written in FYP 1

2. WORK TO BE DONE

Discuss with supervisor about the work and improvement for FYP2

3. PROBLEMS ENCOUNTERED

No problems encountered yet.

4. SELF EVALUATION OF THE PROGRESS

Preparing to start the progress on FYP2 by revising the FYP1



Supervisor's signature



Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Trimester 1, Year 4	Study week no.: 2
Student Name & ID: Hoe Jun Sheng 19ACB03230	
Supervisor: Puan Shakiroh Binti Khamis	
Project Title: Knowledge Visualization of Student Performance Based on Attendance	

1. WORK DONE

[Please write the details of the work done in the last fortnight.]

No work done due to Chinese New Year Celebration

2. WORK TO BE DONE

Discuss with supervisor about the work and improvement for FYP2

3. PROBLEMS ENCOUNTERED

No problems encountered yet.

4. SELF EVALUATION OF THE PROGRESS

-



Supervisor's signature



Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Trimester 1, Year 4	Study week no.: 3
Student Name & ID: Hoe Jun Sheng 19ACB03230	
Supervisor: Puan Shakiroh Binti Khamis	
Project Title: Knowledge Visualization of Student Performance Based on Attendance	

1. WORK DONE

[Please write the details of the work done in the last fortnight.]

No work done due to Chinese New Year Celebration

2. WORK TO BE DONE

Discuss with supervisor about the work and improvement for FYP2

3. PROBLEMS ENCOUNTERED

No problems encountered yet.

4. SELF EVALUATION OF THE PROGRESS

-



Supervisor's signature



Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Trimester 1, Year 4	Study week no.: 4
Student Name & ID: Hoe Jun Sheng 19ACB03230	
Supervisor: Puan Shakiroh Binti Khamis	
Project Title: Knowledge Visualization of Student Performance Based on Attendance	

1. WORK DONE

[Please write the details of the work done in the last fortnight.]

Requested a more complete data from supervisor.

2. WORK TO BE DONE

Discuss with supervisor about the work and improvement for FYP2

3. PROBLEMS ENCOUNTERED

Need more time to collect data.

4. SELF EVALUATION OF THE PROGRESS

Slow progression due to previous weeks of holidays.



Supervisor's signature



Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Trimester 1, Year 4	Study week no.: 5
Student Name & ID: Hoe Jun Sheng 19ACB03230	
Supervisor: Puan Shakiroh Binti Khamis	
Project Title: Knowledge Visualization of Student Performance Based on Attendance	

1. WORK DONE

[Please write the details of the work done in the last fortnight.]

Discussed and decided the work to be done in FYP2 with supervisor

2. WORK TO BE DONE

Update the report.

3. PROBLEMS ENCOUNTERED

No problems encountered yet.

4. SELF EVALUATION OF THE PROGRESS

Have a direction on the work on FYP2 and know what to improve.



Supervisor's signature



Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Trimester 1, Year 4	Study week no.: 6
Student Name & ID: Hoe Jun Sheng 19ACB03230	
Supervisor: Puan Shakiroh Binti Khamis	
Project Title: Knowledge Visualization of Student Performance Based on Attendance	

1. WORK DONE

[Please write the details of the work done in the last fortnight.]

Updated the report content on Chapter 1.

2. WORK TO BE DONE

Update the report.

3. PROBLEMS ENCOUNTERED

Data yet to complete collected due to large data.

4. SELF EVALUATION OF THE PROGRESS

Project progress done 20%, will need to push harder for following weeks.



Supervisor's signature



Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Trimester 1, Year 4	Study week no.: 7
Student Name & ID: Hoe Jun Sheng 19ACB03230	
Supervisor: Puan Shakiroh Binti Khamis	
Project Title: Knowledge Visualization of Student Performance Based on Attendance	

1. WORK DONE

[Please write the details of the work done in the last fortnight.]

Updated the report content for Chapter 2 Literature Review.

2. WORK TO BE DONE

Update the report.

3. PROBLEMS ENCOUNTERED

No problems encountered in this week.

4. SELF EVALUATION OF THE PROGRESS

Slow progression due to assignment from other course.



