A DASHBOARD FOR MONITORING ACADEMIC STAFF TEACHINGS & CITATIONS

By

BRIAN NG YAO SHENG

A REPORT

SUBMITTED TO

Universiti Tunku Abdul Rahman

in partial fulfillment of the requirements

for the degree of

BACHELOR OF COMPUTER SCIENCE (HONS)

Faculty of Information and Communication Technology

(Kampar Campus)

MAY 2020

UNIVERSITI TUNKU ABDUL RAHMAN

REPORT STATUS DECLARATION FORM

Title: A DASHBOARD FOR MONITORING ACADEMIC STAFF TEACHINGS AND CITATIONS

Academic Session: MAY 2020

I BRIAN NG YAO SHENG

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:

62, LORONG TAMBUN INDAH 2

TAMAN TAMBUN INDAH,

14100, SIMPANG AMPAT,

PULAU PINANG

DR PRADEEP ISAWASAN

(Supervisor's name)

Date: 10 SEPTEMBER 2020 **Date**: 10 SEPTEMBER 2020

A DASHBOARD FOR MONITORING ACADEMIC STAFF TEACHINGS & CITATIONS

By

BRIAN NG YAO SHENG

A REPORT

SUBMITTED TO

Universiti Tunku Abdul Rahman

in partial fulfillment of the requirements

for the degree of

BACHELOR OF COMPUTER SCIENCE (HONS)

Faculty of Information and Communication Technology

(Kampar Campus)

MAY 2020

DECLARATION OF ORIGINALITY

I declare that this report entitled "**A DASHBOARD FOR MONITORING ACADEMIC STAFF TEACHINGS & CITATIONS**" 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 : BRIAN NG YAO SHENG

Date : 10/09/2020

ACKNOWLEDGEMENTS

First of all, I would like to express my gratitude to my supervisor Dr.Pradeep Isawasan for the continuous support and giving me guidance in completing my final year project. He puts his faith in me since IIPSPW, believe in me, that is where my motivation and responsibility came in which pushed me forward in times of hardship faced during the development of the whole final year project. Without his support and guidance and his time allowing me to use his wired connection inside his office, this project would not have been possible.

Apart from my advisor, I would like to thank my parents for the endless support, understanding, words of encouragement given throughout these 4 years of university life in UTAR Kampar and the most important one, paying my tuition fees and giving me enough allowance.

I thank my golden trio group, Hoe Yean Sam and Chong Guang Huan who stick around with me throughout all assignments group forming we had in UTAR, all those sleepless nights working together before deadlines, all the fun we shared together for the past 3 years of degree life.

Finally, I would like to thank all of my lecturers that have taught me before, I appreciate all of the knowledge both in life and academic.

ABSTRACT

As data visualization is getting more attention, a lot of organization started to invest in it because data visualization is an essential skill as most of the decisions are reliant upon data. With the recent cases in COVID-19, it is made clear that data visualization played a crucial role in fighting against the disease, ensuring the number of cases is below the curve. Similarly, in order to manage a university effectively, the use of dashboard to monitor the staff teaching performance and citation is inevitable, this would eventually affect the ranking of a university. In this report, the university staff data will be collected, analyzed and come up with a dashboard that can view all of the university staff performance in a glance. The first project objective of this report is to collect data, analyze the total publications, citations and teaching performance, then develop a dashboard that monitors on the number of publications, citations and teaching performance and finally assist the university in making decisions in a quick and timely manner to improve the university ranking. The implementation begins from ground zero which starts from data collection through Scopus API and store it as JSON file to be used later for data visualization, along with the sample teaching survey data provided by the supervisor of this project for the data visualization of teaching survey section. The final product of university staff performance dashboard will be developed and ready to use. The output of the dashboard could be filtered based on the faculty of the university, department and the category of the data, some sentiment analysis of the staff performance based on the comments and reviews given by the students will be displayed. For the future work, the dashboard could be made real time in terms of data collection, collecting data from other academic databases, expanding dashboard service to other higher institutions.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

EIS	Executive Information Systems
OLAP	Online Analytical Processing
KPI	Key Performance Indicator
BPM	Business Performance Management
KSU	Kent State University
THE	Times Higher Education
UT	University of Texas
UOW	University of Wyoming
PDF	Portable Document Format
API	Application Program Interface
CRISP-DM	Cross-Industry Standard Process for Data Mining
UTAR	Universiti Tunku Abdul Rahman
FICT	Faculty of Information and Communication Technology
FYP	Final Year Project
JSON	JavaScript Object Notation
CSV	Comma Seperated Value
HTML	Hypertext Markup Language
DDDM	Data Driven Decision Making
GUI	Graphical User Interface
PYPI	Python Package Index
IP	Internet Protocol
PNG	Portable Network Graphics

CHAPTER 1: INTRODUCTION

1.1 Problem Statement

Data visualization is getting a lot of attention these few years, it is vital to organization success as Littlewood (2018) stated that the data visualization skills is the perfect choice of investment as it has high return rate with high utility and lesser learning time. Berinato (2016) suggests that data visualization is a compulsory skill to have for all executives of the company, as many organization decisions are reliant upon data. Needless to say, the top-level management executives require large amount of information and data to aid in decision-making process to determine the direction of an organization. Therefore, majority of the organization needs a strong information technology team to assist the organization in acquiring information in a timely manner.

A good data visualization will ease up the time taken on decision making, this is something that every organization sought of. Likewise, public and private higher education is deviating towards data-informed decision-making (Bouwa-Gearhart & Collins, 2015) and thus there is a rising need to build data capacity and information literacy in higher education. In order to compete with over 25,000 universities in the world, a proper performance monitoring dashboard will significantly increase the advantage over other competing universities. There is a total of 20 public universities and 60 private universities in Malaysia, most of the private universities are often seek profit in providing tertiary education. Hence, the competition among private universities are stiffer than of public university.

Despite all of the data visualization advancement and importance being mentioned a lot by the articles and researchers in foreign countries, data visualization in local higher education institution is still introductory stage, they have yet to see how data is able to improve the quality of higher education. Matthews (2018) said that with the help of data, universities are able to find out the factors that help students succeed. For instance, the data from Southern Connecticut State University discover that a sense of belonging was one of the two main factors affecting persistence in study. In possession of such data could not only increase the retention rate but able to give advices to students and avoid unnecessary struggles.

Malaysian universities **lacks of a university owned dashboard that monitors on publications and citations** published by their academic staff or post-graduate students. Instead, third party sites like Google Scholar or Scopus were relied on to do the job. Besides, **Scopus can only check the publications of a particular author one at a time**, a platform to view and compare the performance of all of the university staff is required.

Furthermore, **data visualization is rarely adopted by Malaysian Universities.** Think of the possibilities that the universities could do with all these data visualization on their hands, the institutions of higher education could find themselves do exceptionally well to stay competitive by able to fix the issue right away.

Dashboards have been shown to be an effective tool for showing data for decision makers. Besides, in order for the local higher education institution to thrive in the Times Higher Education Ranking which ranks university based on the citations index and become a prestigious university, data visualization with dashboard is a necessity.

1.2 Project Motivation

Higher education institutions worldwide are running in an ever changing and complicated environment. The globalization processes and the rise of artificial intelligence and big data analytics pave way for better data driven decision making, this have caused the society to become competitive not only between companies but also between universities. Leading higher education institutions in western countries have already realised long ago that they need dedicate themselves in creating contemporary management approaches and using dynamic methods and techniques in order to remain competitive. Universities are now focusing not only on their main objective which is providing high quality education but also offering better master programmes benefits to attract more undergraduate students to pursue masters programme to aid in the research of the university, providing more opportunities to PhD and post-doctorate students for working on innovative research topics.

However, most of the Malaysian Universities are still coordinated based on the conventional lecture hall educational system model. University data is often gathered upon the admission of new students, collecting student feedbacks about the university and lecturer, all of the collected data is always stored in electronic format, online course content pages, university portal are usually available for every university in Malaysia. However, there are still no integrated university dashboard being developed for the university board of management. Since most of the universities take their world ranking seriously, it would create a competitive environment between other rival universities, this is where monitoring dashboards comes in handy. The ability of transforming data into information and knowledge allows universities remain competitive and based on the insights of data, they are capable to make quick and competent managerial decisions.

As a result, a dashboard for monitoring academic staff teachings and citations was inspired and developed in this project.

1.3 Project Scope

In this project, a dashboard that monitors on university staff in teachings and citations will be delivered, the data visualization is projected to be fit in one screen without too much of page scrolling needed. The targeted users for this proposed project are the executives of local higher education institution and specifically on the institute of postgraduate studies and research of each university and the publications by the academic staff, whereby publication monitoring is essential for them to keep track on their citations because the performance of a university is credited by the total citations as well.

In pursuance of accurate data on publications and citations, efforts **to establish connection to Scopus API and data retrieval** will be carried out. Metadata like citation count, document count, author names, h-index, co-author count, affiliation-name were required to create this project. As Scopus is the most preferred abstract and citation database all across Asia, it is often used by Malaysian authors to keep track on their citations and publication. However, Scopus is still unreliable as it only focuses on one author at a time, if the executive of a university wishes to check on the staff performance, the executive has to do it one by one, so a platform to review all of the staff is required. Currently, the scope of the data is limited to FICT only, it will gradually expand to the whole university or include comparisons with the other university. Regarding to the issue of data visualization in Malaysian University, this project will **utilize the data visualization tool known as Plotly Dash to create the dashboard**. Apart from that, a sample student review data from a lecturer will be used **to perform sentiment analysis on academic staff review**. As a result, a performance monitoring dashboard specifically owned by the university will be developed, this eliminates the need of the user going to Scopus to search for their publication and citations.

1.4 Project Objectives

In order to tackle the shortcomings listed in the problem statement, the first objective of this project is to **develop a dashboard visualise all publication data and teaching survey data at a glance**. The dashboard should provide an interactive, centralized, intuitive, responsive design yet simple enough for the user. As a result, it cancels off the need of the university executives to filter through each of the university staff profiles one by one using abstract and citation databases, which is inefficient and time-consuming.

This project is aimed **to collect data and analyse the total publications, citations and student feedbacks.** Without a proper platform for the university to keep track on their citations, the ranking of the university might become stalemate, currently UTAR is using cumulative citations and publication to create the annual research and development report and yet to have a proper platform to monitor on each of the academic staff publications. Student feedbacks are the utmost element for a lecturer to improve the teaching performance. However, not all of the student feedbacks are helpful, some of it are irrelevant, this is where sentiment analysis come in place, the dashboard will be able to identify and measure the emotions or feelings given by the students as sometimes humans could not identify the true sentiment value of a piece of text. This is because people have different views on evaluating a piece of text, mainly due to different people having difference way of textual interpretation.

Besides, it is projected to **assist the university in making decisions in a quick and timely manner**. Data driven decision making is what drives a company to success in this competitive world, as (Morris, 2018) said that if you make decisions based on facts and data, it greatly increases the speed of decision making, far better than making decision based on gut or human intuition. Now, with the help of this performance dashboard, the executives are able to oversee the overall performance of all of the staff, thus at the same time improving the quality of decision making for the board of executives of a university.

1.5 Impact, Significance and Contribution

The big data sector is currently a global rising tech trend which has caught huge amount of attention by various corporate giants, since then it has continuously uplifted by professionals. The current technological trend is still cloud computing, whereby majority of the companies already have their cloud storage in place, eliminates the need of an actual database in the company. However, after the downfall of could computing, big data will become the largest influencer for organization across various types of job sector in terms of decision making. This results in the rise of a new systematic approach known as DDDM (data driven decision making), the demand for data visualization techniques was accelerated by data driven decision making, visual patterns will become the norm of every company.

This dashboard will be the perfect platform for the boards of the executives of the university to plan for a better faculty hiring based on the performance of university staff, the data collected and visualized in dashboard ensures the understanding of instructional personnel. The publications and citations of the academic staff is one of the important aspects in raising the overall university ranking. Based on (Ebrahim, 2014), he mentioned that Times Higher Education World University Ranking system, the Citations which is the research influence, citations and teaching wraps up the whopping 90% of the overall ranking score. The university will be excluded from the world university rankings if the university does not teach undergraduates, teaching solely on a subject or the publication of the university is lesser than 50 papers a year. The **short-term significance of this dashboard is to improve the overall academic staff performance especially in teachings and citations** while **the long term significant is to bring the overall ranking of the local university to ascend to a higher position.**

1.6 Background Information

1.6.1 Historical Development

Citation indexing is a way user acknowledging the value of the information by giving it a measurement to determine the quality of the work by calculating the impact of the piece of information makes on the community. Anyone who uses the piece of information and cites the source material determines the influence and impact of the information and its author, due to the simplicity, people usually unaware that citation indexing is actually now a form of information management and retrieval.

There are 3 factors that drives the advancement of citation indexing, the government started to allocate more funds into research and development, public resentments towards the capacity of subject indexing and the rise of automation.

Thanks to the World War II, the government all over the world begin to resort in research and development, thus opening a new path for the research community to begin document its findings though the accepted channel of published scientific journal literature, this causes a spike of demand for a method of indexing and retrieval that is more cost effective and efficient than the current way of managing information.

Apart from that, the capacity of subject indexing is inadequate, unable to fulfil the increasing amount of active researcher, it took months to adding materials to the indexes which is very time consuming.

The third factor is modern computerization created the perfect timing for the development of citation indexing, the founder of bibliometrics and scientometrics realised that the whole point of this citation is to give credit to the author by using this citation method to uphold the idea or methodology and cited as relevant in the bibliography.

Today, citation indexing is considered to be one of the most reliable way to trace the development of an idea across various abstract and citation database such as Scopus and Google Scholar.

1.6.2 Targeted Users

In contemplation of incorporating citation, publication and teaching performance monitoring, dashboard has been used as the platform. A performance dashboard is basically a diagnostic tool created to visualise the overall organization's performance, especially for the busy board of executives. Dashboard is able to provide information about the current condition of any organization with a peace of mind. The key users of performance dashboard for universities will be the president, vice president, dean, deputy dean and head of department. The design of the dashboard is calibrated to suit the needs of the board of executives.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

2.1.1 Definition

First and foremost, there are no definition that best define the term 'Dashboard' because different dashboard developers will have their own definition. Cotgreave et al. (2017) defined a dashboard is a graphical representation of data used to supervise conditions and/or promote understanding, while Few (2006) states that a dashboard is a graphical representation of data with great significance required to fulfil multiple objectives; merged and sorted, utilising every space of the screen so that the information can be conveyed at a glimpse.

The dashboard is an effective medium to view data and information. This is why there are a lot of company have started to use dashboard.

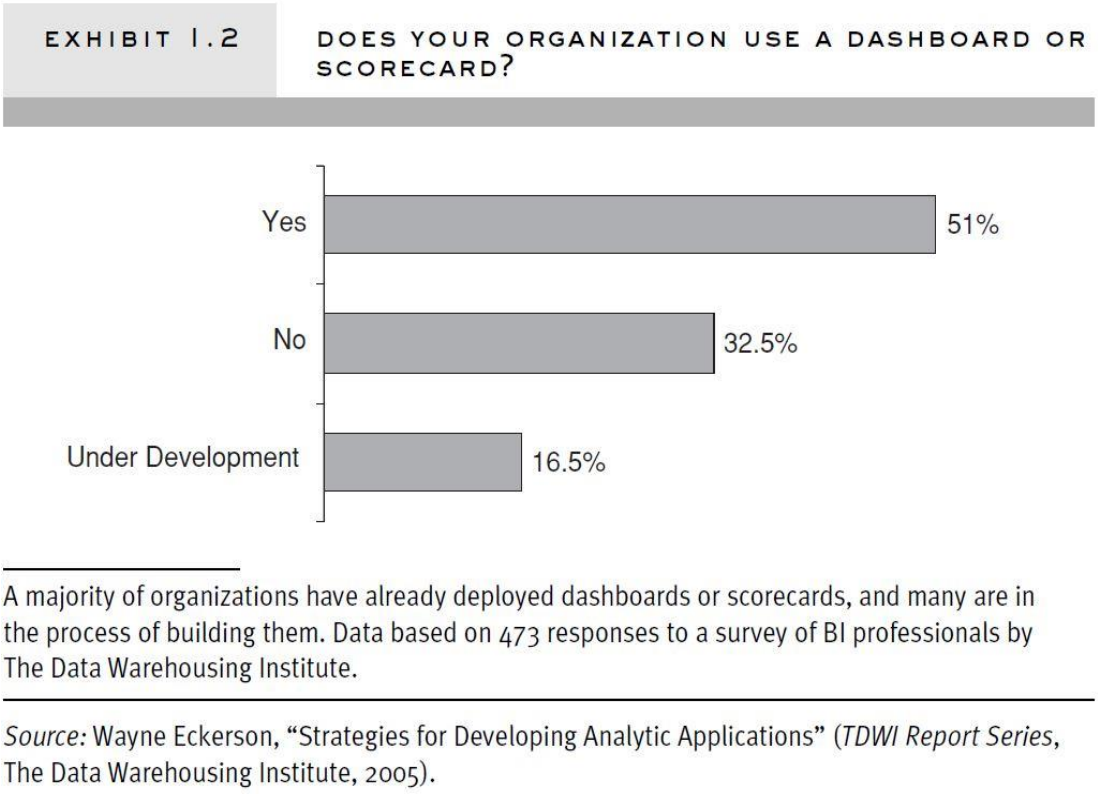


Figure 2.1.1.1: The trend of using dashboard in an organization

2.1.2 Characteristics of Dashboard

Normally, the characteristics of a dashboard includes:

- visual components (charts, bar, maps) to accentuate the data and exceptions that demand action
- show a dynamic, single view of the dashboard with timely data refreshes
- display key indicators in a compact, visceral format
- simple and ease of use
- provides drill-down or drill-through to hidden data sources
- collects data from a variety of source systems

2.1.3 The types of Dashboard

There are 3 types of dashboard:

- **Strategic**
- **Analytical**
- **Operational**

Strategic dashboard is a report-based dashboard that plays a major role in maintaining the company's long-term strategy by determining the critical success factors of a company. The process of creating a strategic dashboard is complicated but it can bring an enterprise-wide impact to a business and it is mostly used by senior level management, this is why it is known as "executive dashboard". It provides a brief outline for the stakeholders to observe the current situation and seize any available business opportunity. This type of dashboard concentrates on performance analysis that are distinguishable, including making estimation for future guidance and reference. (Durcevic, 2018)

Analytical dashboard consists of large amount of data developed by data analysts to aid board of executives in making decision. It delivers comprehensive overview of data with middle management being an essential part of its usage to those business companies. Similar to strategic dashboards, analytical dashboard favours static snapshots of data that are not constantly changing periodically. Nevertheless, more sophisticated display media are beneficial for the analyst who must go through complicated data and relationship.

Operational dashboard is a type of dashboard that handles auditing job and supervise short-term processes. It is commonly operated by middle-level management bodies to keeping track on the business process. These kinds of dashboards are considered the most common dashboard, typically used for monitoring and analysing a company activity. These dashboards often emphasize exception alerting and are based on real-time data, because of this operational dashboard are more precise than strategic dashboards. (Few 2006)

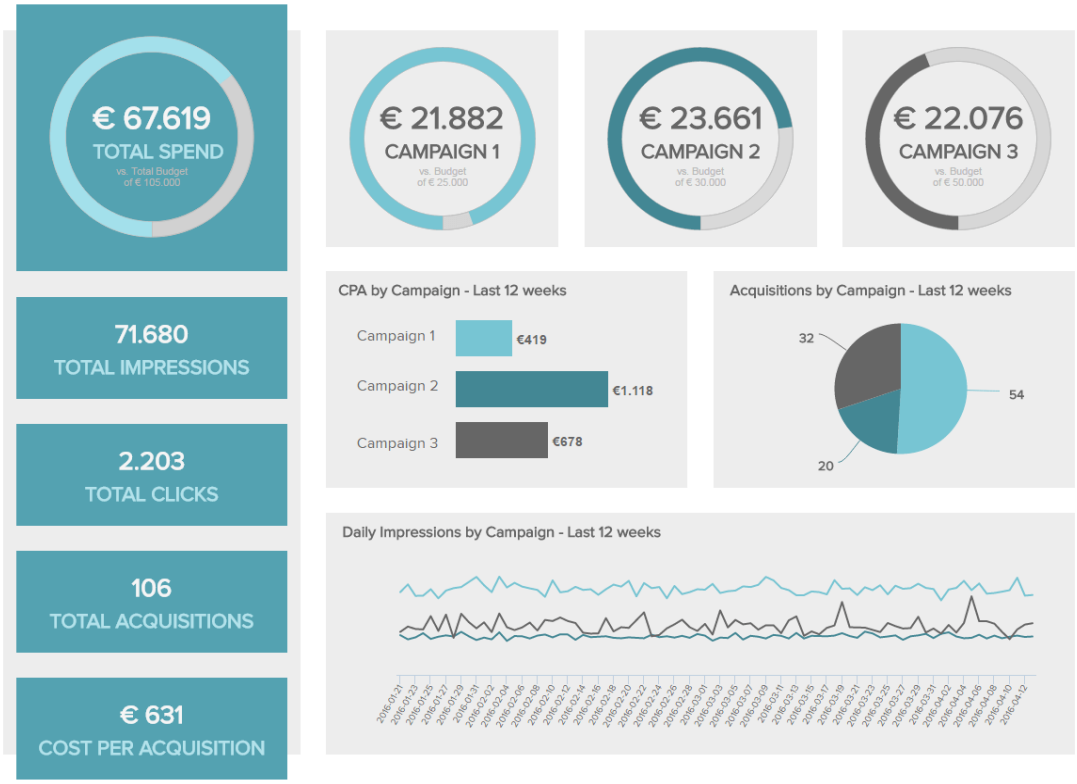


Figure 2.1.3.1: An example of an Operational Dashboard (Durcevic, 2018)

2.2 Kent State University Dashboard

2.2.1 Brief

Kent State University (KSU), a public university in Kent, Ohio, U.S. founded in 1910. It has seven regional campuses in Northeast Ohio. The size of the main campus is comfortably spacious with the value of 953 acres with spinney of trees all over the campus to provide a green and relaxed place for students to study in. (Times Higher Education (THE), n.d.) Kent State University boasts its total enrolment of more than 39,000 students, comprised of more than 2,000 international students from 104 countries. (U.S. News & World Report, 2018) KSU rose into international prominence during the Kent State Shootings on May 4 1970, members of the Ohio National Guard fired a few rounds into a crowd of Kent State University demonstrators on demanding the end of Vietnam War, 4 students were killed and 9 students were wounded in this incident. (Lewis and Hensly, 1998)

The Kent State University have 16 different dashboards to keep track on the institutional progress in order to fulfil the six-year envisagement. (Kent State University, n.d.). The layout of the University Dashboard main page is well organised, it does not leave the user feeling puzzled. Since the logo of the KSU comprised of blue and yellow, the colour of the dashboard was blue and yellow as well, this shows uniformity and consistent use of colour. The source of the peer university data is obtained from IPEDS, Integrated Postsecondary Education Data System.