Supervisor's signature



Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Trimester 1, Year 4	Study week no.: 8
Student Name & ID: Hoe Jun Sheng 19ACB03230	
Supervisor: Puan Shakiroh Binti Khamis	
Project Title: Knowledge Visualization of Student Performance Based on Attendance	

1. WORK DONE

[Please write the details of the work done in the last fortnight.]

Data completely collected, created a new PowerBI dashboard work and input the data into the dashboard.

2. WORK TO BE DONE

Perform data cleaning and filtering, design the dashboard, update report for the following chapter.

3. PROBLEMS ENCOUNTERED

Some misunderstanding for the data, had to ask supervisor.

4. SELF EVALUATION OF THE PROGRESS

Project progress done 40%, slow progress due to assignment from other course.



Supervisor's signature



Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Trimester 1, Year 4	Study week no.: 9
Student Name & ID: Hoe Jun Sheng 19ACB03230	
Supervisor: Puan Shakiroh Binti Khamis	
Project Title: Knowledge Visualization of Student Performance Based on Attendance	

1. WORK DONE

[Please write the details of the work done in the last fortnight.]

Updated the report content for Chapter 3, designed 20% for dashboard.

2. WORK TO BE DONE

Perform descriptive, predictive and prescriptive analysis on the dashboard, update the report.

3. PROBLEMS ENCOUNTERED

No problems encountered.

4. SELF EVALUATION OF THE PROGRESS

Project progress done 60%, not ideal due to few weeks left in the semester



Supervisor's signature



Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Trimester 1, Year 4	Study week no.: 10
Student Name & ID: Hoe Jun Sheng 19ACB03230	
Supervisor: Puan Shakiroh Binti Khamis	
Project Title: Knowledge Visualization of Student Performance Based on Attendance	

1. WORK DONE

[Please write the details of the work done in the last fortnight.]

Updated report content for Chapter 4.

2. WORK TO BE DONE

Perform descriptive, predictive and prescriptive analysis on the dashboard, update the report and design poster.

3. PROBLEMS ENCOUNTERED

Slow progress due to midterm examination week.

4. SELF EVALUATION OF THE PROGRESS

Project progress done 70%, slow progression due to midterm.



Supervisor's signature



Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Trimester 1, Year 4	Study week no.: 11
Student Name & ID: Hoe Jun Sheng 19ACB03230	
Supervisor: Puan Shakiroh Binti Khamis	
Project Title: Knowledge Visualization of Student Performance Based on Attendance	

1. WORK DONE

[Please write the details of the work done in the last fortnight.]

Updated the report content for Chapter 5 and 6.

2. WORK TO BE DONE

Perform visualization and analysis for descriptive, predictive and prescriptive in Power BI and analyze the results, design of poster

3. PROBLEMS ENCOUNTERED

No problems encountered.

4. SELF EVALUATION OF THE PROGRESS

Project report done 80%, work on PowerBI dashboard still in slow progression.



Supervisor's signature



Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Trimester 1, Year 4	Study week no.: 12
Student Name & ID: Hoe Jun Sheng 19ACB03230	
Supervisor: Puan Shakiroh Binti Khamis	
Project Title: Knowledge Visualization of Student Performance Based on Attendance	

1. WORK DONE

[Please write the details of the work done in the last fortnight.]

Completed project report, 50% of the dashboard has been done

2. WORK TO BE DONE

Finish up the Power BI dashboard.

3. PROBLEMS ENCOUNTERED

Some technical and skill issue faced while performing visualization on dashboard.

4. SELF EVALUATION OF THE PROGRESS

A rushing week, project report has been done but the dashboard still in progression.



Supervisor's signature



Student's signature

POSTER



UNIVERSITI TUNKU ABDUL RAHMAN
Faculty of Information and Communication Technology

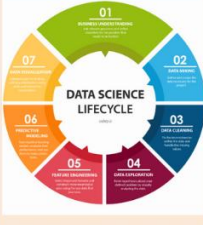
Project Developer: Hoe Jun Sheng
Project Supervisor: Puan Shakiroh
Binti Khamis

KNOWLEDGE VISUALIZATION OF STUDENT PERFORMANCE BASED ON ATTENDANCE

Introduction

In the contemporary educational landscape, student performance evaluation stands as a pivotal facet of pedagogical improvement. Recognizing the significance of attendance as a determinant of academic success, this project embarks on a comprehensive investigation to analyze, visualize, and enhance the understanding of the intricate relationship between attendance and student performance. The research component of this project involves an in-depth analysis of attendance data collected from diverse educational institutions. Through Power BI, we aim to uncover meaningful patterns, relationship, and insights that illuminate the impact of attendance on student outcomes and translates these findings into interactive, visually-rich dashboards and reports, facilitating a deeper understanding of the attendance-performance relationship.

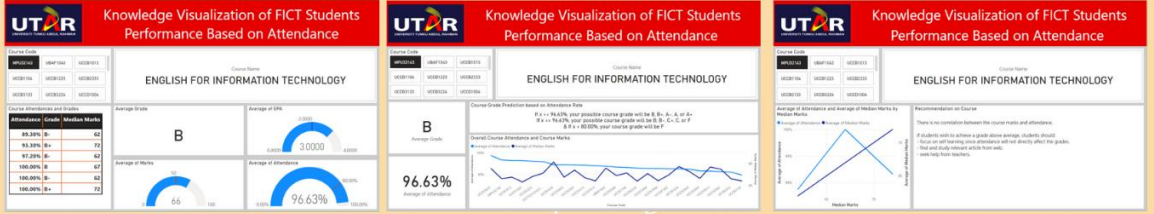
Methodology



DSL

1. Business Understanding: Identify objectives and problems
2. Data Mining: Collect relevant data
3. Data Cleaning: Eliminate null, duplicate and inappropriate values
4. Data Exploration: Analyse pattern using subset of data and make assumption
5. Feature Engineering: Perform feature selection and construction
6. Predictive Modeling: Find the correct model to perform forecast
7. Data Visualization: Create visualization that best describe data relationship

Dashboard Overview



Descriptive Page

- Provide surface information of students information (e.g. attendance, grades, marks)
- Get quick insights
- Easy to make decision
- Best describe the course information

Predictive Page

- Focus on main variables relationship (attendance and grades)
- Provide estimated prediction on students grade
- Show the pattern of all courses attendance and grade

Prescriptive Page

- Show the correlation between attendance and grade for each course.
- Provide different recommendation for different courses
- Allow educational institutions adapt recommendation to enhance students performance

Conclusion

To address the important role that student attendance plays in academic success. Many educational institutions use traditional attendance tracking techniques and don't have access to real-time data. This study aims to close this gap examining the connection between attendance and student achievement using knowledge visualization approaches in Power BI. We utilize descriptive analytics, predictive analytics and prescriptive analytics to perform visualization. The proposed visual analytic tool will visualize student data through dashboard containing graphs, charts, and diagrams by utilizing the use of Excel sheets and Power BI to show the relationship of students performance and attendance and help educators make easier decisions.

Future Works

- Adding data from other variables (assignment score, test score, etc)
- Collect data from other different sources
- Perform visualization using advanced analytics tools
- Provide real-time monitoring features for educators to make quick adjustment

PLAGIARISM CHECK RESULT

ORIGINALITY REPORT			
6%	6%	2%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS
PRIMARY SOURCES			
1	Submitted to Universiti Tunku Abdul Rahman Student Paper		1%
2	slejournal.springeropen.com Internet Source		1%
3	www.utar.edu.my Internet Source		1%
4	www.coursehero.com Internet Source		1%
5	www.geeksforgeeks.org Internet Source		1%
6	towardsdatascience.com Internet Source		<1%
7	fict.utar.edu.my Internet Source		<1%
8	hdl.handle.net Internet Source		<1%
9	Submitted to Australian National University Student Paper		<1%

PLAGIARISM CHECK RESULT

10	Submitted to Southern New Hampshire University - Continuing Education Student Paper	<1 %
11	Submitted to Camarines Sur Polytechnic Colleges Student Paper	<1 %
12	OFER ARAZY, CARSON C. WOO. "Analysis and design of agent-oriented information systems", The Knowledge Engineering Review, 2002 Publication	<1 %
13	dspace.lib.cranfield.ac.uk Internet Source	<1 %
14	docplayer.net Internet Source	<1 %
15	chubuweb.hosp.pref.okinawa.jp Internet Source	<1 %
16	Benjamin Kettner, Frank Geisler. "Chapter 9 Power BI", Springer Science and Business Media LLC, 2022 Publication	<1 %
17	digital.wpi.edu Internet Source	<1 %
18	fdp.spit.ac.in Internet Source	<1 %