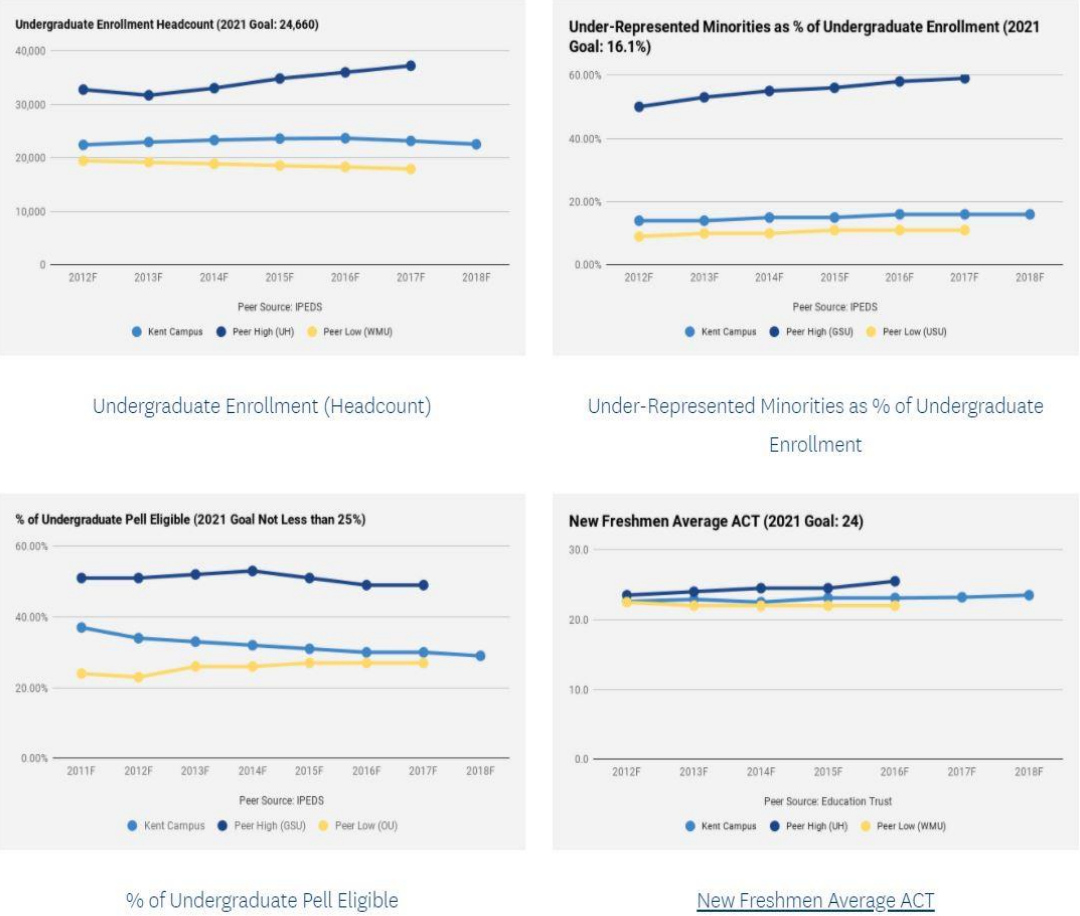
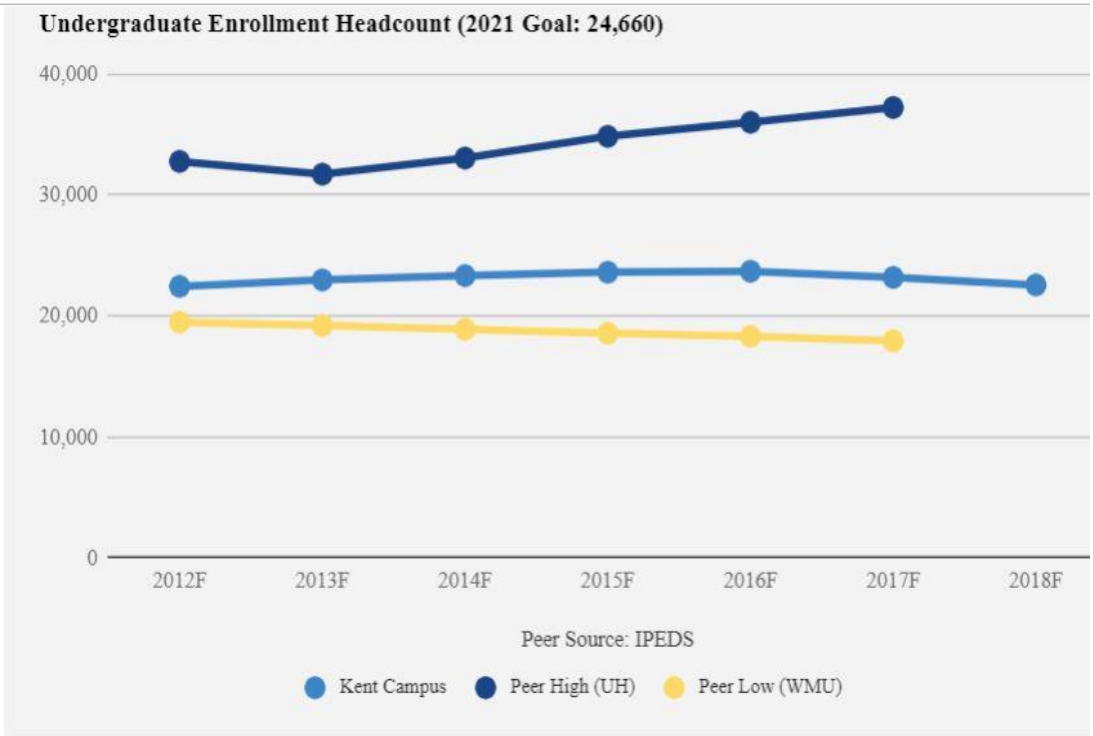


Figure 2.2.1.1: Dashboard overview of Kent State University

2.2.2 Strength

In the main dashboard overview of Kent State University, the best part of the KSU dashboard is that it **compares its data with peer institution** thus enabling the people to make better decision based on this dashboard without the need of referring to the other university's dashboard. Besides, the dashboard uses **"colour-blind friendly" colours**, using blue-yellow palette works well because people with colour-blind will be able to distinguish each other. (Wexler, Shaffer and Cotgreave, 2017). In addition, since all of the dashboards of KSU is yearly based, **line graph was fully utilised to show change over time**, the height and slopes of the line allow the user to see trends and making comparisons to the previous year. The layout of the main overview of KSU dashboard is **well-organised and easy to understand**, it gives the user a feeling to pick any one of it to view the details.



Peer Institution Concurrent Undergrad Enrollment
Source: IPEDS (Undergraduate Enrollment)

Institution Name	2012F	2013F	2014F	2015F	2016F	2017F	2018F
University of Houston	32,760	31,706	33,037	34,830	35,995	37,215	
University of North Texas	30,593	31,243	29,758	30,533	31,241	31,448	
Georgia State University	24,656	24,868	25,328	25,140	25,228	25,770	
Utah State University	25,445	24,385	24,271	25,259	24,838	24,618	

Figure 2.2.2.1: Kent State University Undergraduate Enrolment Dashboard

2.2.3 Weakness

The main limitation to KSU dashboard is that **it only uses line chart** for all of the 16 KSU dashboards which is rather boring for the user because every category of KSU dashboard basically looks the same, just that the data and attribute name is different. In addition, the only interaction of the dashboard between the user is by hovering the mouse over the plot showing detailed information of that piece of data. This results in **lack of user interactive features.** Users are not provided with more buttons or slicers to view the data in a more detailed way, instead it limits the user to only view it in static

manner. Once the user chooses the desired category and view it in details, the **dashboard does not fit in one view**, the user has to scroll through to see all of the information.

2.2.4 Recommendation

Kent State University should **consider including bar chart** to the dashboard in order to show the diversity in usage of the type of chart. As (Wexler, Shaffer and Cotgreave, 2017) stated that bar chart is simply the most efficacious way to compare data, judging from the perspective of psychology, human beings are remarkably well at looking at small differences in length from a common baseline. Moving on, the dashboard **should be made in one view, without the need of scrolling.** (Few, 2006) mentioned the users assume any piece of information that is not seen in the first glance is less important than of what is immediately visible, users are less often to check on what lies off the screen.

2.3 The University of Texas System Dashboard

2.3.1 Brief

The University of Texas System is undeniably having one of the largest public university systems in the United States. This system handles 14 institutions with a staggering amount of enrolment which exceeds 234,000 students, University of Texas System is responsible for more than one-third of the Texas state's undergraduates, nurtures roughly two-thirds of the state's medical practitioner yearly and contributed for almost 70 percent of research funds given to public institutions in Texas. (The University of Texas System, 2017)

The UT System Dashboard was created with the purpose to bring current data into one place, showing trends over time and add comparative benchmarking in order to ease decision making and policy making. (The University of Texas System, 2017)

The UT System Dashboard is made using Microsoft Power BI, a business analytics service by Microsoft. It is a software providing interactive visualization and business intelligence with a user-friendly interface to create reports and dashboards. (Powerbi.microsoft.com, 2019)

Figure 2.3.1.1: Dashboard overview of The University of Texas System

2.3.2 Strength

The developers of University of Texas Dashboard System understand humans organise visual information as groups, According to Gestalt Principles, human brains work better by looking at an object representing a group. (Aanderud, 2018) This is why before dashboard is shown, it allows the user to choose from a choice of 8 different dashboard topics. Once the user chose a topic, the user will be able to **look at the dashboard in one view.** This is a major advantage for a dashboard, as people tend to forget what they see if they are not looking at the part of data anymore, as a dashboard developer, never expect the users to be able to recall the piece of information that is no longer in their area of vision. (Few, 2006)

Next, it **prioritises user interactivity**. For instance, in the fall enrolment system, the users not only able to see detailed data by hovering on the bar or pie chart, there are buttons and slicers for the user to play with. If the users with to see how many Asians are enrolled to University of Texas in the fall intake trimester, by just clicking the Asian button provided below, the colour intensity of Asian data will remain and the colour intensity of the other data will be on low contrast.

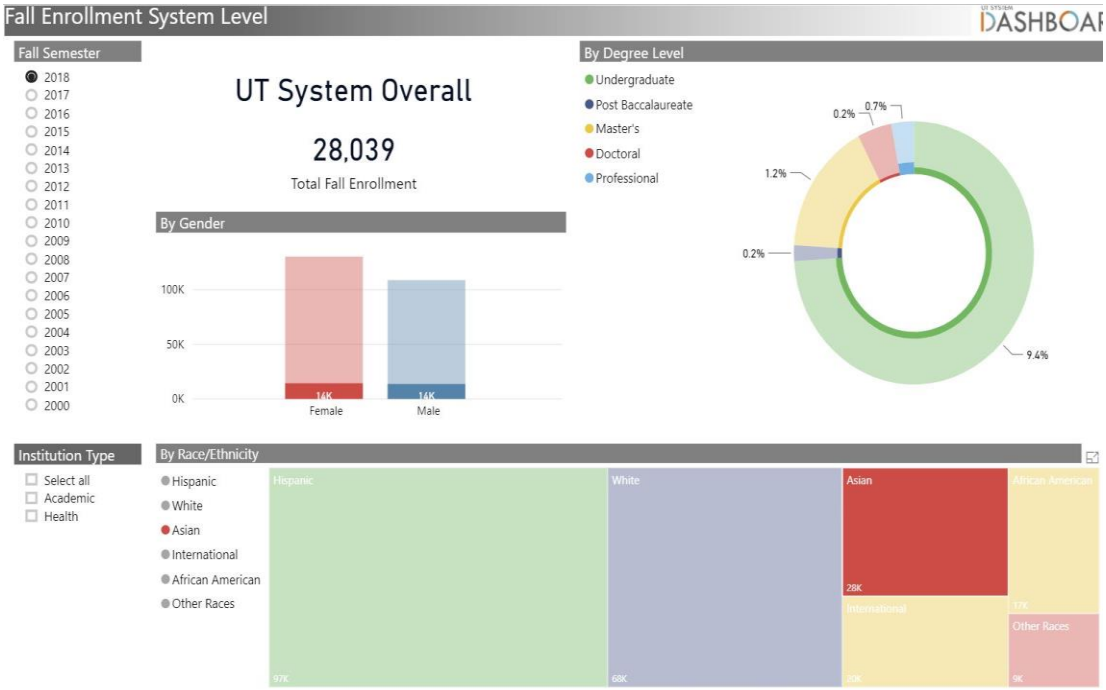


Figure 2.3.2.1: Fall Enrolment System of University of Texas System Dashboard with Asian attribute selected

Moving on, the University of Texas System Dashboard **obeys F-Shaped Pattern.** When the users read in a first horizontal movement, the users will see the total fall enrolment with a total of 239,086. This layout is vital for a dashboard, the reason behind people click and view the enrolment dashboard is to know the total enrolment and what the user sees it first will determine whether the user will continue to look on other content. (Pernice, 2017)

2.3.3 Weakness

Overall, it is considered a good dashboard system being compared to so many peer institutions. However, the users **are unable to download the dataset** in csv or png format. even if some of the dashboard data is not sensitive nor confidential.

2.3.4 Recommendation

University of Texas System should consider sharing the dataset used to create all these dashboards, so that people could study on how they create dashboard.

2.4 University of Wyoming Dashboard

2.4.1 Brief

University of Wyoming, UOW is a public institution located 7220 feet above sea level in the city of Laramie, founded in 1886. The total undergraduate enrolment is 9,791 with the campus size of 785 acres. UOW offers more than 80 undergraduate courses in the fields of Agriculture, Arts & Sciences, Business, Education, Engineering, Health Sciences and Law. University of Wyoming's ranking in the 2019 edition of Best Colleges & Universities is 183. The students of University of Wyoming often host an alcohol-free Friday night fever open to all students to join their fun filled activities including, sand art, balloon games, tie-dye etc.

The University of Wyoming Dashboard has a special name, known as UW Brown & Gold Report. It was just developed recently by the developers with new interface, more interactive functions and the dashboard are able to view in a glance. However, it is strictly for the students and lecturer of UOW only as it requires the user to sign in to view the dashboard.

Nevertheless, there are old version of UOW dashboard that are still accessible by the public. UOW prepares their dashboard in pdf form and releases 2 version annually on

average. The dashboard is generated and consolidated into one file with 6 major category and 38 sub-category dashboards. The latest accessible version of the dashboard being released by UOW is Fall 2016, but the link is not working, so April 2016 was considered to be the latest version.

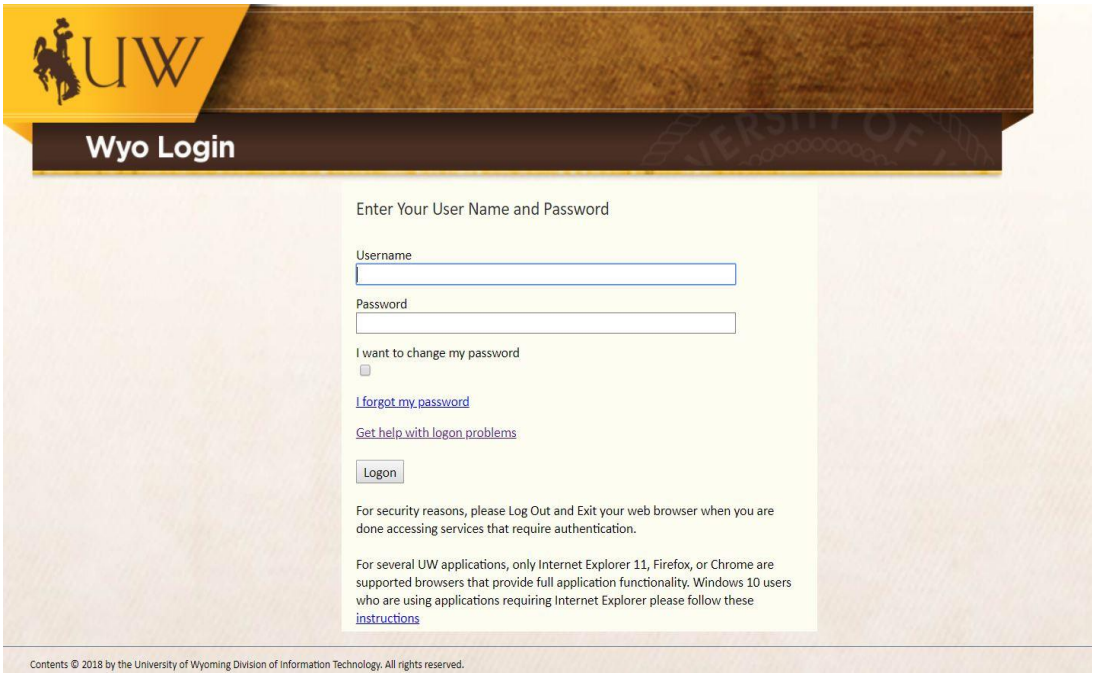


Figure 2.4.1.1: Login prompt after the UW Brown & Gold Report link was clicked.

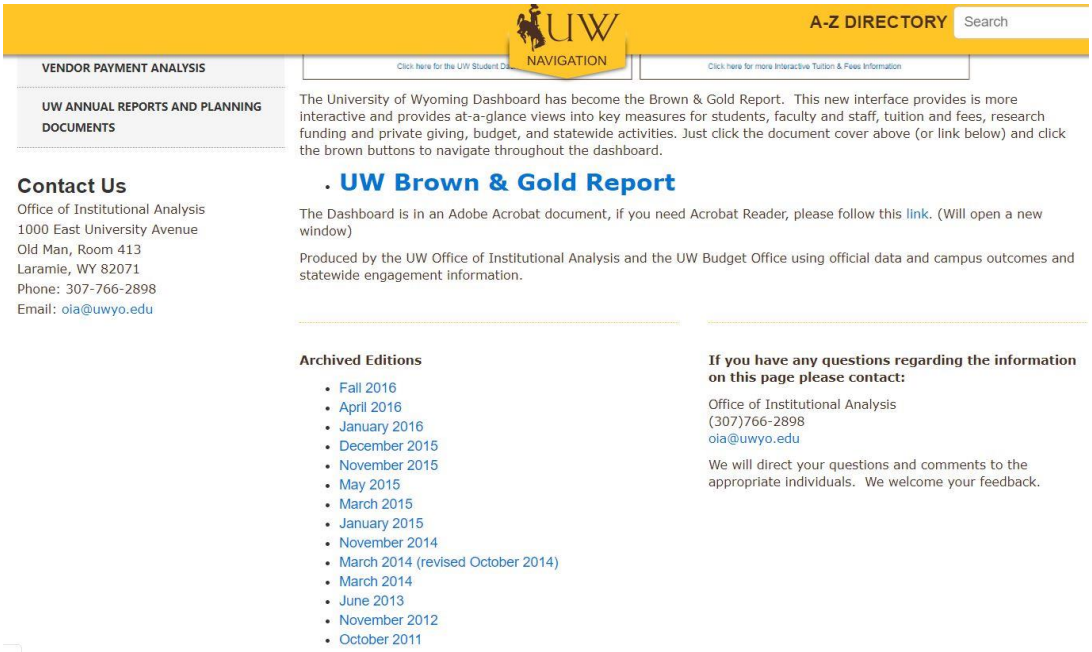


Figure 2.4.1.2: The dashboard is available to download in pdf.

2.4.2 Strength

In order to match with the colour theme of UOW, the colour of the dashboard was mainly comprised of brown and gold with a touch of black, blue and yellow, The dashboard **design was aesthetically pleasing**, despite the usage of different colour is high, the **differences in the colour of is meaningful as each one has a label below** it. Each of the dashboard was displayed in one full page thus **showing all the information in one glance without the need of scrolling** unless the user wish to view the next dashboard. Besides, with the dashboard being created as a PDF, **it can be printed or emailed** which **promotes flexibility and convenience.** Moreover, it **does not require internet access** to view the dashboard.

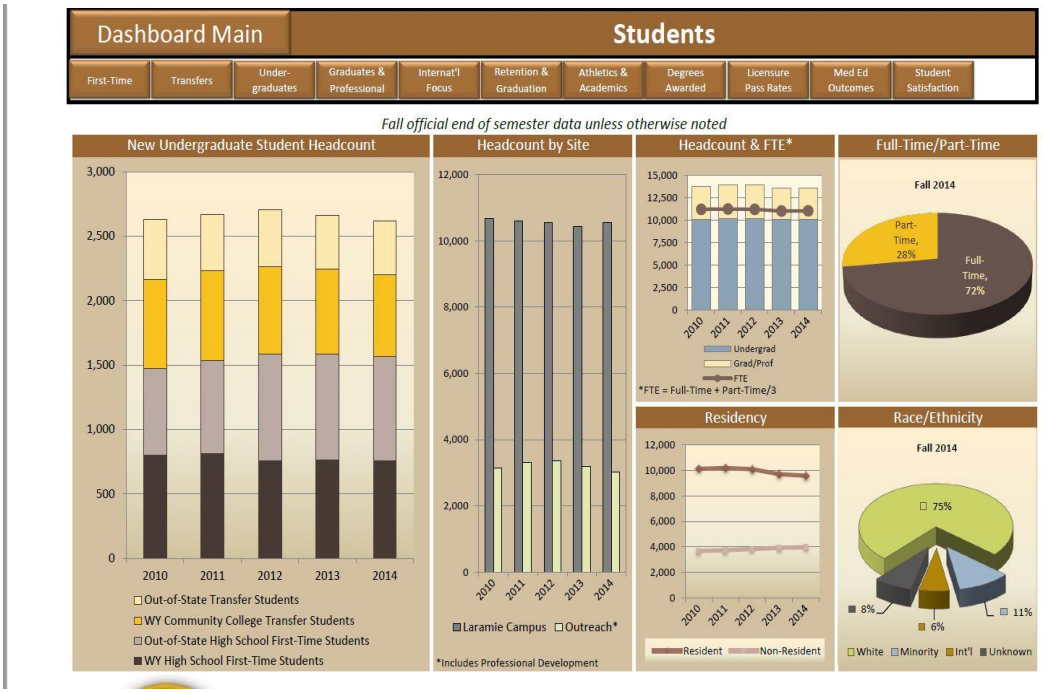


Figure 2.4.2.1: The dashboard is able to view it in one glance

2.4.3 Weakness

This UOW dashboard is **static, it does not have any interactive features.** The users are unable to view any detailed information even the user hovers over them. Needless to say, slicers and buttons are not available for the user to only show desired results, the user have to view the piece of information in whole. The colour used for this dashboard is **not colour-blind friendly,** especially brown colour is being used extensively all over the dashboard including the interface and the graph. Certain graphs are **unable to show**

data properly due to bad choice of chart type. The chart is too densely packed, the users could not really see the difference of the data anymore, this causes the dashboard to lose its main purpose that is to tell a story without context.

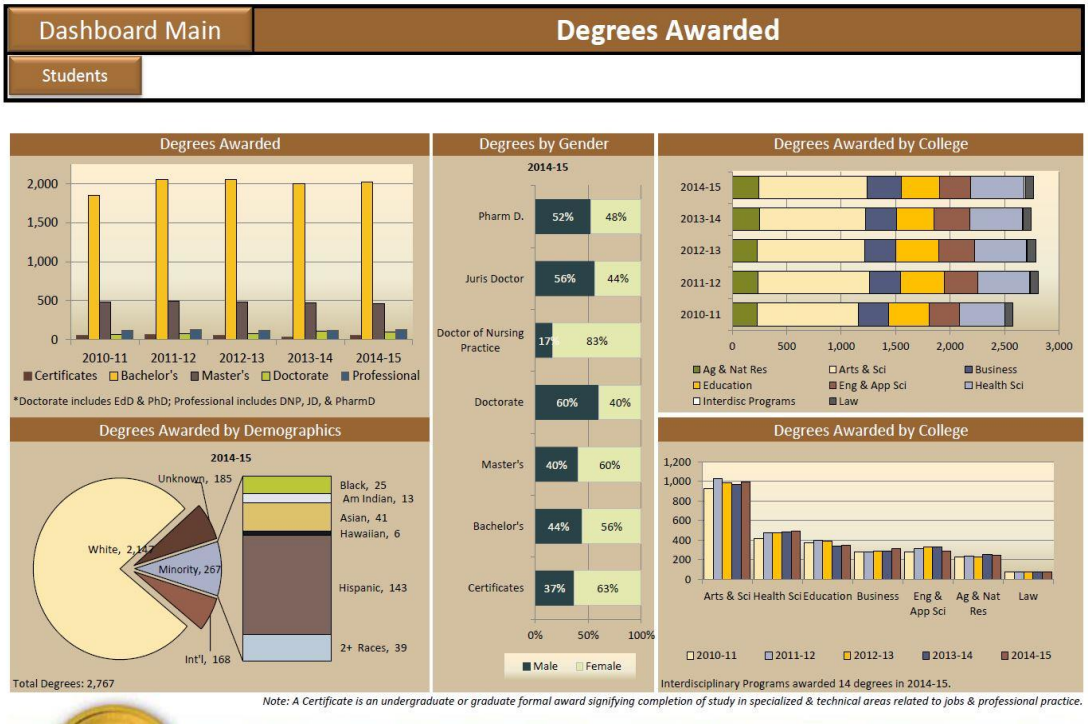


Figure 2.4.3.1: The bar chart is too densely packed and does not really show the trend

2.4.4 Recommendation

UOW should consider **opening up their latest dashboard to the public** so that people can make better decisions on whether to study at University of Wyoming, so that the dashboard people using it is interactive and able to convey the message even more clearly. It is suggested that UOW to **add a colour-blind friendly drop-down box** to their dashboard. It is best to set the colour palette to red/green paring for the majority of the user and swap the green to blue once it is in Colour-Blind Friendly mode. (Wexler, Shaffer and Cotgreave, 2017)

2.5 University of California, Los Angeles Dashboard

2.5.1 Brief

University of California – Los Angeles, or commonly known as UCLA, it is a public institution established in 1919, one of the prestigious universities in the world, it holds the record of most applied university in the United States. Undergraduates takes up to 70% of the student body, there are more than 20% of international graduate students in UCLA. The university undergraduate studies splits into 5 different divisions: letters and science, arts and architecture, engineering and applied science and theatre, film and television. The university has always receive an average of $1 billion for research annually for 11 years since then with more than 350 research labs, which explains why UCLA has always maintain such high ranking in the world, as research accounts most of the ranking criteria.

UCLA understands how important data driven decision making as the chancellor's office itself is a place for data analytics which have basic dashboard for public to access. However, if you wish to view more detailed data, you have to be a UCLA student or staff which is a very good way to protect sensitive information. There is a email prepared for those people who are interested of the data and would like to perform more in depth analytics are allowed to send an email to request for permission for the data.

UCLA Dashboard is created using Tableau, an interactive data visualization software, the best part of Tableau software is that it does not require technical skills to operate which explains why it is quite expensive to use Tableau services, Despite their price, the tool still garnered interest among people from each sectors such as business, university and researchers.

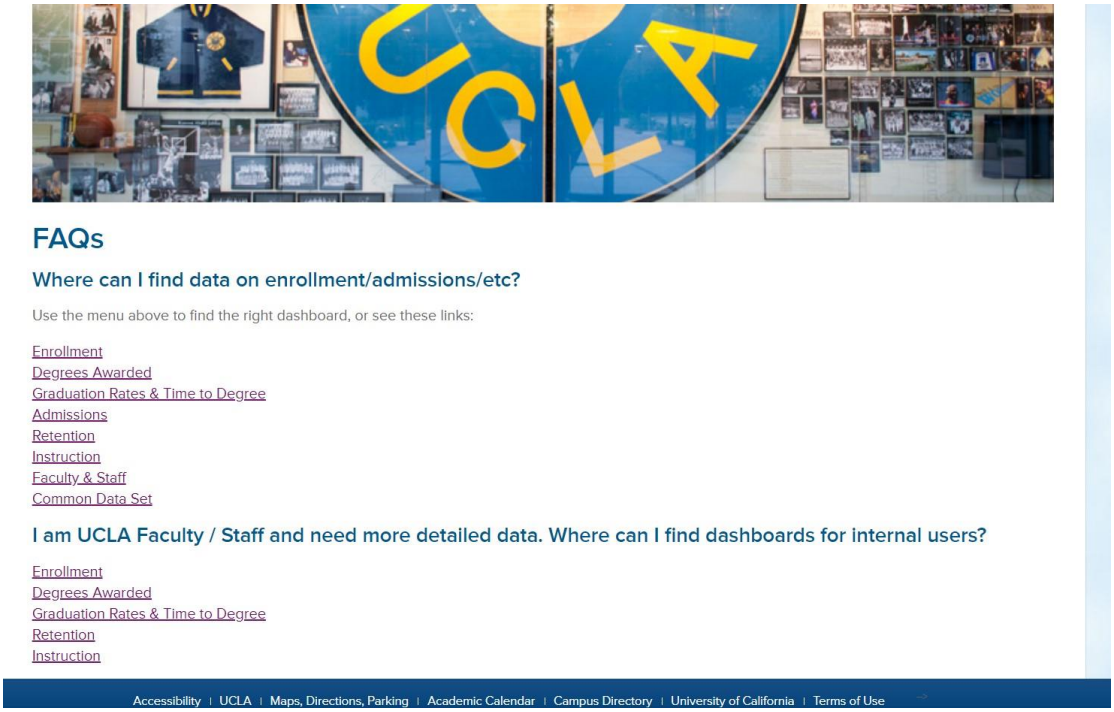


Figure 2.5.1.1: UCLA Dashboard Webpage

2.5.2 Strength

The design of the dashboard was undeniably simple and user friendly, the best part is that it includes a tab known as historic overview in some of their dashboard which **allows user to browse through old data** ranging from 1920 to 2019. This shows that UCLA turn their old paper documents into digital version, keeping their data in check in a well-maintained database would ensure no data breaches and leakage. The user **interactivity of the dashboard is good especially the hover over data is brief and concise**. Apart from that, the data is available to download as TXT file, PDF, Image and PowerPoint which is very convenient.

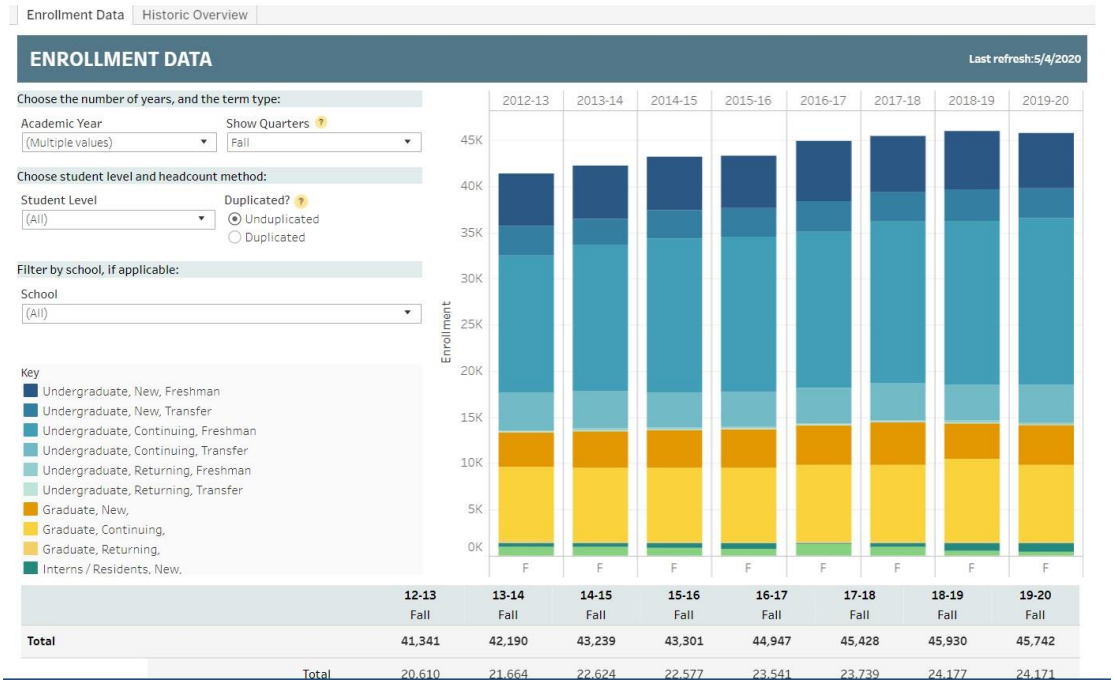


Figure 2.5.2.1: UCLA Enrolment Dashboard

2.5.3 Weakness

The **dashboard is not colour blind friendly**, even people without colour blind have to check twice to confirm the category of the data, especially the blue colour part. If the browser was minimized, the dashboard does not follow the size of the browser, which makes it **not responsive to the user**.

2.5.4 Recommendation

Overall, the UCLA dashboard was well developed, since it is one of the best universities in the world. The only improvement is that it needs to account for colour-blind users, a **colour-blind button** was suggested to add in as another feature.

2.6 Universiti Tunku Abdul Rahman Research & Development Annual Report

2.6.1 Brief on UTAR Institute of Postgraduate Studies and Research

IPSR acts as a central coordinating body for both postgraduate programmes and research and development activities inside UTAR. IPSR plays a major role in today's UTAR Ranking, it is a place for researchers to retrieve information on related research, regulatory and legal requirements, publications, etc. Since research activities would often involve significant resource, this unit focuses on seeking sources of external funding for research projects and monitoring applications for funding. UTAR funding comes from public sector, private sector and international agency sources. The number of postgraduate graduands and research output have been increasing over the years and it is a clear sign for UTAR is heading the right way to become a leading centre of learning and research.

Despite UTAR do not have a proper dashboard to monitor on the publication and teaching performance of the academic staff, UTAR was able to rank 501-600 in the latest Times Higher Education 2021, imagine all the possibilities if UTAR have equipped itself with a monitoring dashboard. Since UTAR is a non-profit organization, it would come up with research & development annual report to show the public on how UTAR spends all the funding given by other organization, which promotes transparency to their funders, communities and students.

Figure 2.6.1.1: UTAR External Funding Report 2018

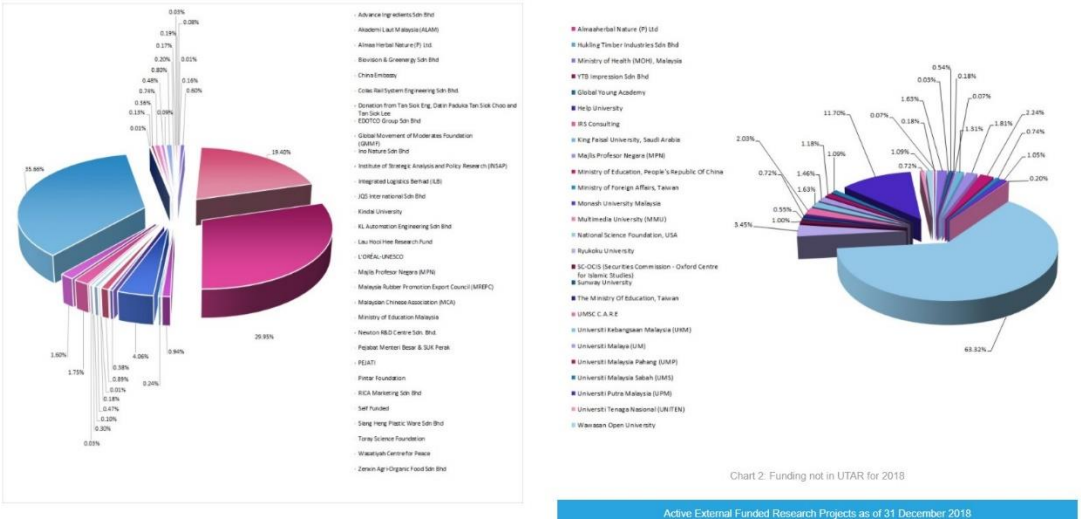


Figure 2.6.1.2: UTAR External Funding Pie Chart 2018

2.6.2 Strength

The **publication report of UTAR is responsive if minimized**, this is important for a website as users might have different screen size, the content was automatically scaled and match the screen size so that users do not need to do extra work to view the report content which indirectly **improves user experience** an analysis report was written based on the data. Besides, the **list publications produced by UTAR academic staff is available to download** in PDF format, especially the report in both external and internal funding, the amount of money and the name of the research was statement clearly. Since the list of publication is available in PDF format, users do **not need internet connection** to view the data once it is downloaded.

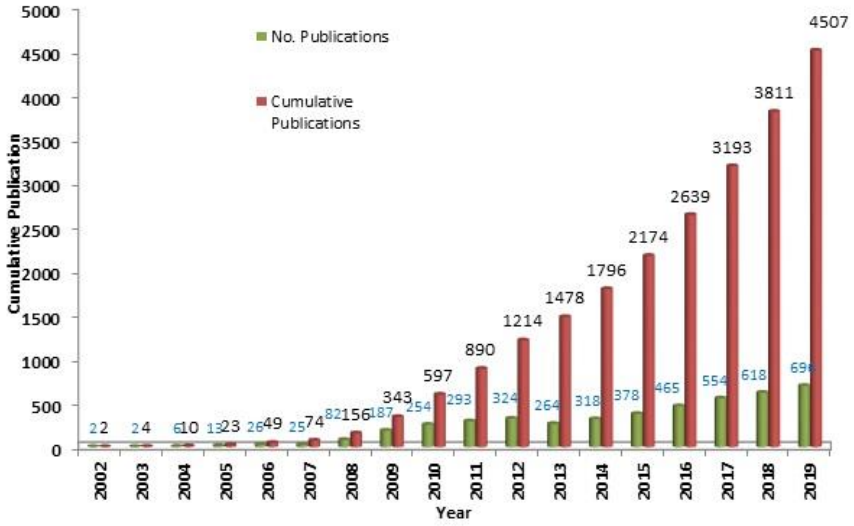


Figure 2.6.2.1: UTAR Publication Bar Chart 2019

2.6.3 Weakness

Since it is an annual report, all of the charts are generated and then uploaded as JPEG image, there are **no filter available** for the user to look at certain category of the data causing the whole annual report chart **lack of user interactivity**. Besides, the **poor choice of graph in showing trend of data**, the UTAR publication uses bar chart and did not account for the standard deviation of the data. The users can barely see the difference of small values like in the year 2002 to 2006. Apart from that, **UTAR uses the same design and same graph for the past few years** without changing other method, the only thing changed was the value of the data.

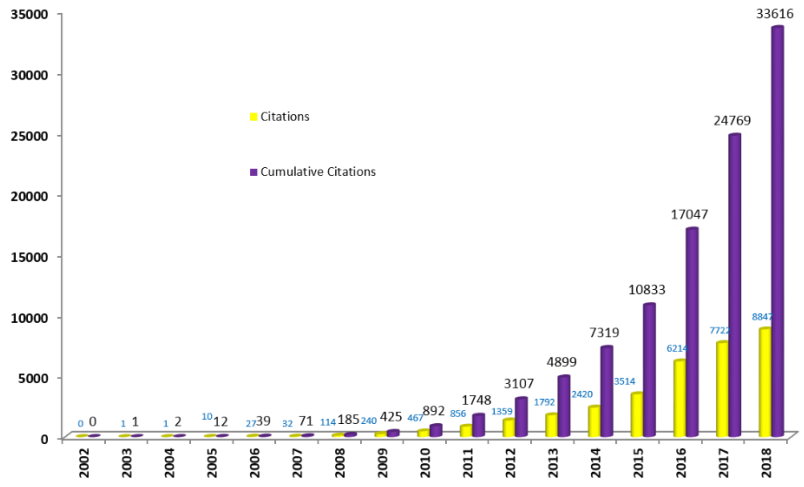


Figure 2.6.3.1: UTAR Citation Bar Chart 2018

2.6.4 Recommendation

UTAR are advised to adopt data visualization tool to create interesting dashboards, there are expensive tools such as Microsoft Power BI or Tableau or UTAR can opt for free tools such as RShiny or Plotly Dash. This way both the board of management can gain insight and make better decision while the user who view the report are able to look at the data with more user interactivity and detailed. Besides, in order to perform trend analysis, line graph would be the better candidate compared to bar, as you can see in the Figure 2.6.3.1 above, users can barely see the difference from 2002 to 2010.

2.7 Sentiment Analysis on Teaching Survey

Most of the universities, including UTAR are using evaluation questionnaire as an tool for data collection, the teaching survey consists of quantitative and qualitative parts. The quantitative data was collected by close-ended questions, whereby the students rate their lecturers based on multiple choice questions, while the qualitative data was collected by open-ended questions such as comments and suggestions given by the students in textual form. Faculties generally having difficulties in understanding students' written feedbacks in teaching survey. Despite such open-ended comments are usually the core purpose of the whole teaching survey, the feedback sometimes is rich with observations and insights, faculty administrative often struggles to draw conclusions from it. Undeniably, there are certain comments that are biased, contradictory, half of the students agrees while the other half says otherwise, this can leave faculty members baffled and causing them to think that it is impossible to please everyone. As a result, the faculty often ignore the all of the written comments, which might include the important messages written by the students.

Sentiment analysis is a utilization of natural language processing, computational linguistics and text analytics that classifies and return sentiment polarity from the text by studying the opinion. A sentiment polarity is normally either positive or negative, there are neutral as well. (Altrabsheh, Cocea and Fallahkhair, 2014).

Zied Kechaou et al(2011) created a sentiment classification of e-learning blogs and forums using a supervised hybrid technique that integrated hidden Markov models with support vector machines. Three selection methods were used and chi-statistics method was the best method compared to mutual information and information gain.

Nabeela Altrabsheh, et al (2014) implemented sentiment analysis of student feedback using Naïve Bayes, complement Naïve Bayes, SVM and maximum-entropy classifiers using unigrams as features. The results were SVM and complement naïve bayes having higher accuracy in real-time feedback analysis, the algorithm works best without including the neutral polarity in the data.

Guadalupe Gutierrez Esparza et al (2016) created a model for sentiment analysis on student tweets about teaching performance in Spanish. SVM algorithm was used to classify the tweets into positive, negative and neutral, in order to compare their results, a syntactic pattern model was proposed using SVM and syntactic patterns.

V.Dhanalakshmi et al (2016) carried out sentiment analysis on feedbacks from a teaching survey of Middle East College in Oman. They have utilised RapidMiner tool to classify the comments into positive and negative polarity based on features like teacer, exam, module, content and resources. Comparison was made using Naïve Bayes, SVM, k-nearest neighbours and neural-network classifiers.

Balahadia et al (2016) developed a teacher's performance evaluation tool using opinion mining with the help of sentiment analysis, the tool was used to evaluate teacher performance based on teaching survey in both English and Filipino language. The sentiment values are calculated from the qualitative and quantitative response ratings using naïve bayes algorithm and then was plotted into graphs to help university administrators to take student feedback seriously.

\

CHAPTER 3: SYSTEM DESIGN

3.1 System Overview

The development of a citation, publication and teaching performance dashboard is based on CRISP-DM data mining. In general, CRISP-DM data mining methodology is a hierarchical model consists of four levels of abstraction which is, *phase, generic task, specialized tasks* and *process instance.*

The development starts from the top level which is the *phase*. All kinds of possible tasks were planned and written down in this phase, starts with selection of the abstract and citation database and ends with deployment of the dashboard. Each *phase* is comprised of multiple second-level *generic* tasks. The generic tasks have to be complete and stable as possible. In the *generic* tasks level, the tasks are brief and enough to satisfy all possible data mining situations. For instance, data selection is one of the generic tasks, the data have to be complete and not having too much missing values that is beyond the ability to be cleaned. Stable means that the data model should be able to cater certain unexpected developments like new modelling techniques. Moving on to the third level, *specialised task,* this is the place to give all of the *generic tasks* an in-depth explanation and actions to be carried out in specific situations. Take 'data cleaning' as an example, the third level would specify the details of carrying out this task such as cleaning numeric values or categorical values and whether the problem is clustering or predictive modelling. In the final level, the *process instance,* is the final depiction of all the actions did in second level and third level, the data visualization part. The dashboard will be completed at this level.

3.2 Methodology

The methodology that best suits to develop the dashboard in this project is Cross-industry standard process for data mining, usually known as CRISP-DM.

The CRISP-DM methodology defined in terms of hierarchical process model, it involves six levels of process: *Business Understanding, Data Understanding, Data Preparation, Modelling, Deployment* and *Evaluation.* In some cases, it hinges on the outcome of each phase whether it is suitable to proceed to the next stage or going back to the previous.