PLAGIARISM CHECK RESULT

19	Doug Rose. "Data Science", Springer Science and Business Media LLC, 2016 Publication	<1%
20	Karthik Ramasubramanian, Abhishek Singh. "Machine Learning Using R", Springer Science and Business Media LLC, 2017 Publication	<1%
21	S. Suriya, J. Shyamala Devi, R. Agusthiyar. "chapter 2 Data Visualization in Large Scale Based on Trained Data", IGI Global, 2024 Publication	<1%
22	ee.bradley.edu Internet Source	<1%
23	rspb.royalsocietypublishing.org Internet Source	<1%

Exclude quotes On

Exclude matches < 8 words

Exclude bibliography On

PLAGIARISM CHECK RESULT

Form Title: Supervisor's Comments on Originality Report Generated by Turnitin for Submission of Final Year Project Report (for Undergraduate Programmes)			
Form Number: FM-IAD-005	Rev No.: 0	Effective Date: 01/10/2013	Page No.: 1 of 1



FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY

Full Name(s) of Candidate(s)	HOE JUN SHENG
ID Number(s)	19ACB03230
Programme / Course	BACHELOR OF INFORMATION SYSTEMS (HONOURS) BUSINESS INFORMATION SYSTEMS
Title of Final Year Project	KNOWLEDGE VISUALIZATION OF STUDENT PERFORMANCE BASED ON ATTENDANCE

Similarity	Supervisor's Comments (Compulsory if parameters of originality exceed the limits approved by UTAR)
Overall similarity index: <u>6</u> % Similarity by source Internet Sources: <u>6</u> % Publications: <u>2</u> % Student Papers: <u>2</u> %	
Number of individual sources listed of more than 3% similarity: <u>0</u>	
Parameters of originality required, and limits approved by UTAR are as Follows: (i) Overall similarity index is 20% and below, and (ii) Matching of individual sources listed must be less than 3% each, and (iii) Matching texts in continuous block must not exceed 8 words <i>Note: Parameters (i) – (ii) shall exclude quotes, bibliography and text matches which are less than 8 words.</i>	

Note: Supervisor/Candidate(s) is/are required to provide softcopy of full set of the originality report to Faculty/Institute

Based on the above results, I hereby declare that I am satisfied with the originality of the Final Year Project Report submitted by my student(s) as named above.

Signature of Supervisor

Signature of Co-Supervisor

Name: SHAKIROH BINTI KHAMIS

Name: _____

Date: 25/04/2024

Date: _____

FYP 2 CHECKLIST

UNIVERSITI TUNKU ABDUL RAHMAN
FACULTY OF INFORMATION & COMMUNICATION
TECHNOLOGY (KAMPAR CAMPUS)
CHECKLIST FOR FYP2 THESIS SUBMISSION

Student Id	19ACB03230
Student Name	Hoe Jun Sheng
Supervisor Name	Dr Shakiroh Binti Khamis

TICK (√)	DOCUMENT ITEMS
	Your report must include all the items below. Put a tick on the left column after you have checked your report with respect to the corresponding item.
√	Title Page
√	Signed Report Status Declaration Form
√	Signed FYP Thesis Submission Form
√	Signed form of the Declaration of Originality
√	Acknowledgement
√	Abstract
√	Table of Contents
√	List of Figures (if applicable)
√	List of Tables (if applicable)
√	List of Symbols (if applicable)
√	List of Abbreviations (if applicable)
√	Chapters / Content
√	Bibliography (or References)
√	All references in bibliography are cited in the thesis, especially in the chapter of literature review
√	Appendices (if applicable)
√	Weekly Log

FYP 1 CHECKLIST

√	Poster
√	Signed Turnitin Report (Plagiarism Check Result - Form Number: FM-IAD-005)
√	I agree 5 marks will be deducted due to incorrect format, declare wrongly the ticked of these items, and/or any dispute happening for these items in this report.

***Include this form (checklist) in the thesis (Bind together as the last page)**

I, the author, have checked and confirmed all the items listed in the table are included in my report.



(Signature of Student)

Date: 23 April 2024