The larger outer circle represents the cyclic nature of the data mining process. Data mining process is not completed just because a solution is being deployed. The information gained during the development process often triggers new and focused business questions. The continuation of data mining process profits from the previous data mining experience. (Hipp & Wirth, 2000)

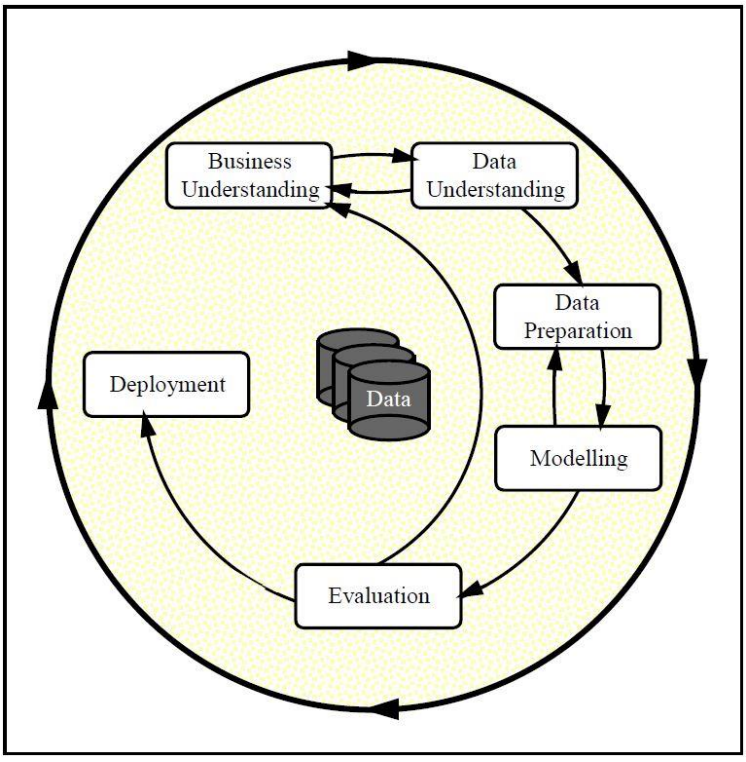


Figure 3.2.1: The CRISP-DM Methodology (Hipp & Wirth, 2000)

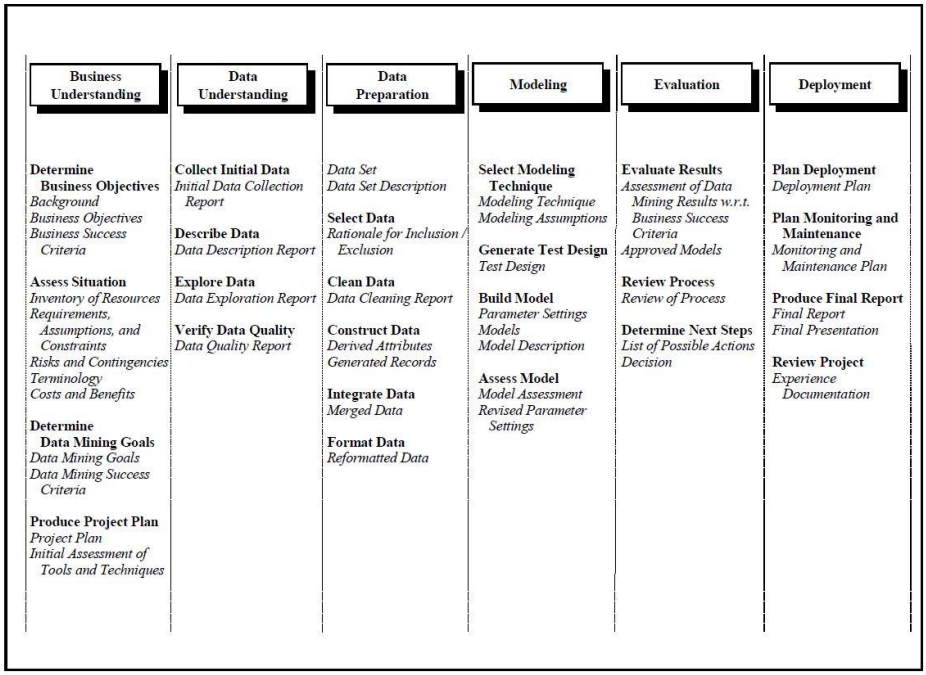


Figure 3.2.2: Overview of the CRISP-DM tasks and their outputs (Hipp & Wirth, 2000)

The CRISP-DM focus on the problems by determining a process model which is mutual exclusive of both the industry sector technology used. The CRISP-DM process model intend to develop large data mining projects with lesser costs, more reliable, easier to manage and faster. The application of the CRISP-DM to this project will be discussed further in the following subsection.

3.2.1 Business understanding

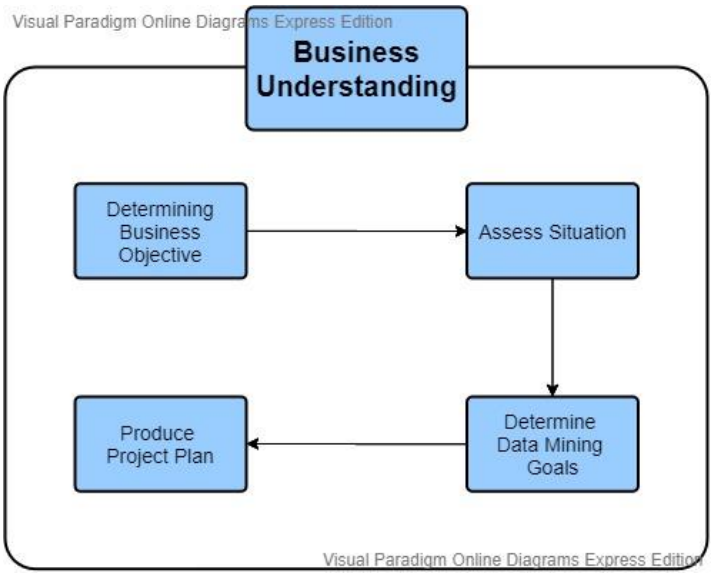


Figure 3.2.1.1: Business Understanding

This phase focuses on recognising the business domain, ensure the capability to detecting possible issues from the business domain viewpoint and eventually transforms such business issues into a data analytics problem.

Determine Business Objective

As the project objective mentioned in Chapter 1, the first objective of this project is to develop a dashboard to visualise publication data and teaching survey data, the publication data is planned to be retrieved via Elsevier Developer API while the teaching survey data will be provided by the supervisor of this project. Since most of the university currently are still relying on Scopus to monitor on their publication and citation, an interactive university owned dashboard is recommended to assist the university in making better decision based on the data collected. Student feedback in textual form are often ignored by the faculty because of the contradicting feedbacks given by the students, hence sentiment analysis on the student feedback was suggested to improve how the teaching survey being presented to the lecturers, instead of just showing lines of comments without any text analysis.

Access Situation

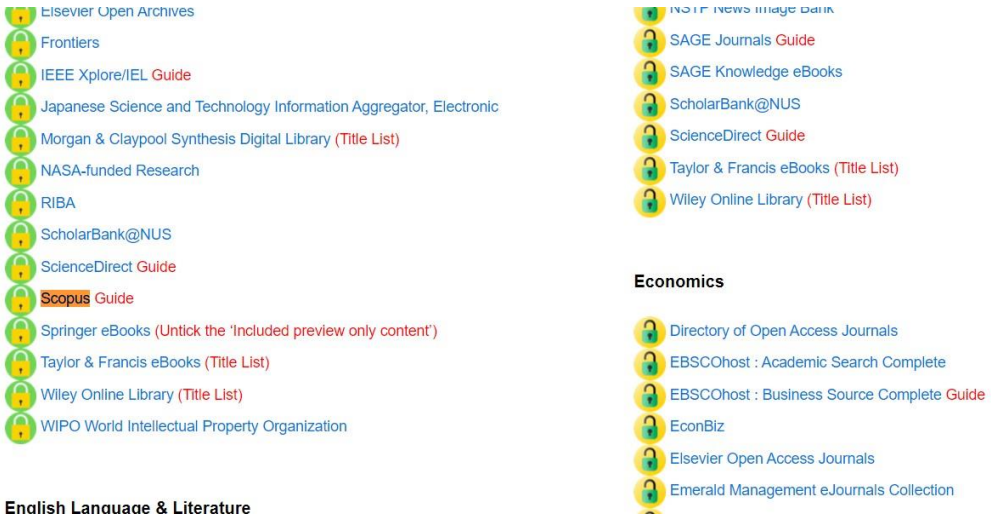


Figure 3.2.1.2: UTAR is Scopus Subscriber

First of all, data collection only works if UTAR is a subscriber of Scopus, the figure above shows all of the e-databases UTAR have subscribed, it requires UTAR authentication to gain access to all of the databases provided at UTAR e-database portal.

According to the Elsevier Developers documentation, Elsevier distinguishes the user as subscriber or non-subscriber by the IP address from which the requests made to the API. For example, if the user wishes to access the Abstract Retrieval API, the user has to make sure the request comes from within a network that is already configured for access to Scopus.

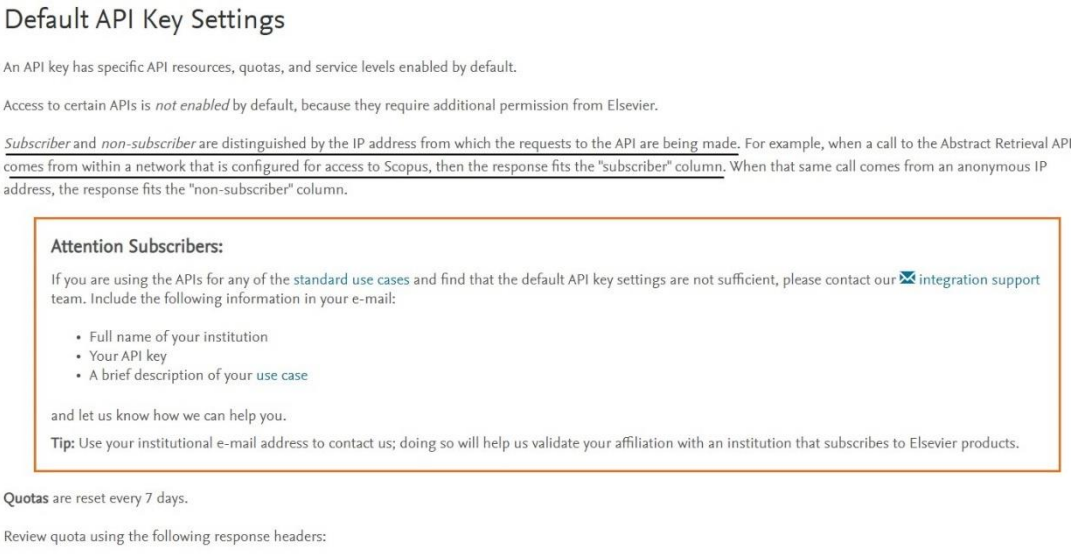


Figure 3.2.1.3: Scopus identifies IP Address

Since Scopus API is a paid subscription service, unauthorised access from unknown API is limited to only retrieving serial title, the API required for this project is Affiliation Retrieval and Author Retrieval, which both only available for subscriber only, even for subscribers there are limits to retrieve the data, which in this project is only allowed to retrieve a maximum 5000 data weekly with a limit of 6 requests per second, which is more than enough for this project since the number of academic staff in UTAR did not exceed 5000.

Scopus APIs

#	API Name	Enabled or Disabled	Non-subscriber	Subscriber	Weekly Quota	Requests/second
1	Serial Title	Enabled	STANDARD, COVERIMAGE views / Default 25 results / Max 200 results	STANDARD, COVERIMAGE, ENHANCED Default 25 results / Max 200 results	20,000	6
2	Citations Count Metadata	Disabled	N/A	STANDARD view / Default 25 results / Max 200 results	50,000	24
3	Citations Overview	Disabled	N/A	STANDARD view / Default 25 results / Max 200 results	20,000	6
4	Subject Classifications	Enabled	No restrictions	No restrictions	N/A	N/A
5	Abstract Retrieval	Enabled	META view	All views, default FULL view	10,000	9
6	Affiliation Retrieval	Enabled	N/A	All views, default STANDARD view	5,000	9
7	Author Retrieval	Enabled	N/A	All views, default STANDARD view	5,000	6
8	Affiliation Search	Enabled	N/A	Default 25 results / Max 200 results 5000 item result limit	5,000	6
9	Author Search	Enabled	N/A	Default 25 results / Max 200 results 5000 item result limit	5,000	6
10	Scopus Search	Enabled	STANDARD view / Default 25 results	STANDARD view / Max 200 results	20,000	9

Figure 3.2.1.4: Scopus API quotas and condition

Determine Data Mining Goals

The goals of data mining for this project are to retrieve data from the Scopus API to be visualised using Plotly Dash and retrieve all of the textual data from teaching survey, convert it into string format to perform sentiment analysis on the data.

Project Plan

The project plan will be explained in depth later at the timeline section of this chapter,

3.2.2 Data understanding

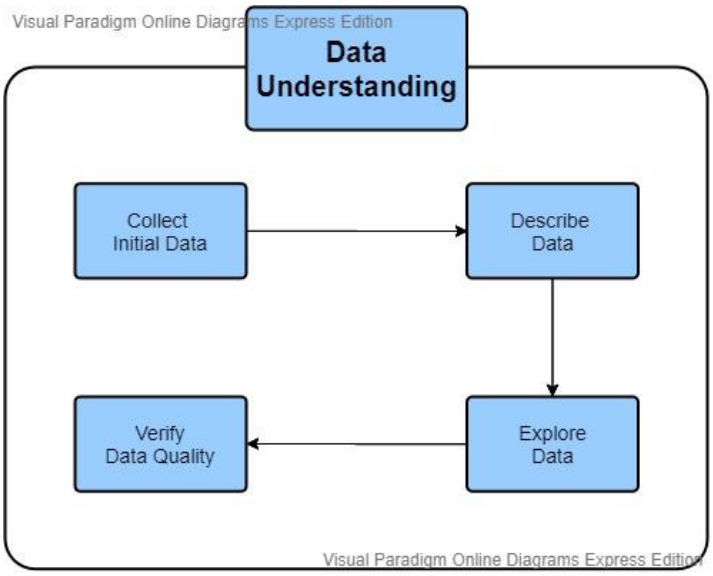


Figure 3.2.2.1: Data Understanding

This phase involves a set of the necessary data and their initial visualization/summarization in order to gain the first insights, particularly but not exclusively about data quality problems such as missing data or outliers.

For this phase, most of the explanations will be on how to work on the Scopus API to get the data required. Scopus validates its user by respective IP address from which requests to the API being made. For example, when a call to the Subject Classification API came from a network that is configured for access to Scopus, then the response is classified as the "subscriber". In contrast, when the same call comes from an anonymous IP address, it is then classified as the "non-subscriber". An API key is required to gain access to different types of Scopus API, the API for this project is author retrieval API.

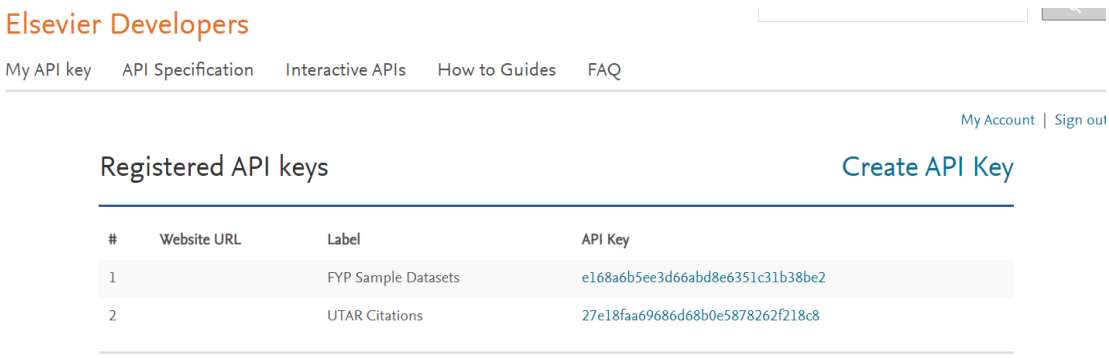


Figure 3.2.2.2: The registered API key from Elsevier Developers

Collect Initial Data

Publication Data Collection

Moving on to the author retrieval API, there will be limited access for those who are not subscribed to Scopus, while subscribers gain full access to the data which was required in this project.

Author Retrieval Views

Access to views or fields marked with an * may be restricted due to entitlements

Field	BASIC	METRICS	LIGHT	STANDARD *	ENHANCED *
dc:identifier *(Author identifier)*	X	X	X	X	X
eid *(Electronic Identifier)*	X		X	X	X
orcid *(Author identifier)*	X		X	X	X
link ref=scopus-author *(Scopus author details page URL)*	X		X	X	X
link ref=self *(Content Author Retrieval API URI)*	X		X	X	X
prism:url *(Content Author Retrieval API URI)*	X	X	X	X	X
link ref=search			X	X	X
document-count *(Number of documents authored)*		X	X	X	X
cited-by-count *(Number of citing documents)*		X	X	X	X
citations-count *(Number of citations)*		X	X	X	X
preferred-name **surname**			X	X	X

Figure 3.2.2.3: Author Retrieval Views for data collection

According to the figure above, the non-subscribers can only access to elements like dc-identifier, eid, prism-url which are totally irrelevant to the project objective. Only document-count, cited-by-count, surname, given-name, affiliation-name, h-index and co-author-count are needed for this project with the condition of retrieving under UTAR network. However, there are major difficulty in retrieving the data, further details will be shared in the implementation issues and challenges subsection.

Since the author retrieval view does not involve faculty, department and gender, then the data will have to be added manually in order to be able to filter the data. First of all, the HTTP requests URL would require unique scopus author_id, every person that has publication will have a unique id, to retrieve all of the author_id that belongs to UTAR academic staff, the faculty,department and gender is available to retrieve at UTAR staff directory.

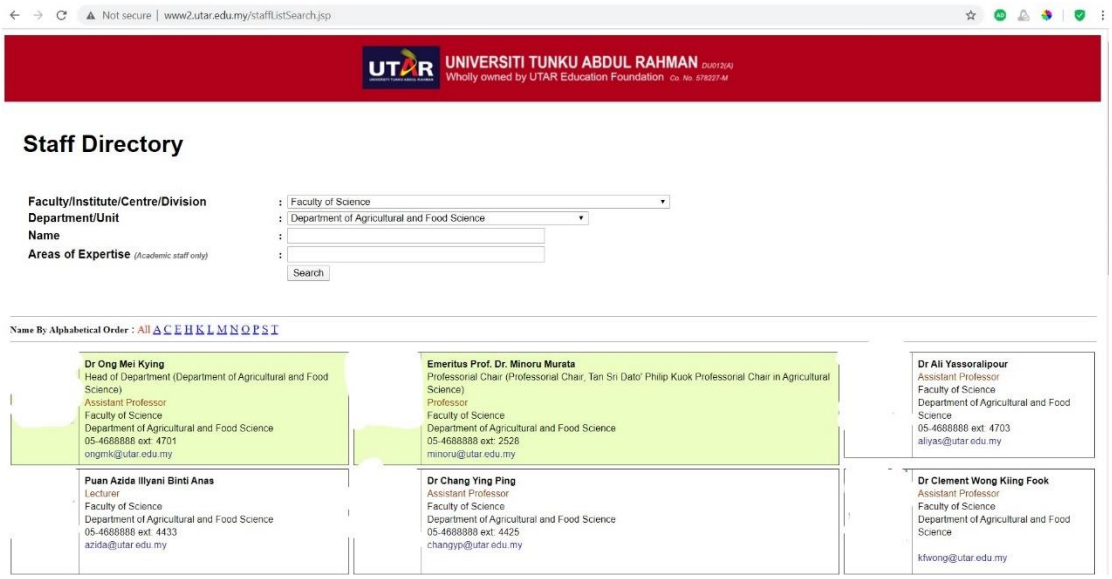


Figure 3.2.2.4: UTAR Staff Directory

Once all of the academic staff names, gender, faculty, department is collected, it is then searched at Scopus Author Search.

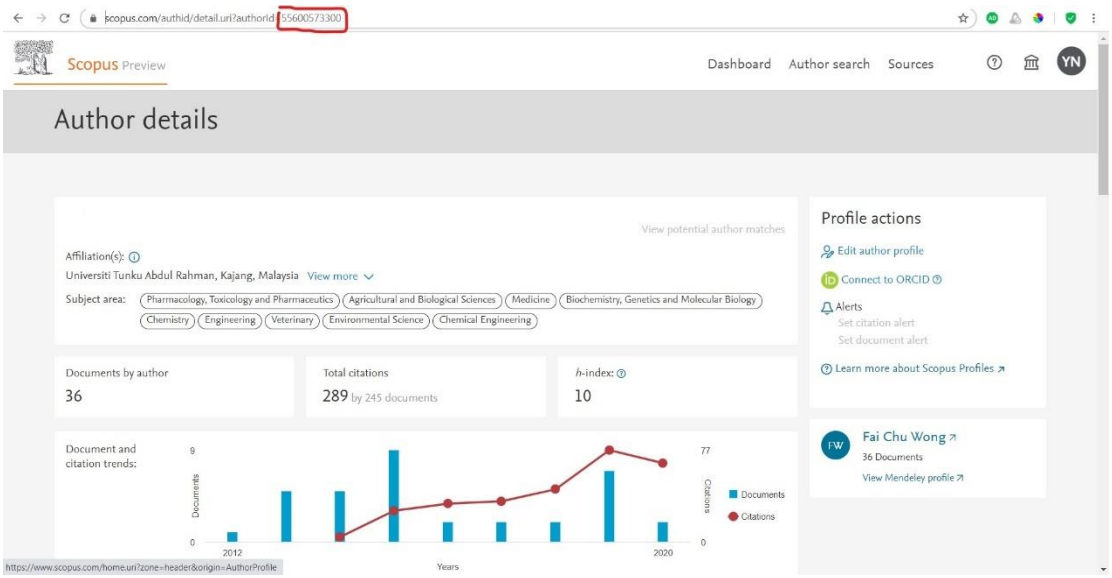


Figure 3.2.2.5: Scopus Search Academic Staff

The unique author_id will be available at the URL link, which is then recorded in a text file in order for us to replace each of the author_id in the data retrieval via Scopus API. Initially, the Scopus API is able to return all of the author_id from Universiti Tunku Abdul Rahman, but this way would require to identify whom is the author_id belongs to, which is extremely tedious, more details on this issue will be mentioned in implementation issues and challenges.

```
FSc Faculty of Science

        Department of Agricultural and Food Science
        Ong Mei Kying 55508517000 female
        Ali Yassoralipour 28268200000 male
        Chang Ying Ping 55370572400 female
        Ee Kah Yaw 17342218800 male
        Heng Mei Hsuan 52663572700 female
        Lye Huey Shi 25230350100 female
        Ooi Ai Lin 35264402100 female
        Ooi Aun Chuan 57202535414 male
        Phuah Eng Tong 25630907100 male
        Sim Kheng Yuen 57188697770 male
        Tan Yen Nee 57191989815 female
        Teo Kah Cheng 37010013500 female
        Teoh Chaiw Yee 36998160600 female


        Department of Allied Health Sciences
        Sit Nam Weng 57200646500 male
        Chee Huei Phing 56340522800 female
        Annalechumy a/p Loganathan 56963917000 female
        Chew Choy Hoong 7102800372 female
        Kokila a/p Thaigarajah 57190838841 female
        Michelle Ng Yeen Tan 16158067100 female
        Ng Kock Wai @ Tony 7402229714 male
        Ng Wen Jie 57044038700 male
        Sangeetha a/p Arullappan 36619684200 female
        Sinouvassane Djearamane 57202375412 male
        Tan Chin Xuan 57195587723 male
        Teh Lai Kuan 54384268900 female
```

Figure 3.2.2.6: Collected Academic Staff Scopus ID

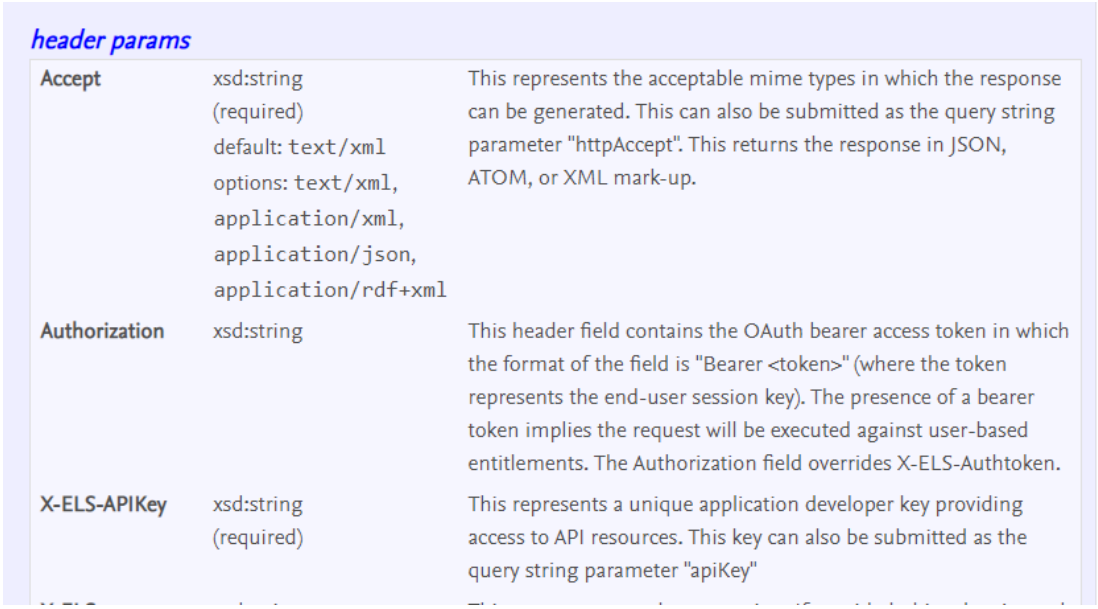
The process is repeated until every academic staff that have publication is recorded inside the text file, the number next to lecturer's name is the unique author_id of Scopus which will be used later in data retrieval.

Initially, the data retrieval from the Scopus API was using elsapy, a python module to interact with all of the data from Elsevier, but it is able to return only 1 data of a particular author per request, it is not enough to fulfil the purpose of this project. In the end, a simple HTTP library from Python known as 'requests' was utilised to capture the data.

```python
import requests
import json

with open('C:/Users/User/Anaconda3/Scripts/ScopusApi/elsapy-master/scopus_sample_author.json') as access_json:
    read_content = json.load(access_json)

save_data = []
for author in read_content:
    for author_access in author:
        for author_data in author_access:
            authid = author_data['authid']
            save_data.append(authid)


def getAuthorInfo():
    url = ("https://api.elsevier.com/content/author/author_id/36629511500?field=h-index,document-count,cited-by-count,citations-count,coauthor-count,surname,given-name,af
    resp =  requests.get(url, headers= {'Accept' : 'application/json',
                                        'X-ELS-APIKEY': '27e18faa69686d68b0e5878262f218c8' })
    print(json.loads(resp.text.encode('utf-8')))

    with open("scopus_author_data_36629511500.json", "w") as write_file:
        json.dump(resp.json(), write_file)

    return json.loads(resp.text.encode('utf-8'))

getAuthorInfo()
```

Figure 3.2.2.7: Coding for data retrieval

The API key will be placed inside the header parameter of the request function, the parameters used was ‘accept’ and ‘X-ELS-APIKey’. ‘Accept’ determines which format will the response being generated, there are 5 different choices, application/json was chosen for this project. The ‘X-ELS-APIKey’ accepts unique key in string format to gain access to API resources.

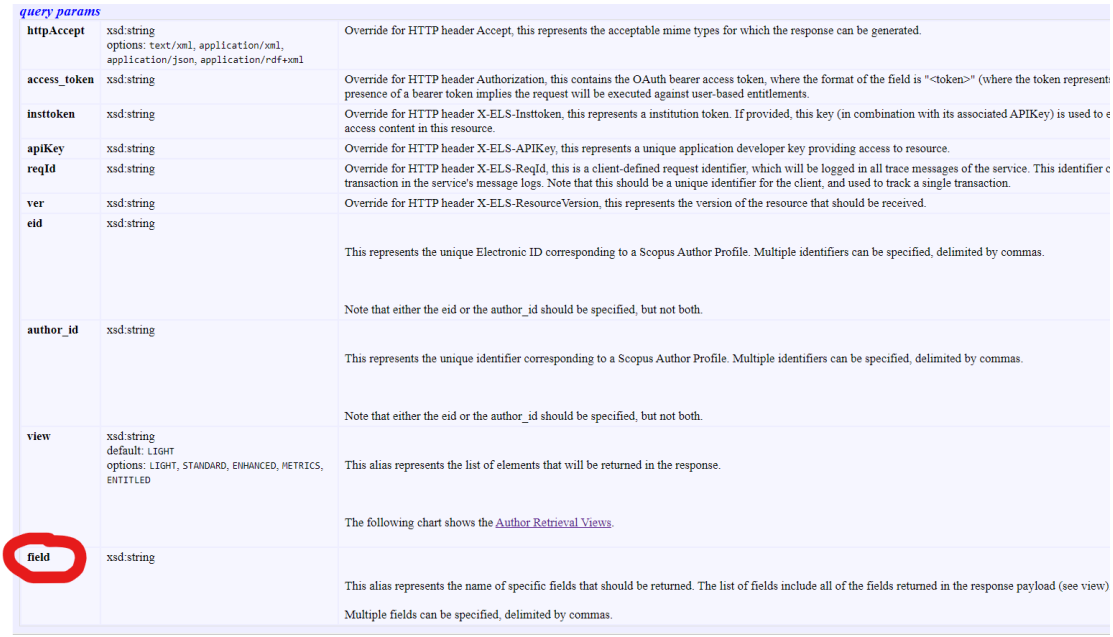


header params

Accept	xsd:string (required) default: `text/xml` options: `text/xml`, `application/xml`, `application/json`, `application/rdf+xml`	This represents the acceptable mime types in which the response can be generated. This can also be submitted as the query string parameter "httpAccept". This returns the response in JSON, ATOM, or XML mark-up.
Authorization	xsd:string	This header field contains the OAuth bearer access token in which the format of the field is "Bearer <token>" (where the token represents the end-user session key). The presence of a bearer token implies the request will be executed against user-based entitlements. The Authorization field overrides X-ELS-Authtoken.
X-ELS-APIKey	xsd:string (required)	This represents a unique application developer key providing access to API resources. This key can also be submitted as the query string parameter "apiKey"

Figure 3.2.2.8: Header parameters in author retrieval documentation

The URL 'https://api.elsevier.com/content/author/author_id/{author_id}' is a get method sending request to the author retrieval API. All of the author-id for each of the respective lecturers are manually recorded using default author search bar. The {author_id} was replaced with different id collected previously until all of the author_id was retrieved.



query params

httpAccept	xsd:string options: `text/xml`, `application/xml`, `application/json`, `application/rdf+xml`	Override for HTTP header Accept, this represents the acceptable mime types for which the response can be generated.
access_token	xsd:string	Override for HTTP header Authorization, this contains the OAuth bearer access token, where the format of the field is "<token>" (where the token represent presence of a bearer token implies the request will be executed against user-based entitlements.
insttoken	xsd:string	Override for HTTP header X-ELS-Insttoken, this represents a institution token. If provided, this key (in combination with its associated APIKey) is used to e access content in this resource.
apiKey	xsd:string	Override for HTTP header X-ELS-APIKey, this represents a unique application developer key providing access to resource.
reqId	xsd:string	Override for HTTP header X-ELS-ReqId, this is a client-defined request identifier, which will be logged in all trace messages of the service. This identifier c transaction in the service's message logs. Note that this should be a unique identifier for the client, and used to track a single transaction.
ver	xsd:string	Override for HTTP header X-ELS-ResourceVersion, this represents the version of the resource that should be received.
eid	xsd:string	This represents the unique Electronic ID corresponding to a Scopus Author Profile. Multiple identifiers can be specified, delimited by commas. Note that either the eid or the author_id should be specified, but not both.
author_id	xsd:string	This represents the unique identifier corresponding to a Scopus Author Profile. Multiple identifiers can be specified, delimited by commas. Note that either the eid or the author_id should be specified, but not both.
view	xsd:string default: `LIGHT` options: `LIGHT`, `STANDARD`, `ENHANCED`, `METRICS`, `ENTITLED`	This alias represents the list of elements that will be returned in the response. The following chart shows the Author Retrieval Views.
field	xsd:string	This alias represents the name of specific fields that should be returned. The list of fields include all of the fields returned in the response payload (see view) Multiple fields can be specified, delimited by commas.

Figure 3.2.2.9: Query parameters in author retrieval documentation

Moving on to the query parameters for the URL, according to the code provided above, the 'field' is one of the parameters provided in the Elsevier documentation. It is used to represents the names of specific fields that should be returned, multiple fields are allowed, just have to separate it using commas.

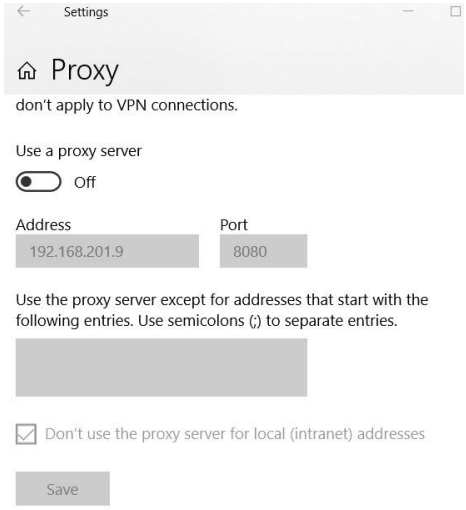


Figure 3.2.2.10: Proxy Configuration for Data Collection

The data was collected inside this project supervisor's room, the proxy IP address was configured to 192.168.201.9 with port number 8080 along with LAN cable of UTAR network connected to the laptop, this way the Scopus API recognised the connection to the API as subscriber and then the requested data was returned and saved as JSON file. Due to the data collection had to be done inside UTAR network, real-time data retrieval from the Scopus database is unable to be done. Hence, the data must be saved down as json file so that the data can be loaded into the dashboard with ease. This explains why the code above using json.dumps to store all of the data returned from the server into a file known as 'scopus_author_data'. There are several challenges during data collection inside UTAR network, further details will be discussed in the implementation issues and challenges.

Teaching Survey Data Collection

As individual lecturer evaluation analysis report is considered private and confidential, the sample piece of data was provided by the supervisor of this project. One of the future works of this project is to increase the quantity of the lecturer evaluation analysis report and perform sentiment analysis.

The data comes in PDF format, it will have to convert into string format to be able to perform sentiment analysis, this will be done later in data cleaning phase.

Describe Data

Publication Data Describe

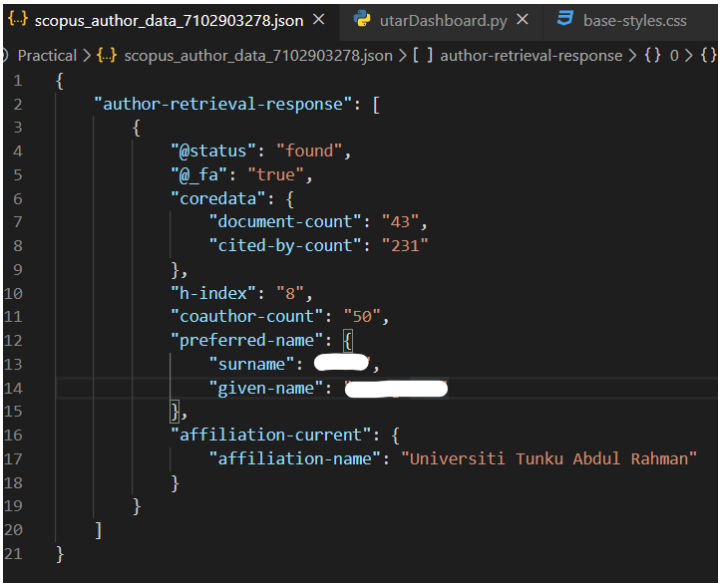


Figure 3.2.2.11: Sample uncleaned data

As Scopus is a paid subscription citation database, the data quality returned by the Scopus API was not clean as expected, there are unnecessary metadata such as '@status' and '@_fa' to be cleaned otherwise the data structure of the data would not fit into data frame. The attributes required was document-count, cited-by-count, h-index, co-author-count, surname and given-name, several columns of attributes would still need to be added to make the data meaningful such as faculty, department and gender. There is no need to show python commands like 'df.describe()' or 'df.info()', because the result is meaningless and would not help in understanding the data better. Besides, all of the data collected is separated individually and not consolidated in one JSON file, efforts on merging all the data will be done in the Data Preparation phase.

Teaching Survey Data Describe

```
#textract is a package that reads any file format and converts to bytes for text analysis
import textract
from string import punctuation
#The text is in bytes form
text = textract.process("C:/Users/User/Documents/TeachingSurvey_Prob.pdf")

text= text.decode(encoding="utf-8") #Convert bytes to string format

text
```

Figure 3.2.2.12: Teaching Survey Data bytes form (Data is Confidential)

In order to show the teaching survey data, textract a python package was used to read and convert the contents into bytes form, it is applicable to any documents in any format, the package is commonly used with the purpose to detect any useful information that is hidden inside Word documents, PowerPoint presenations, PDF, the content might be valueable for further textual analysis and visualization. The bytes form content is then converted to string format in order to print out to the jupyter notebook.

Explore Data

Data Exploration on Publication Data

Figure 3.2.2.13: Publication Data Messy Format

The data is unable to be read into data frame, since it involves a lot of unnecessary attribute and the JSON syntax is out of shape, the pandas data frame can only show what is under author-retrieval-response, data cleaning has to be made in order to explore the data properly.

Data Exploration on Teaching Survey Data

The teaching survey requires sentiment analysis to come up with new column of data. Hence, without data cleaning, the data is unable to be explored.

Verifying Data Quality

Data Quality of Publication Data

The data quality of publication data is considered above average since it does not involve empty data or missing values but only data refactoring is required since the structure of data is unorganized. The data attributes required for this project is available inside all of the individual academic staff data, which are document-count, cited-by-count, citations-count, h-index, co-author-count, surname, given-name. The problem of this data is that does not have columns like faculty, department and gender which is crucial in applying filter later in dashboard development, this could be resolved later in data cleaning stage.

Data Quality of Teaching Survey Data

Since the teaching survey data is all textual, additional columns of data have to be created using sentiment analysis and data cleaning in order to be able to plot the data to dashboard, which is also a part of data cleaning phase.

3.2.3 Data preparation

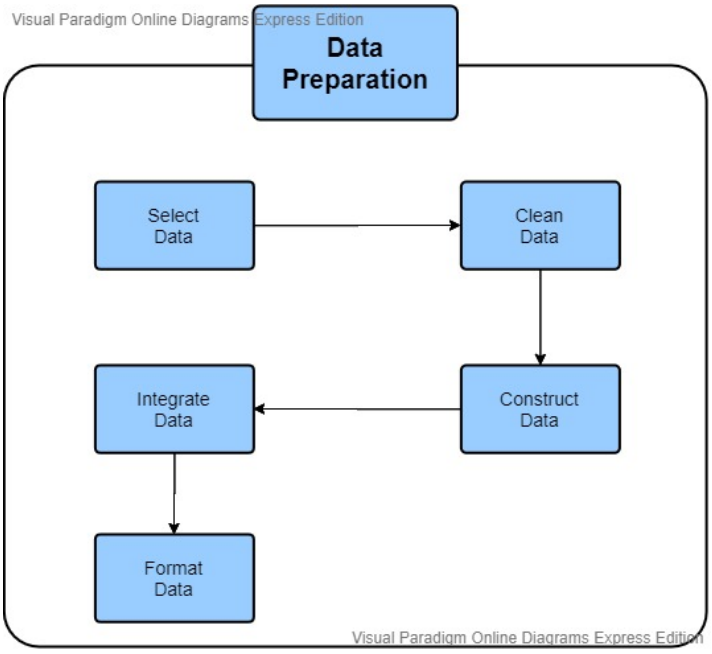


Figure 3.2.3.1: Data Preparation

This phase prepares the data set for the modelling tool and includes data transformation, feature construction, outlier removal, missing data fulfilment and incomplete instances removal. However, this project does not involve data modelling, once the data is cleaned it will be able to plot to dashboard which is the deployment stage.

Select Data

Publication Data Selection

The attributes selected from Scopus publication data to be used for dashboard development are document-count, cited-by-count, citations-count, h-index, co-author-count, surname, given-name, which already chosen during the data collection phase, whereby 7 attributes are chosen from the author retrieval view. The reason these 7 data

was chosen is because it is able to show the individual publication performance of an academic staff, the average performance of a department, faculty and a whole campus.

Teaching Survey Data Selection

The data of teaching survey is restricted to collect only comments given by the students in the individual lecturer evaluation report. The reason the comments section is chosen is to match the objective of this project, perform sentiment analysis on the student feedback in teaching survey.

Clean Data

Cleaning Publication Data

Before the development of dashboard and plotting of the graphs, the data have to be cleaned in order to fit inside Pandas Data Frame.

```
df = pd.read_json (r'C:/Users/User/Documents/UTAR STUFFS/Artificial Intelligence (Python)/Practical/scopus_author_data_merged_unc
df
```

	author-retrieval-response
0	{'@status': 'found', '@_fa': 'true', 'coredata...
1	{'@status': 'found', '@_fa': 'true', 'coredata...
2	{'@status': 'found', '@_fa': 'true', 'coredata...
3	{'@status': 'found', '@_fa': 'true', 'coredata...
4	{'@status': 'found', '@_fa': 'true', 'coredata...
5	{'@status': 'found', '@_fa': 'true', 'coredata...
6	{'@status': 'found', '@_fa': 'true', 'coredata...

Figure 3.2.3.2: Merged Unclean Publication Data of UTAR academic staff

The figure above shows the uncleaned and merged collected data which does not fit inside a data frame due to unwanted metadata and the data structure of the JSON file is incompatible. Generally, the JSON file only allows two data structures which is objects and arrays. Rearrangement of data structure will be done in this phase. Objects are enclosed in braces ({ }), the attribute name and value pair are separated by a comma, the attribute name and value are separated by a colon (:) in a middle. Arrays are enclosed in brackets ([ ]) and the values are separated by a comma, using the information above, the data was restructured and refactored to a version whereby it is readable using pandas data frame. All of the 7 attributes were collected from Scopus Author Retrieval API. The data structure of JSON was cleaned using Visual Studio Code, since python codes unable to make changes and then add a new open bracket or braces.

There are two new attributes being added to the original data, department, faculty and gender, this is because the data will involve every faculty in Universiti Tunku Abdul Rahman, the three new attribute is important to differentiate and compare the performance between faculties. The department attribute is used to classify different academic staff into groups, not just comparing the performance between faculties, it can be made to compare between inter-department.

Figure 3.2.3.3: Cleaned data with added attributes

```
In [12]: import json
         import pandas as pd

         dataFrame = pd.read_csv("C:/Users/User/Anaconda3/Scripts/ScopusApi/elsapy-master/dash-utar-dashboard/data/mainAuthorDataCleaned.c

         dataFrame.tail()

         #dff = df.groupby(['department']).mean()

         #dff
```

Out[12]:

	document-count	cited-by-count	citations-count	h-index	coauthor-count	surname	given-name	faculty	department	gender	affiliation-name
459	2	4	5.0	2	3	Goh	Kai Way	CFSKPR	DSEKPR	male	Universiti Tunku Abdul Rahman
460	3	14	15.0	3	5	Ho	Wai Yew	CFSKPR	DSEKPR	male	Universiti Tunku Abdul Rahman
461	3	69	69.0	1	5	Tan	Lee Siew	CFSKPR	DSEKPR	female	Universiti Tunku Abdul Rahman
462	5	26	26.0	3	5	Kan	Donica Pei Xin	CFSSGL	DSESGL	female	Universiti Tunku Abdul Rahman
463	3	2	3.0	1	11	Tan	Chun Yuan	CFSSGL	DSESGL	male	Universiti Tunku Abdul Rahman

Figure 3.2.3.4: Cleaned data inside data frame ready for dashboard

After all the data was refactored and free from unwanted attributes, it is now ready for data visualization in dashboard development, especially the filter option in dashboard heavily relies on the attribute of the data. Since this is a citation data, the data is not suitable to go through data modelling, it is unable to predict how much a person will publish a research paper.

Cleaning Teaching Survey Data

```
In [8]: #textract is a package that reads any file format and converts to bytes for text analysis
        import textract
        from string import punctuation
        #The text is in bytes form
        text = textract.process("C:/Users/User/Documents/TeachingSurvey_Prob.pdf")

        text= text.decode(encoding="utf-8") #Convert bytes to string format

        text=text.strip()
        text=re.sub("\n","",text) #Remove \n from string
        text=re.sub("\r","",text) #Remove \r from string
        text=re.sub(r'\b\d+\b', '', text) #Remove numbers from string
        text=re.sub(r'\w*\d\w*', '', text).strip() #Remove words that contain numbers
        text=re.sub(r'"(-*\d+(?:\.\d+)?)"', r"\1", text)
        text=re.sub(r'[^\w\s]',' ',text)

        sentences = re.split(r' *[\.\?!][\'"\)\]]* *', text) #Form sentences
        sentences
        studentFeedback= json.dumps(sentences)

        with open("feedbackData.json", "w") as write_file:
            json.dump(sentences, write_file)
```

Figure 3.2.3.5: Cleaned teaching survey data (Data is Confidential)

The teaching survey data was cleaned using python in Jupyter Notebook. The data quality issues of this teaching survey data mentioned in Section 3.2.2 will be solved in this phase. The code was converted to string and applying regular expressions to remove all the unwanted values from the string which includes \n, \r, numbers, numbers between words and all of the symbols inside the string.

Construct Data

Constructing Publication Data

The publication data is already having enough attributes once it is cleaned and added with 3 new columns, would not need any more derived attributes.

Constructing Teaching Survey Data

```
In [5]: df = pd.read_json("C:/Users/User/Documents/UTAR STUFFS/Artificial Intelligence (Python)/Practical/feedbackData.json")


def getSubjectivity(text):
    return TextBlob(text).sentiment.subjectivity

def getPolarity(text):
    return TextBlob(text).sentiment.polarity

df['Subjectivity'] = df['StudentFeedback'].apply(getSubjectivity)
df['Polarity'] = df['StudentFeedback'].apply(getPolarity)

df
```

Out[5]:

	StudentFeedback	Subjectivity	Polarity
0	really like the way you teach sir	0.200000	0.200000
1	I'll try my best to score sir	0.300000	1.000000
2	Thank you sir	0.000000	0.000000
3	666 hope you get your bonus	0.000000	0.000000
4	Tutorial are able understand than lecture	0.625000	0.500000
5	thanks for the fun lectures and the demos on h...	0.233333	0.233333
6	The experiments were greattoo, since I prefer ...	0.750000	0.625000
7	The experiments show me how amazing the formul...	0.900000	0.600000
8	It feels magical is a good lecturer who puts ...	0.800000	0.600000
9	suggest more explanation in lecture	0.500000	0.500000
10	He can improvise his lecture notes a bit by ad...	0.437500	0.281250
11	a good lecturer to teach on this subject, hope...	0.420000	0.140000
12	ways to help students to achieve better	0.500000	0.500000
13	Provides basic learning skills to encourage st...	0.250000	-0.062500

Figure 3.2.3.6: Constructing teaching survey data new column

There are two functions that integrate TextBlob library to perform sentiment analysis on the student feedback collected and cleaned earlier. The getSubjectivity function utilises TextBlob and return the subjectivity value and then it is saved as a new column name known as Subjectivity, same goes to the getPolarity function which is also returning a polarity value and then derived a new attribute known as Polarity.

In [10]:
```
#Create a function to compute the negative, neutral and positive analysis
def getAnalysis(score):
    if score < 0:
        return 'Negative'
    elif score == 0:
        return 'Neutral'
    else:
        return 'Positive'

df['TextAnalysis'] = df['Polarity'].apply(getAnalysis)


df
```

Out[10]:

	StudentFeedback	Subjectivity	Polarity	TextAnalysis
0	really like the way you teach sir	0.200000	0.200000	Positive
1	I'll try my best to score sir	0.300000	1.000000	Positive
2	Thank you sir	0.000000	0.000000	Neutral
3	666 hope you get your bonus	0.000000	0.000000	Neutral
4	Tutorial are able understand than lecture	0.625000	0.500000	Positive
5	thanks for the fun lectures and the demos on h...	0.233333	0.233333	Positive
6	The experiments were greattoo, since I prefer ...	0.750000	0.625000	Positive
7	The experiments show me how amazing the formul...	0.900000	0.600000	Positive
8	It feels magical is a good lecturer who puts ...	0.800000	0.600000	Positive
9	suggest more explanation in lecture	0.500000	0.500000	Positive
10	He can improvise his lecture notes a bit by ad...	0.437500	0.281250	Positive
11	a good lecturer to teach on this subject, hope...	0.420000	0.140000	Positive
12	ways to help students to achieve better	0.500000	0.500000	Positive
13	Provides basic learning skills to encourage st...	0.250000	-0.062500	Negative
14	Take note of student's performance statistically	0.000000	0.000000	Neutral

Figure 3.2.3.7: Constructing teaching survey data another new column

Another function was created to sort the data into a new column whereby when the polarity value is lesser than zero, it will be a negative statement, zero is neutral and more than zero is positive statement.

3.2.4 Modelling

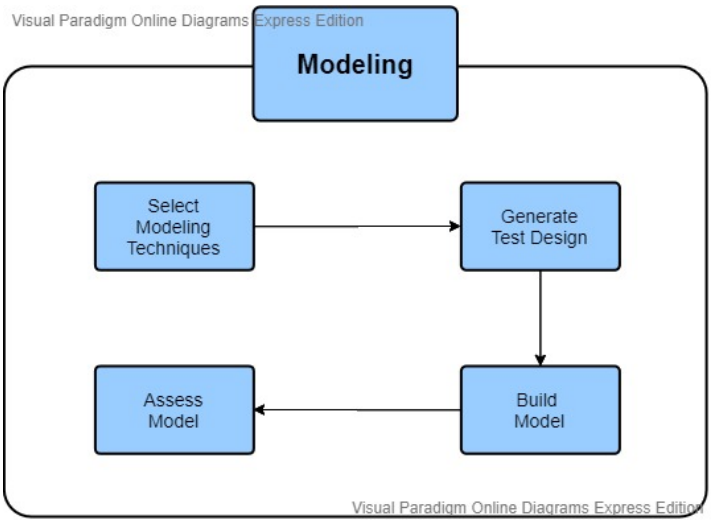


Figure 3.2.4.1: Modelling

The publication data is not suitable to be used in data modelling. The reason is that it is unable to correctly predict how often a person or a university decide to publish a paper, the research might take months to a few years to complete, so there will be no modelling for publication data, it is ready to be used in deployment of dashboard.

The teaching survey data is able be used in data modelling. However, there is only one sample teaching survey provided because this teaching survey data is private and confidential, it is not easy to retrieve the data if it was not provided by the supervisor of this project. Hence, insufficient data to perform machine learning since it involves to split the data into training set and testing set. This phase will be the future work of this project, the collection of data will be the new challenges for the future work and most importantly both publication and teaching survey data can only be viewed by the board of executives of UTAR only, to assist them to make better decision for the university and help to increase the ranking of UTAR.

3.2.5 Evaluation

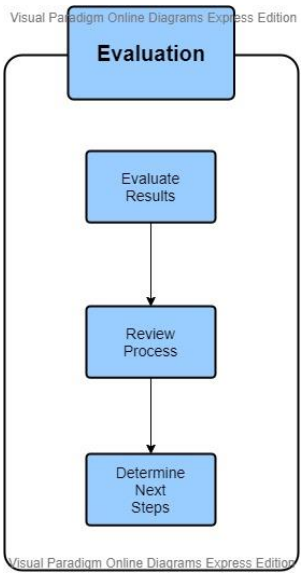


Figure 3.2.5.1: Evaluation

Evaluate Results

This phase is supposed to evaluate the results from the modelling stage, since the project currently still does not involve modelling, the future work in expanding the teaching survey data will certainly involve data modelling. The results of publication data and teaching survey data after data cleaning is eligible for dashboard deployment.

Review Process

The process from business understanding up to modelling are checked regularly, there are no essential elements that has been overlooked.

Determine Next Steps

All of the findings are now ready to be deployed to the dashboard, the data will be uploaded to github in order to avoid retrieving data locally, as it could cause failure in deployment of the dashboard.

3.2.6 Deployment

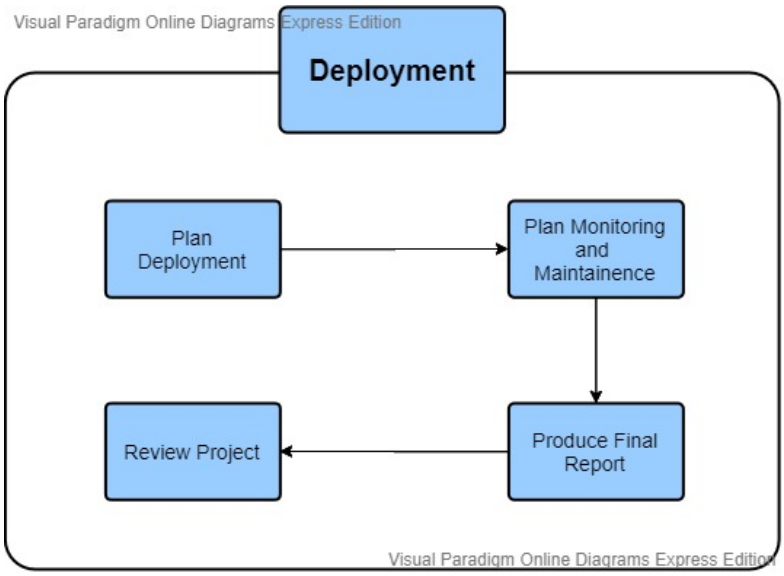


Figure 3.2.6.1: Deployment

Plan Deployment

Plotly Dash recommends the user to deploy all of the works to Heroku, a web application works similar than of github. Here are the steps to deploy the app to Heroku.

1. Sign up for an account on Heroku
2. Give your app a name
3. Installing Heroku CLI in your computer
4. Under your main app.py file, adding this line of code “server= app.server” under the “app=dash.Dash(__name__)”.
5. Open Anaconda Prompt, cd into the project folder
6. Pip install all of the library with specific version which already include inside the requirements.txt file

Figure 3.2.6.2: requirements.txt

7. Create .gitignore file inside the project folder and then add 4 separate lines of venv *.pyc .env .DS_Store

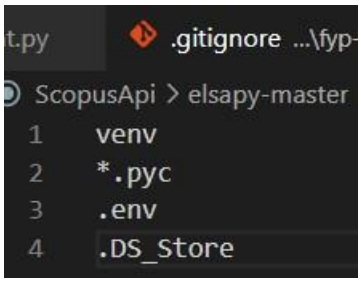


Figure 3.2.6.3: gitignore

8. Create a Profile inside the same folder and add this line "web: gunicorn wsgi:application --preload --workers=3"

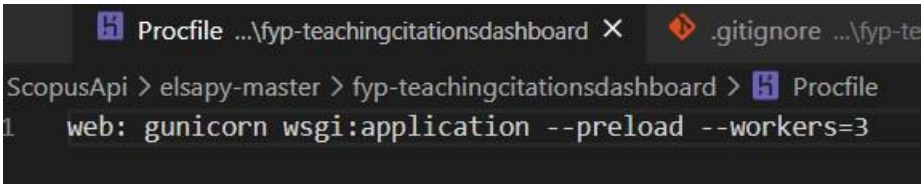


Figure 3.2.6.4: Procfile

9. Create requirements. Go back to anaconda prompt, cd to project folder and type the command ‘pip freeze > requirements.txt.
10. Use the command heroku login to your account
11. Type git init
12. type git add . to add all of your files
13. git commit -am “your commit message”
14. Finally, git push heroku master

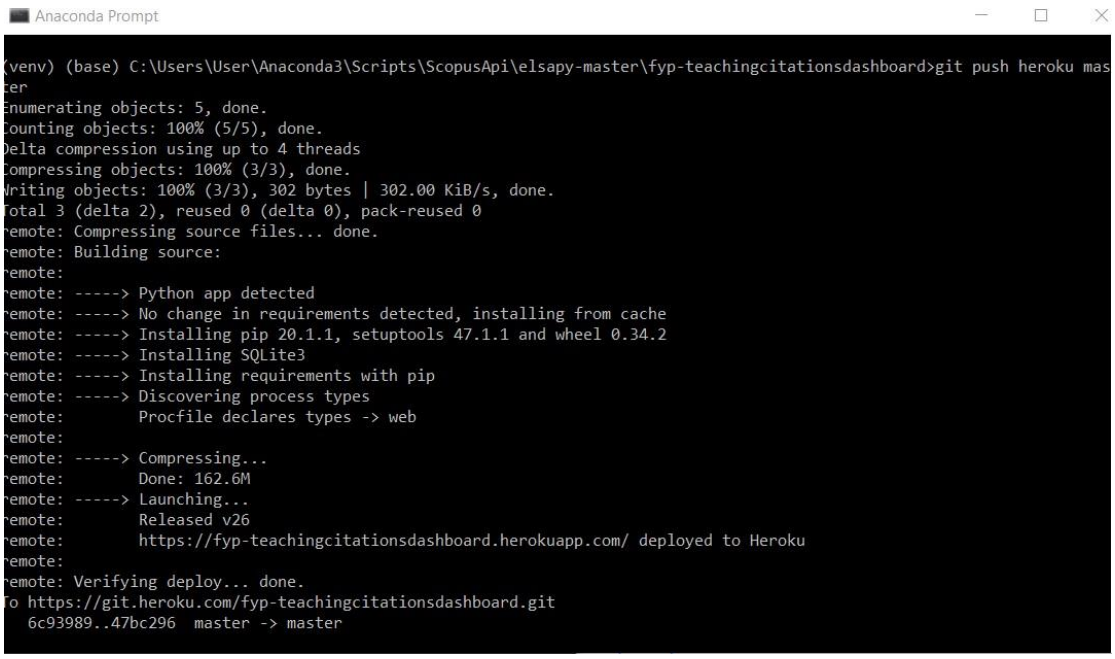


Figure 3.2.6.5: Heroku Deployment

Judging by the looks of it, it seems like the dashboard was successfully deployed to heroku. However, there are dependencies issue. Dash is rejecting Flask related attribute.

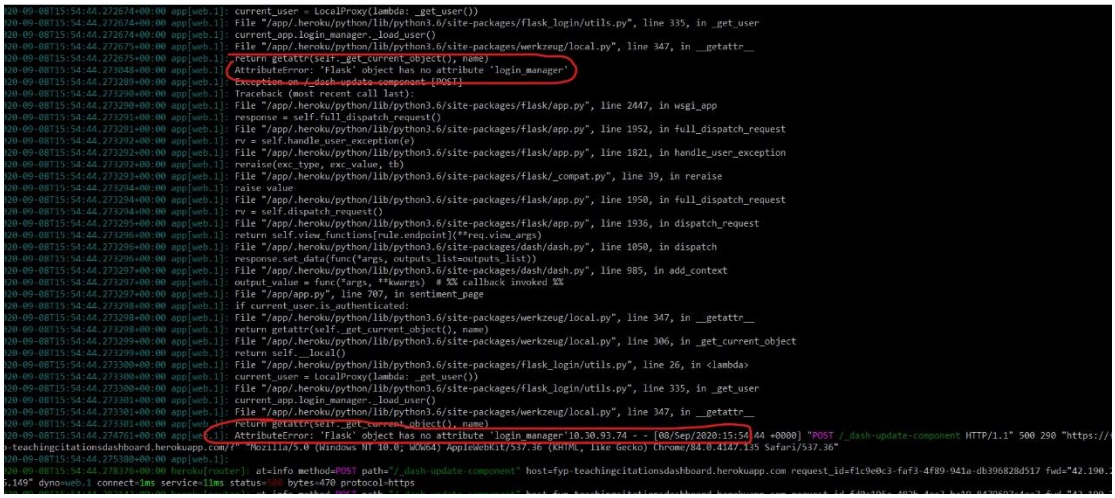


Figure 3.2.6.6: Heroku Log Error

Despite the attribute login_manager exists inside the code, the error does not make sense, causing the dashboard unable to login to the main dashboard, it is only able to see the login part.

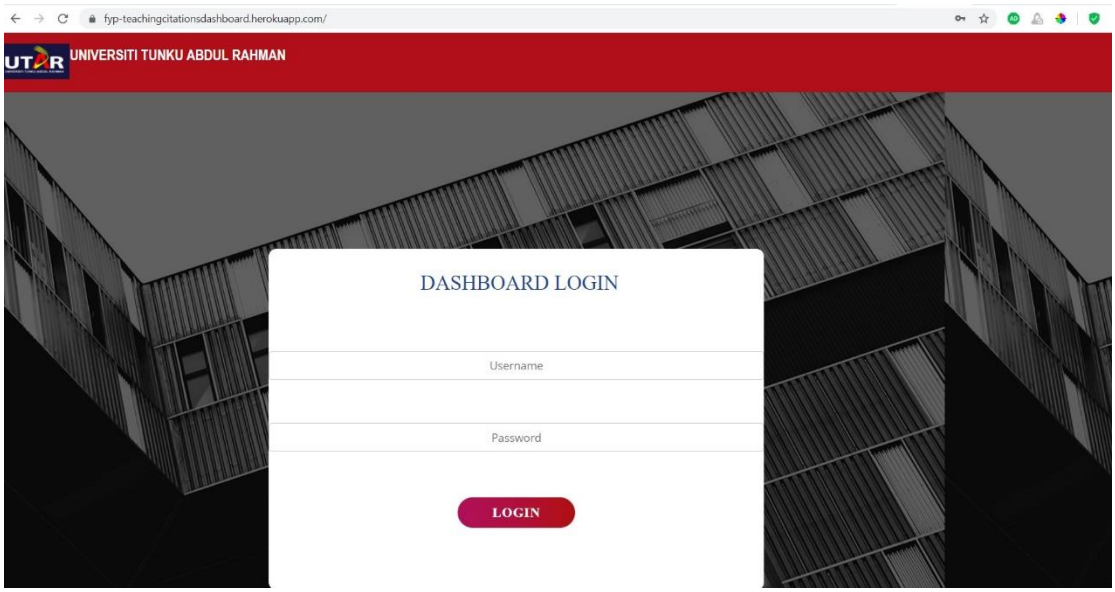


Figure 3.2.6.7: Heroku Dashboard

Hence, ngrok deployment method was used as backup plan, it is able to bring local host connection to the internet, since it is free version, the service is limited to 40 connections per minute and having time limit before the session expires, to increase the connections will have to subscribe a paid hosting plan. Since it is only used for demonstration, free version would be sufficient.

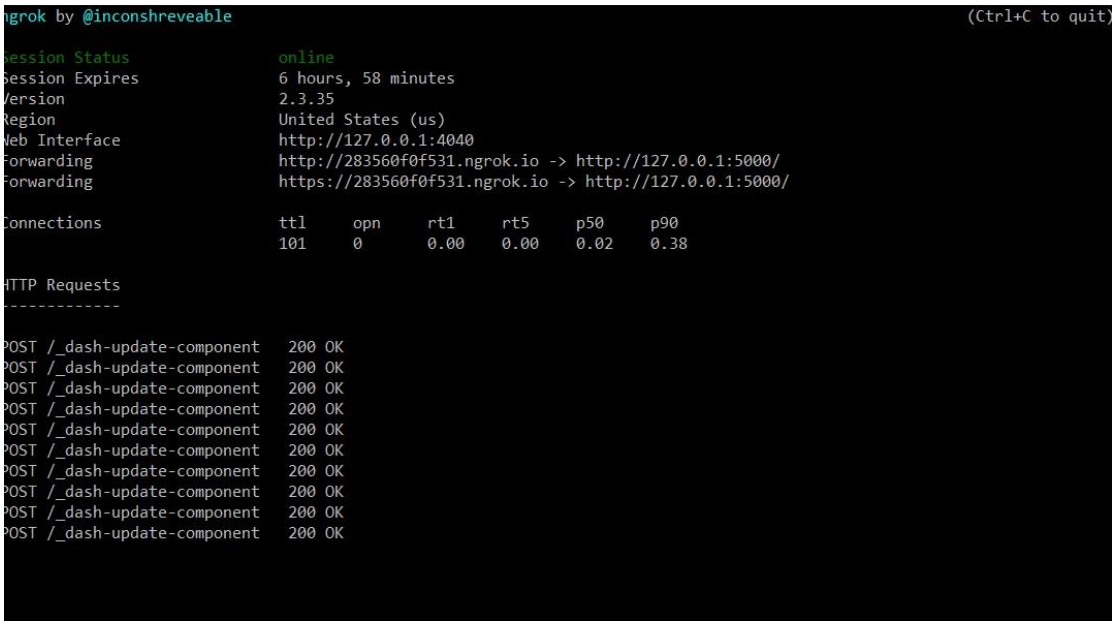


Figure 3.2.6.8: ngrok deploy

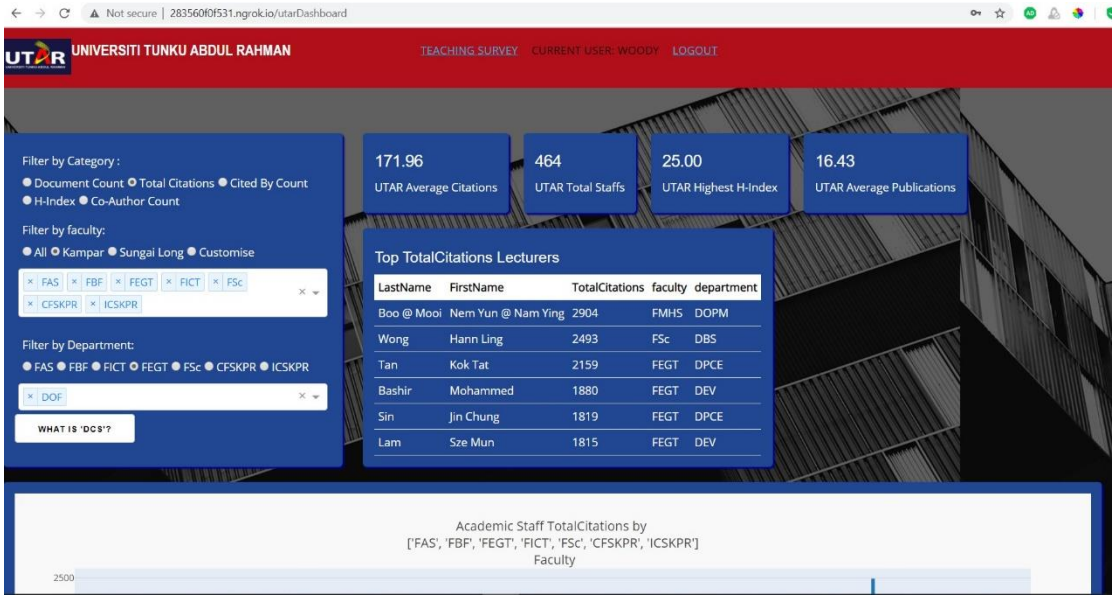


Figure 3.2.6.9: ngrok dashboard

As you can see, the dashboard was successfully logged in using the ngrok hosting.

Plan Monitoring and Maintenance

The dashboard performance heavily relies on the hosting service provider, if there is too much connection or the hosting service is under maintenance, the dashboard might be temporarily unavailable.

Produce Final Report and Review Project

As a result, a dashboard that monitors academic staff in teachings and citations is developed and report is being written. The data collection for this project is relatively tedious, since it requires subscriber IP address, it is best the dashboard is able to implement within UTAR, then the dashboard is able to be done in real-time, since it is always inside UTAR network and available to the management board of the university.

CHAPTER 4 SYSTEM ARCHITECTURE & TESTING

4.1 Use Case Design

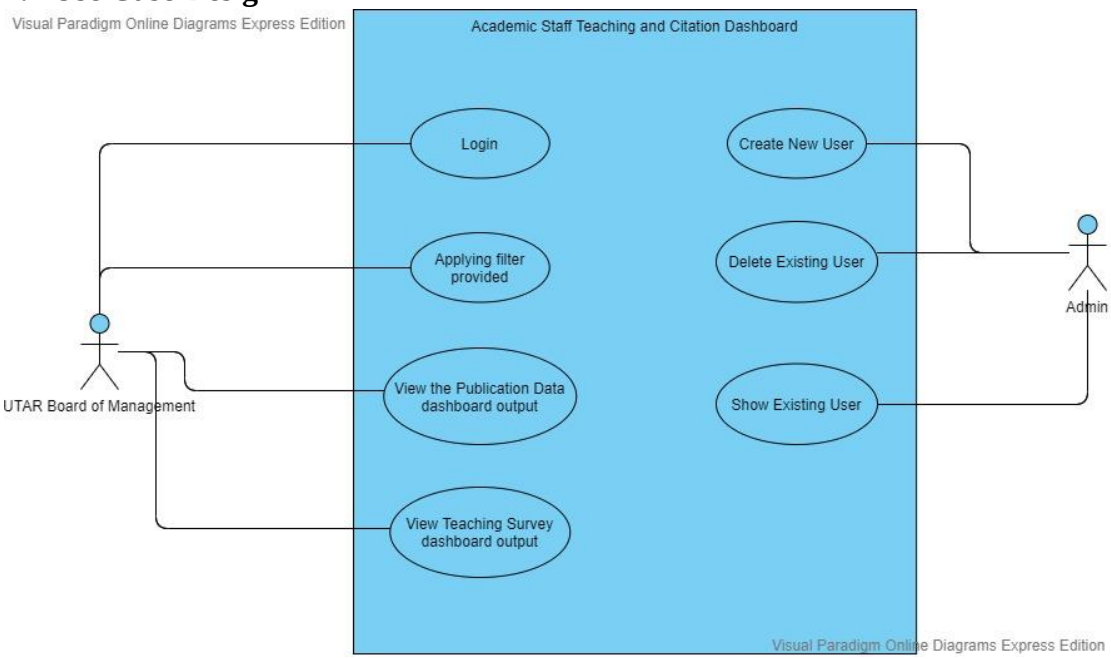


Figure 4.1.1: Academic Staff Teaching and Citation Dashboard Use Case Diagram

The use case diagram above showing how UTAR board of management interact with the dashboard. The management would able to login to the system, applying filter on the dashboard, view publication data and teaching survey. The admin will create new user, delete existing user and show existing user, which controls the amount of people having access to this dashboard.

4.2 Use Case Description

Use Case ID	UC001	Version	1.0
Use Case	Login to the Dashboard		
Aim	To gain access to the dashboard layout		
Actor	UTAR Board of Management		
Trigger	UTAR Board of Management log in to the dashboard using username and password and able to view dashboard layout.		
Precondition	Admin must create a username and password for the board of management		
Main Flow	1. Admin create a new user with username and password 2. The board of management enters the username and password 3. The board of management presses login button or enter 4. The system will return the dashboard layout and display.		
Alternative Flow	2.1 Fails to log into dashboard and return to Login Page		
Author	Brian Ng Yao Sheng		

Table 4.2.1 Use Case Diagram UC001

Use Case ID	UC002	Version	1.0
Use Case	Applying filters provided on the dashboard		
Aim	To view graphs generated with different type of data		
Actor	UTAR Board of Management		
Trigger	UTAR Board of Management uses the radio button or dropdown provided on the dashboard		
Precondition	UTAR Board of Management will have to logged in successfully into the dashboard.		
Main Flow	1. The management clicks on the buttons or dropdown provided. 2. The system filters the data, plots the graph and then return the graph to the user		
Alternative Flow			
Author	Brian Ng Yao Sheng		

Table 4.2.2 Use Case Diagram UC002

Use Case ID	UC003	Version	1.0
Use Case	Viewing teaching survey data dashboard		
Aim	To view teaching survey data dashboard		
Actor	UTAR Board of Management		
Trigger	UTAR Board of Management uses the link on the header to view teaching survey dashboard		
Precondition	UTAR Board of Management will have to logged in successfully into the dashboard.		
Main Flow	1. The management clicks on the link on the header provided. 2. The system returns the teaching survey dashboard to the management. 3. The management will able to interact with the dashboard		
Alternative Flow			
Author	Brian Ng Yao Sheng		

Table 4.2.3 Use Case Diagram UC003

Use Case ID	UC004	Version	1.0
Use Case	Viewing publication data dashboard		
Aim	To view publication data dashboard		
Actor	UTAR Board of Management		
Trigger	UTAR Board of Management will see publication data dashboard after login to the dashboard		
Precondition	UTAR Board of Management will have to logged in successfully into the dashboard.		
Main Flow	1. The management login to the dashboard 2. The system returns the publication dashboard to the management. 3. The management able to interact with the dashboard		
Alternative Flow			
Author	Brian Ng Yao Sheng		

Table 4.2.4 Use Case Diagram UC004

Use Case ID	UC005	Version	1.0
Use Case	Creating new user to access to the dashboard		
Aim	To control the amount of people that allowed to access the dashboard		
Actor	Admin		
Trigger	Admin uses the jupyter notebook file provided known as add_remove_users.ipynb, calling the add_user function, parsing in values		
Precondition	The add_user() function is developed inside user_mgt.py and connection made to the SQLlite database file.		
Main Flow	1. Admin opens the jupyter notebook file 2. Admin uses add_user() function and types in 3 parameters, username, email and password 3. The system adds the new user into the users.db file		
Alternative Flow	2.1 The username is used, return error.		
Author	Brian Ng Yao Sheng		

Table 4.2.5 Use Case Diagram UC005

Use Case ID	UC006	Version	1.0
Use Case	Deleting an existing user from the database		
Aim	To control the amount of people that allowed to access the dashboard		
Actor	Admin		
Trigger	Admin uses the jupyter notebook file provided known as add_remove_users.ipynb, calling the del_user function, parsing in values		
Precondition	The del_user() function is developed inside user_mgt.py and connection made to the SQLlite database file.		
Main Flow	1. Admin opens the jupyter notebook file 2. Admin uses del_user() function and types in the existing username 3. The system deletes the existing user from the users.db file		
Alternative Flow	2.1 The username does not exist, return error.		
Author	Brian Ng Yao Sheng		

Table 4.2.6 Use Case Diagram UC006

Use Case ID	UC007	Version	1.0
Use Case	Show all of the user registered in the database		
Aim	To control the amount of people that allowed to access the dashboard		
Actor	Admin		
Trigger	Admin uses the jupyter notebook file provided known as add_remove_users.ipynb, calling the show_users function, parsing in values		
Precondition	The show_users() function is developed inside user_mgt.py and connection made to the SQLlite database file.		
Main Flow	1. Admin opens the jupyter notebook file 2. Admin uses show_users() function and types in the existing username		
Alternative Flow			
Author	Brian Ng Yao Sheng		

Table 4.2.7 Use Case Diagram UC007

4.3 Sequence Diagram

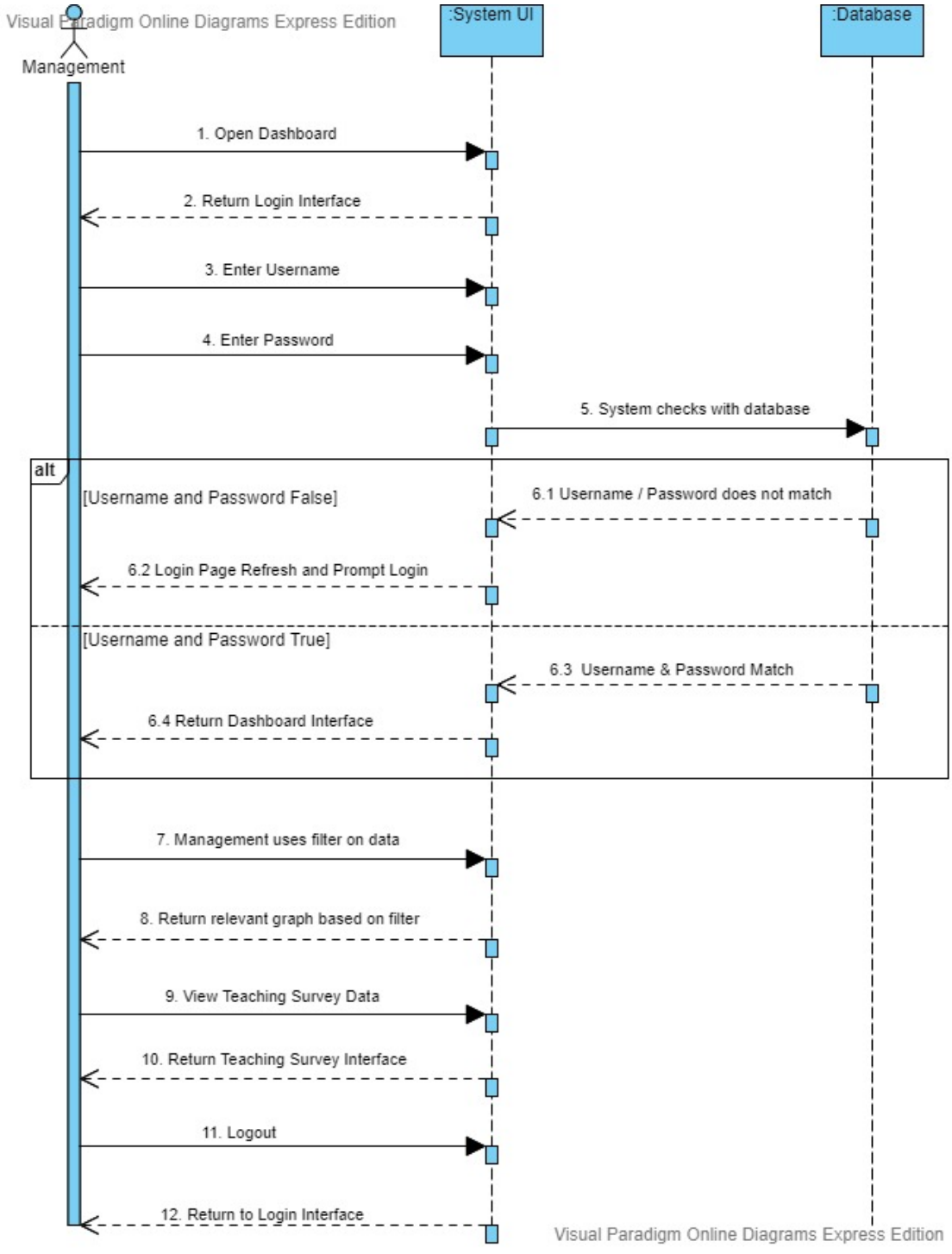


Figure 4.3.1: Teaching and Citations Dashboard Sequence Diagram

4.4 Timeline

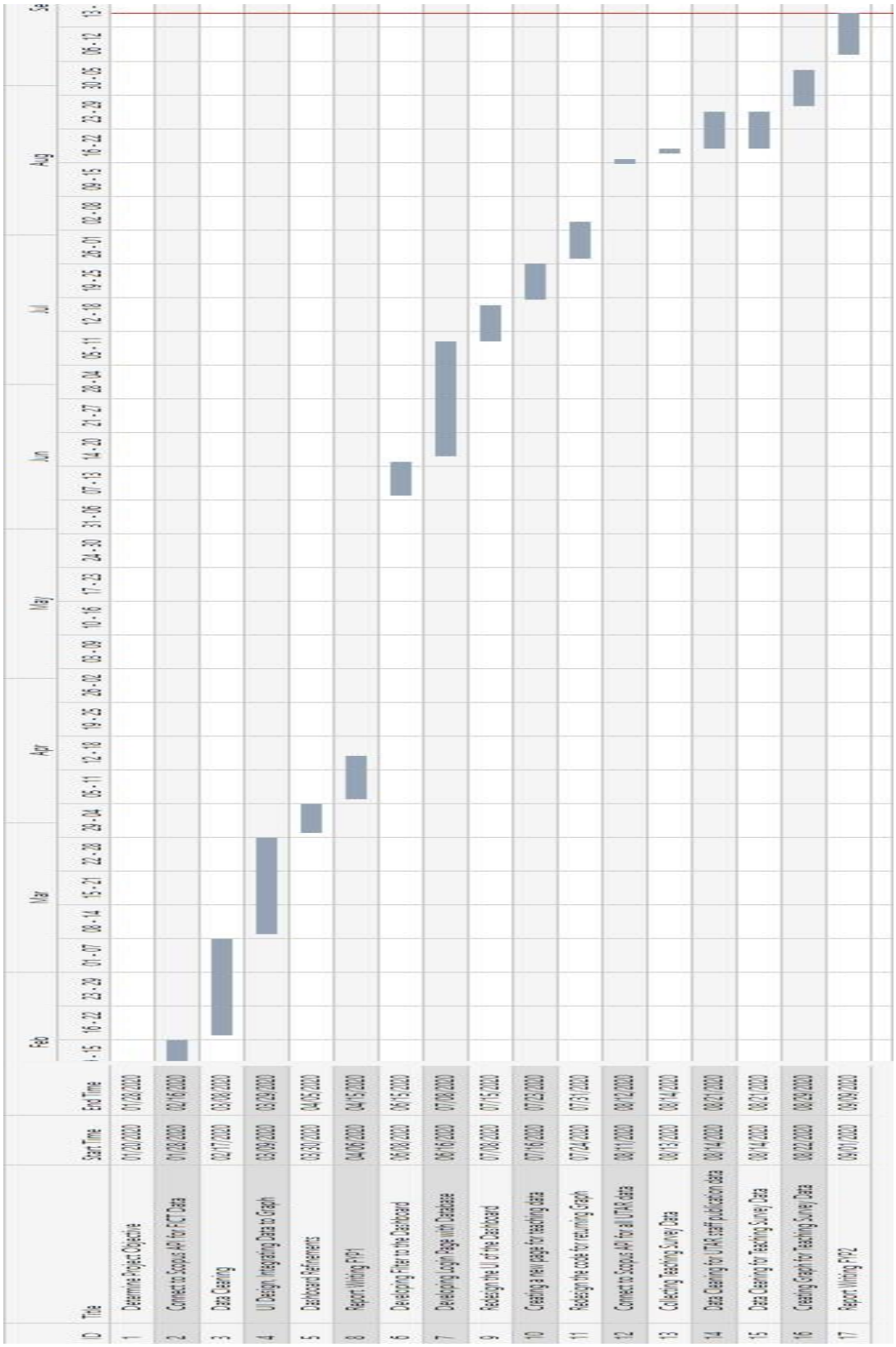


Figure 4.4.1: FYP1 & FYP2 Timeline Gantt Chart

4.5 Use Case Testing

	Input	Expected Output	Actual Output	Result
UC001 Login to the dashboard	User entered valid username	User is able to see the valid username they entered	User is able to see the valid username they entered	Passed
	User entered valid password	User is unable to see the hashed password form	User is unable to see the hashed password form	Passed
	User entered valid username or password and pressed login button	User is able to see the main dashboard layout	User is able to see the main dashboard layout	Passed
	User entered invalid username or password and pressed login button	User failed to login into the dashboard and remains at the Login Page	User failed to login into the dashboard and remains at the Login Page	Passed

Table 4.5.1 Use Case Testing UC001

	Input	Expected Output	Actual Output	Result
UC002 Applying filters provided on the dashboard	User chooses document count button	User is able to see the dashboard return all the values of document count	User is able to see the dashboard return all the values of document count	Passed
	User chooses total citations button	User is able to see the dashboard return all the values of total citations	User is able to see the dashboard return all the values of total citations	Passed
	User chooses cited by count button	User is able to see the dashboard return all the values of cited by count	User is able to see the dashboard return all the values of cited by count	Passed
	User chooses h-index button	User is able to see the dashboard return all the values of h-index	User is able to see the dashboard return all the values of h-index	Passed
	User chooses Co-author count button	User is able to see the dashboard return all the values of co-author count	User is able to see the dashboard return all the values of co-author count	Passed
	User choose the all button in faculty	User is able to see the dashboard return all the values of all faculty, the dropdown menu is updated with all faculty and the radio button is updated under department filter	User is able to see the dashboard return all the values of all faculty, the dropdown menu is updated with all faculty and the radio button is updated under department filter	Passed
	User choose the Kampar button in faculty	User is able to see the dashboard return all the values of kampar faculty, the dropdown menu is updated with kampar faculty and the radio button is updated under department filter	User is able to see the dashboard return all the values of kampar faculty, the dropdown menu is updated with kampar faculty and the radio button is	Passed

			updated under department filter	
	User choose the Sungai Long button in faculty	User is able to see the dashboard return all the values of Sungai Long faculty, the dropdown menu is updated with Sungai Long faculty and the radio button is updated under department filter	User is able to see the dashboard return all the values of Sungai Long faculty, the dropdown menu is updated with Sungai Long faculty and the radio button is updated under department filter	Passed
	User choose the Customise button in faculty	User will be able to see the dropdown menu is empty, user have to choose the faculty, the radio button under department filter will be empty, dashboard is empty	User will be able to see the dropdown menu is empty, user have to choose the faculty, the radio button under department filter will be empty, dashboard is tmpty	Passed
	User uses the dropdown under faculty filter	User will be able to add multiple faculty and the dashboard will updated accordingly	User will be able to add multiple faculty and the dashboard will updated accordingly	Passed
	User uses the dropdown under department filter	User will be able to add multiple department and the dashboard will updated accordingly	User will be able to add multiple department and the dashboard will updated accordingly	Passed

Table 4.5.2 Use Case Testing UC002

	Input	Expected Output	Actual Output	Result
UC003 Viewing the teaching survey dashboard	User clicked the teaching survey link at the header section	User is redirected to teaching survey dashboard	User is redirected to teaching survey dashboard	Passed
	User hovers over a data	User is able to see the data values of the particular attribute	User is able to see the data values of the particular attribute	Passed
	User clicks the zoom in or zoom out function at the graph	User is able to see the enlarged or minimized graph	User is able to see the enlarged or minimized graph	Passed
	User clicks the camera icon at the graph	User is able to download the plot as png file	User is able to download the plot as png file	Passed

Table 4.5.3 Use Case Testing UC003

	Input	Expected Output	Actual Output	Result
UC004 Viewing the publication data dashboard	User hovers over a data	User is able to see the data values of the particular attribute	User is able to see the data values of the particular attribute	Passed
	User clicks the zoom in or zoom out function at the graph	User is able to see the enlarged or minimized graph	User is able to see the enlarged or minimized graph	Passed
	User clicks the camera icon at the graph	User is able to download the plot as png file	User is able to download the plot as png file	Passed

Table 4.5.4 Use Case Testing UC004

	Input	Expected Output	Actual Output	Result
UC005 Creating new user to access the dashboard	Admin enters username, email and password to add_user() function	User is able to login to the dashboard	User is able to login to the dashboard	Passed

Table 4.5.5 Use Case Testing UC005

	Input	Expected Output	Actual Output	Result
UC006 Delete an existing user from the database	Admin enters username, del_user() function	The deleted user is unable to login to the dashboard	The deleted user is unable to login to the dashboard	Passed

Table 4.5.6 Use Case Testing UC006

	Input	Expected Output	Actual Output	Result
UC007 Show all the user registered in the database	Admin uses show_users() function	All of the existing users that can access the dashboard is shown	All of the existing users that can access the dashboard is shown	Passed

Table 4.5.7 Use Case Testing UC007

4.6 Tools to Use

Visual Studio Code

A powerful source-code editor that supports a library of languages including HTML, Python, CSS, Javascript, JSON, etc. It a very versatile code editor and compatible with Anaconda Prompt which is the main environment used for this project. Besides, the code snippets and syntax highlighting ease up debugging process and simplifies code refactoring. In this project, the coding for collection of data and development of dashboard was done using Visual Studio Code, the code was executed using the terminal inside visual studio code.

Plotly Dash

Dash is an open source Python and R framework for creating responsive, web-based analytic application especially data science projects. All of the applications are able to run web servers locally using Flask and communicating JSON packets over HTTP requests. All of the user interface in this project is developed using dash core components including graphs, it is very similar to HTML but with a dcc or html in front of the coding. Every aesthetic element is of the app is fully customizable with the help of CSS, thus it is very suitable for this project.

Anaconda

Anaconda is a free and open-source distribution of the Python and R programming languages used for scientific computing including data science, it serves as a package management which includes Anaconda Navigator serves as the GUI of the system. Anaconda prompt, jupyter notebook and visual studio code provided in anaconda are the main tools used for this project, the project environment was creating using anaconda prompt with the help of PyPI. All of the dependencies required was installed using Anaconda Prompt within the virtual environment created.

Jupyter Notebook

An open-source web application that supports multiple programming languages that allows users to create interactive web application that contain visualizations, equations code. For publication and citations data, most of the trial and error coding for applying filter to the dashboard was performed in Jupyter Notebook before it was applied to the

main code. The data preparation and cleaning of teaching survey data and sentiment analysis was done in Jupyter Notebook, then it was saved as JSON file to be plotted into graph in dashboard.

SQLAlchemy

A Python SQL toolkit and Objective Relational Mapper that allows the web or application developers the flexibility of SQL, it is designed for high performance and efficient database access, the SQL commands was adapted into simple Python based language. The login user data database was created using SQLAlchemy.

CHAPTER 5: SYSTEM IMPLEMENTATION

5.1 Dashboard Overview

The user interface of this dashboard is divided into 3 pages, login page, publication dashboard page and teaching survey dashboard page. The left side of the dashboard are equipped with different kinds of filter such as filter by category, faculty, department. There are 4 boxes showing the overall performance of all UTAR academic staff, the first column displayed the average citation counts, followed by total number of staffs having publication, highest h-index acquired and average publications published by UTAR. There is another data table below showing all of the publication details of the academic staff, users are allowed to apply filter to see different staff appearing at the data table.

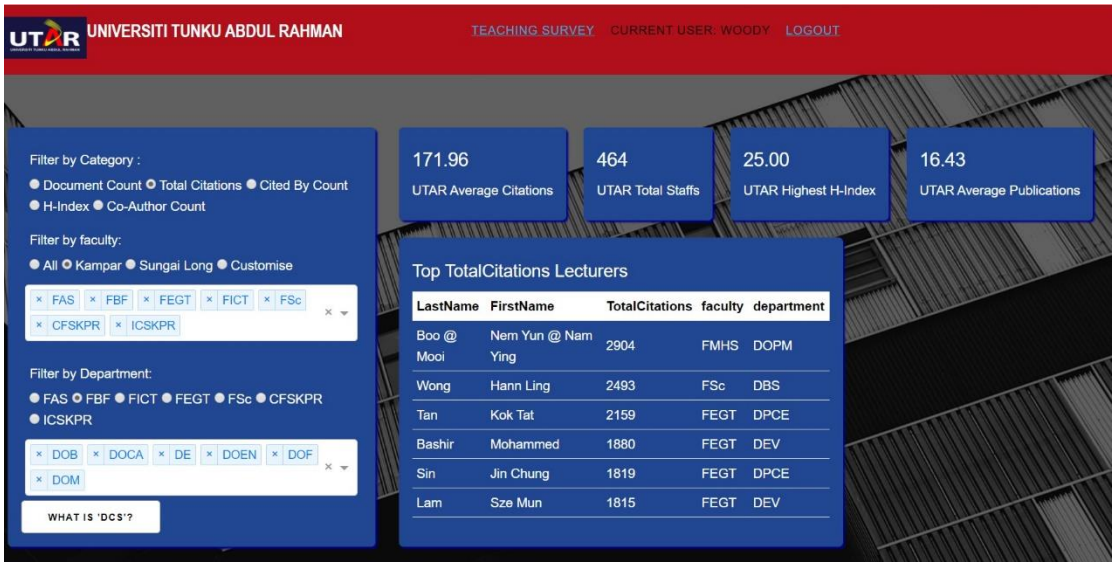


Figure 5.1.1: User Interface of Dashboard

5.2 Publication Dashboard Features

5.2.1 Login Page

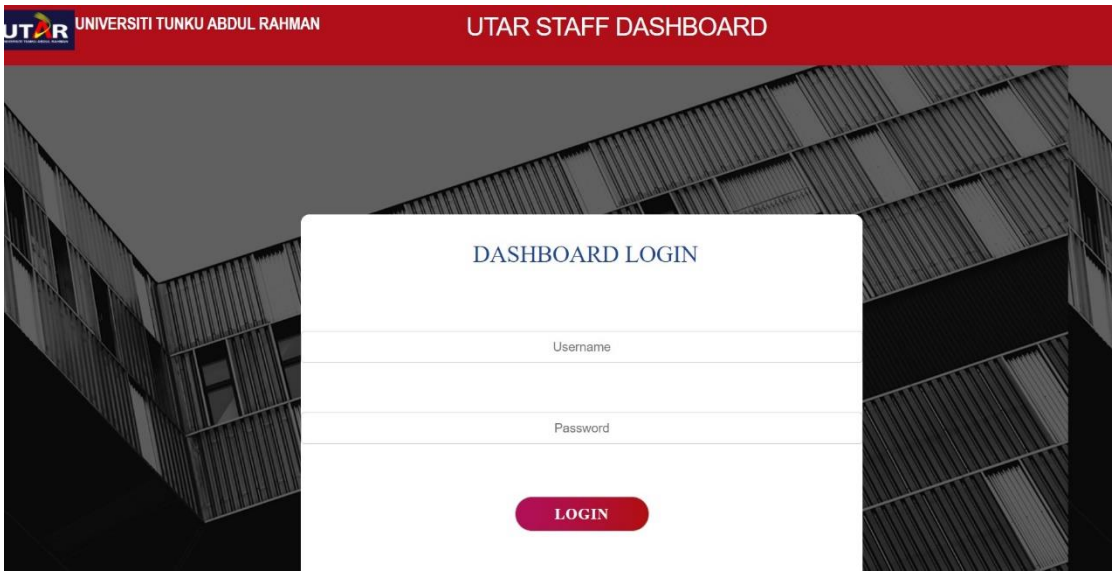


Figure 5.2.1.1: Login Page of Dashboard

The login page having interaction as well, if the username or password column is selected, the login button is highted once the mouse is hover over the login button. The dashboard is intended to assist the UTAR board of executives in making better decisions by monitoring the teaching performance and publication performance of academic staff, which in turn increasing the ranking of UTAR in Times Higher Education. Hence, the dashboard is not made to be publicly accessible. SQL Injection attack was performed on the login and it is proved does not work, so the security of the data is guaranteed.

5.2.2 Bar Charts

There are 4 bar charts available, 2 involve, co-author, total citations and h-index. The reason behind using bar charts to visualise the data is because bar charts are the best to compare things between groups, especially performance between the academic staff. All of the charts are equipped with user interactivity, despite the bar chart has been allocated with a lot of space, the number of columns is still too much to fit in the bar chart, users will have to use the zoom function and then using the pan function to move through the bar chart to see how the individual academic staff fare in each of the category of the data. Once the user is done with traversing the data, the auto-scale button

automatically resizes the chart into default appearance. Besides, if the user hovers over the data, it will display the exact value and the name of the academic staff which is very convenient and centralised place to visualise staff data. Apart from that, user have the option to download the plot as PNG.

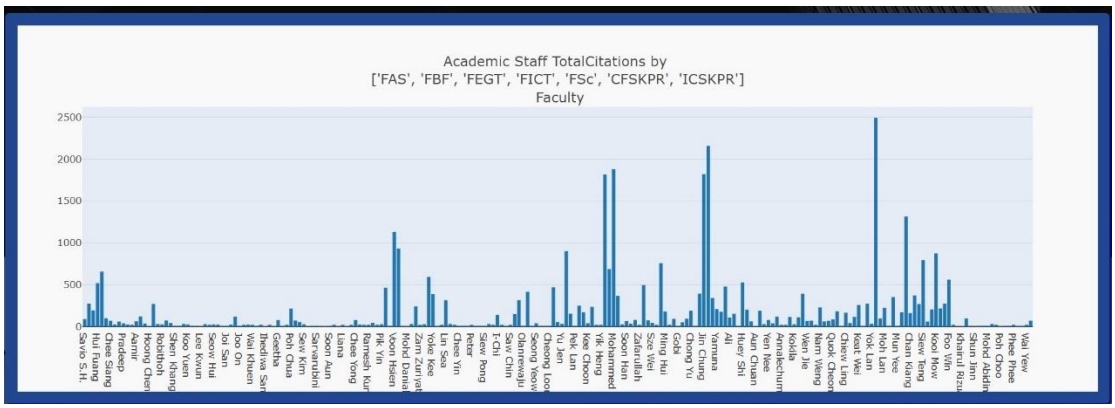


Figure 5.2.2.1: Academic Staff with category by Faculty Bar Chart

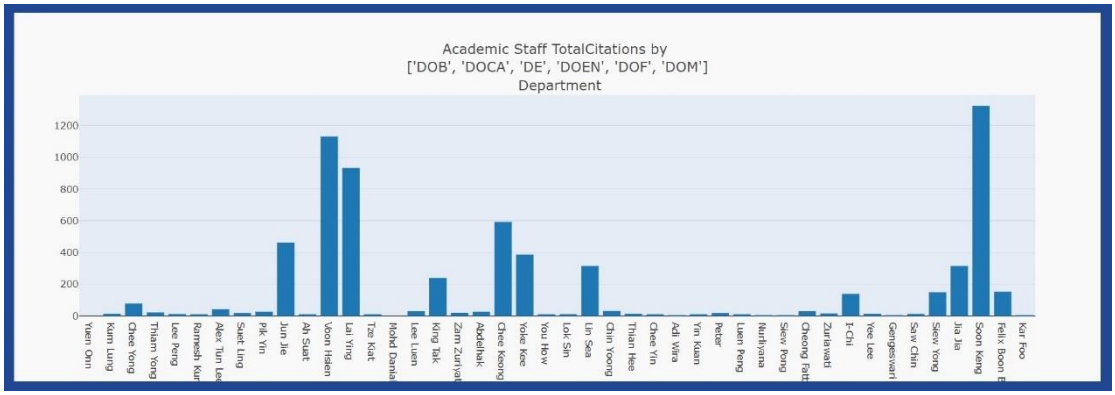


Figure 5.2.2.2: Academic Staff with category by Department Bar Chart

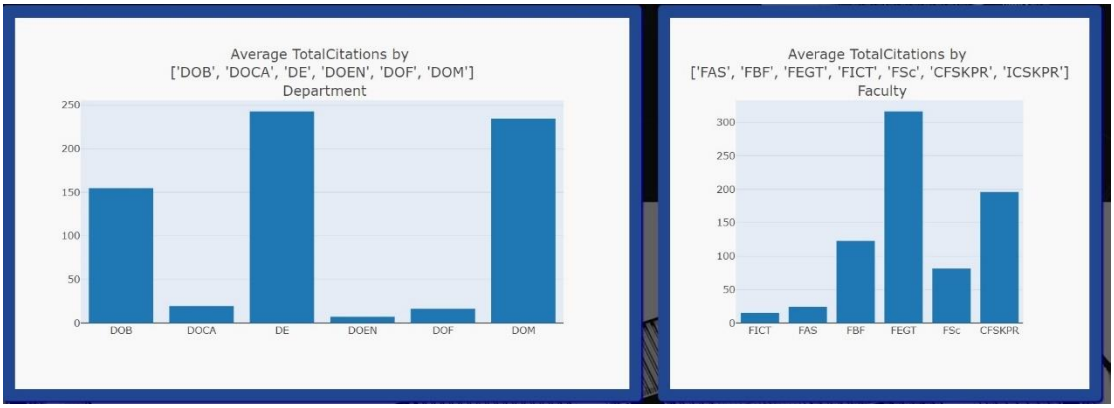


Figure 5.2.2.3: Average category by Department and Faculty Bar Chart

5.2.3 Pie Chart

A simple pie chart that is interactive showing the number of male lecturers and female lecturers according to the faculty filter applied.

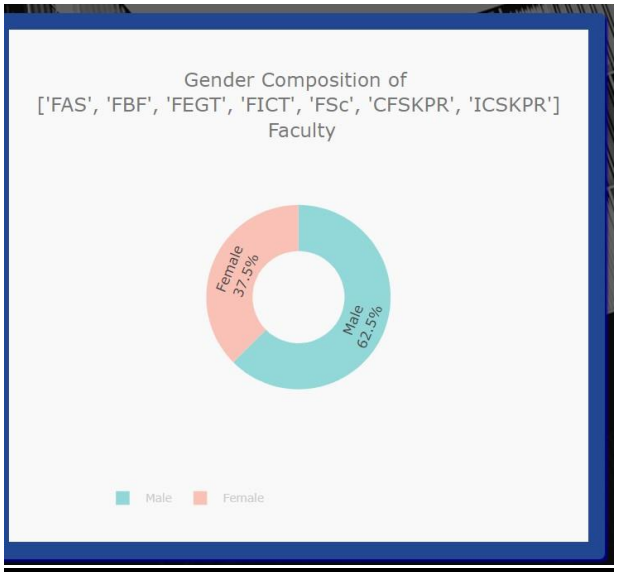


Figure 5.2.3.1: Gender by faculty Pie Chart

5.3 Teaching Survey Dashboard

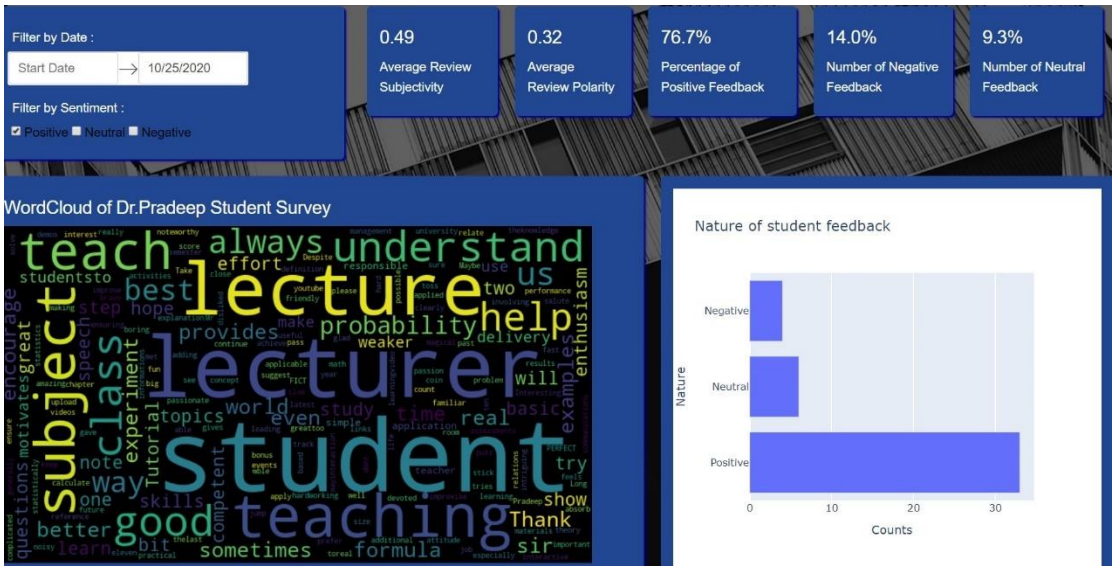


Figure 5.3.1: Top part of teaching survey dashboard

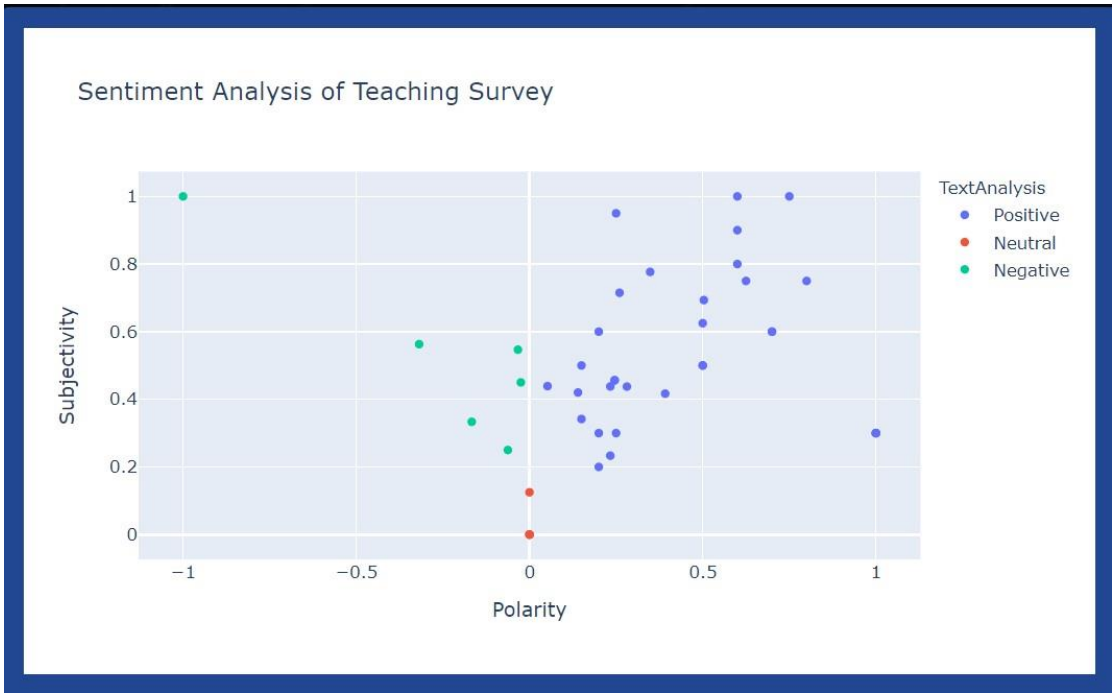


Figure 5.3.2: Bottom part of teaching survey dashboard

There are 5 small boxes showing the average Subjectivity, average Polarity, percentage of positive, negative and neutral feedbacks of that academic staff

Word Clouds is one of the low-cost alternatives to analysing text from student feedback or customer feedback. Basically, word clouds generators work by splitting the text down into component words and counting how frequent those words appear in the body of the text. Next, the font size of the words is based on the frequency the word appears inside the text, the more frequent the word appears, the larger the word is shown. However, there is no word cloud component in Plotly Dash, so in order to show word cloud in Plotly Dash, word cloud library is needed to import into the code, then convert the word cloud result into image format, then display the word cloud using image component. This way of analysing the teaching survey data is much better than the conventional way whereby an individual evaluation report was generated with the comments listed as rows of text without any sentiment analysis performed, which are often ignored.

The bar chart simply shows the count of positive, neutral and negative student feedback, the bar chart does have interactivity, users can hover over the bar chart to view more details of the data.

The scatter plot shows the user clearly about the frequency of positive, neutral and negative comments acquired by an academic staff, the zero value at the middle acts as the trend line, left is negative, right is for positive.

5.4 Implementation Issues and Challenges

5.4.1 FYP1 Challenges

One of the challenges for this project is to collect publication and citation data from Scopus API. Despite the Scopus API Documentation have stated that the API can only accessed from IP address that have deemed as subscribers only, not every UTAR network is truly accessible to the API. Initially, the first attempt of data retrieval was done using UTAR public Wi-Fi, obviously the first attempt failed as the response from the API was as figure below, the user was not authorised to retrieve the data.

```
(base) C:\Users\User\Anaconda3\Scripts\ScopusApi\elsapy-master>python scopusSearch.py
{'service-error': {'status': {'statusCode': 'AUTHORIZATION_ERROR', 'statusText': 'The requestor is not authorized to access the requested view or fields of the resource'}}}
```

Figure 5.4.1: Unauthorised access to API

The second attempt was done inside UTAR FYP lab, using the LAN cable provided by the lab officer with the specific proxy IP and port. Unfortunately, it is still showing the same error which is unauthorised access to the Scopus API, this is probably something to do with the network organization of UTAR causing trouble to recognise the eligible UTAR network, with the help of the supervisor of this project, the third attempt on data collection was done inside Dr.Pradeep's room using the LAN cable provided and successfully retrieved the data after weeks of trying.

The reason it took about 3 weeks to perform data collection is because the data can only be retrieved if and only if Dr.Pradeep is inside the office and there are 3 different modules that is able to access to the Scopus API. The first module used was known as 'elsapy', the module only returns one data per attribute, not able to support multiple attribute request, which is very time consuming to collect every lecturer data in FICT. The next module is 'pybliometrics' which also another module that is not able to return several attributes at a time. Finally, the request library from Python works and successfully get the data required.

There were challenges to develop the callback function to increase the interactivity of the whole data visualization using Plotly Dash, as the filter function fully relies on how the callback function works, more trial and error development will be done to overcome this issue.

5.4.2 FYP2 Challenges

The fact that data collection has to be done inside the project supervisor's room is still a major problem in FYP2 due to COVID-19. Initially, before the announcement UTAR does not allow students to come back to campus until official announcement was made that students are allowed to enter the campus as early as week 4. After requests for permission to enter UTAR was approved, all of the publication data which includes every faculty of UTAR and teaching survey data was finally collected during Week 9, since the data was collected late, it has become another challenge to clean and reformat the data for graph plotting and the development of dashboard in short time, as the submission of FYP report is Week 13. Against all odds, the dashboard was developed with all the objectives achieved.

CHAPTER 6: CONCLUSION

6.1 Project Review and Contributions

As mentioned earlier, publication, citation, teaching performance and research influence are the 4 main pillars to bring a university to a higher ranking as it accounts for 90% of the overall evaluation criteria in the Times Higher Education Ranking System. This has brought up the main motivation of this project, whereby Malaysian universities lacks of a university owned dashboard that monitors on publications and citations published by their academic staff or post-graduate students. Instead, third party sites like Google Scholar or Scopus were relied on to do the job. Besides, Scopus can only check the publications of a particular author one at a time, a platform to view and compare the performance of all of the university staff is required. Apart from that, the UTAR annual research and development report involves merely graphs generated in image form, does not have any interactivity and the comments in lecturer evaluation report does not go through further analysis, just showing rows of comments. If a university is able to grasp publication data, it would bring significant advantage. In order to solve the problem above, a dashboard is the best solution as it is a centralised platform that collects all publication data and teaching survey data into one place without the need of third-party software or website.

In brief, the dashboard is developed using data retrieved from the Scopus API, and stored as JSON file to be further cleaned and removing unwanted attributes along with teaching survey data in PDF form, converted to string to perform data cleaning and then perform sentiment analysis. Then, the user interface was created using Plotly Dash to visualise the data collected.

6.2 Novelties

The novelty in this project is the development of a publication and teaching monitoring dashboard for UTAR which is not widely adopted by Malaysian Universities, not even Universiti Malaya which is also relying on annual report as well. Thus, it is about time for UTAR to utilize and reap the benefits of implementing dashboard to monitor all of the publication data and teaching survey.

6.3 Future Work

There are a lot of potential in this project, the scope of teaching survey data can be expanded to every academic staff of UTAR, but the dashboard is only for the dean of respective faculty and board of management of UTAR since the data is considered to be private and confidential. In addition, UTAR did not adopt a dashboard to monitor in their publication monitoring and teaching survey, instead a static graph was shown in the annual research & development report without any interactivity and the comments given to a particular lecturer in teaching evaluation is often ignored and not analyzed. Hence, this dashboard is the perfect solution to solve all the issues currently UTAR is having.

Besides, the scope of the dashboard data can even expand broader, developing a dashboard that is accessible to the public, as this was never done by Malaysian Universities. For example, UTAR can start from developing a dashboard in admission data, student demographics, degrees awarded to students etc.

Bibliography

Aanderud, T. (2018). *Working on Your Dashboard Layout?*. [online] Medium. Available at: https://towardsdatascience.com/working-on-your-dashboard-layout-9b7c38d7b61e

[Accessed 4 August. 2019].

Altrabsheh, N., Cocea, M. and Fallahkhair, S., 2014. Sentiment Analysis: Towards a Tool for Analysing Real-Time Students Feedback. *2014 IEEE 26th International Conference on Tools with Artificial Intelligence*, [online] Available at: https://sci-hub.tw/10.1109/ictai.2014.70

[Accessed 27 August 2020].

Balahadia, F., Fernando, M. and Juanatas, I., 2016. Teacher's performance evaluation tool using opinion mining with sentiment analysis. *2016 IEEE Region 10 Symposium (TENSYMP)*, [online] pp.95-97.

Available at: <https://sci-hub.tw/10.1109/tenconspring.2016.7519384> [Accessed 27 August 2020].

Berinato, S., 2016. *Harvard Business Review*. [Online] Available at: https://hbr.org/2016/06/visualizations-that-really-work [Accessed 30 July 2019].

Bouwma-Gearhart, J. & Collins, J., 2015. *What We Know About Data-Driven Decision Making In Higher Education: Informing Educational Policy and Practice.* Florence, IISES.

Durcevic, S. (2018). *Types of Dashboards: Strategic, Operational & Analytical Examples*. [online] BI Blog | Data Visualization & Analytics Blog | datapine. Available at: https://www.datapine.com/blog/strategic-operational-analytical-tactical-dashboards/

[Accessed 30 Jul. 2019].

Ebrahim, N., 2014. Citations and its Impact to University Ranking. [online]

Available at:<https://www.researchgate.net/publication/285057833_Citations_and_its_Impact_to_University_Ranking>

[Accessed 15 March 2020].

Few, S. (2006). *Information Dashboard Design: The Effective Visual Communication of Data*. 1st ed. [ebook] Gravenstein Highway North Sebastopol: O'Reilly Media, Inc., pp.14-15,26,30-32. Available at: http://93.174.95.29/main/674000/0dcb4dc6fbaf7eb862efccc0404759eb/Stephen%20Few%20-%20Information%20Dashboard%20Design_%20The%20Effective%20Visual%20Communication%20of%20Data-O%27Reilly%20Media%20%282006%29.pdf

[Accessed 22 Jul. 2019].

Friendly, M., 2008. The Golden Age of Statistical Graphics. *Statistical Science*, 23(4), pp. 502-535.

Hipp, J. & Wirth, R., 2000. *CRISP-DM: Towards a Standard Process Model for Data MMining*, Tübingen: -.

Kabakchieva, D., 2015. Business Intelligence Systems for Analyzing University Students Data. *CYBERNETICS AND INFORMATION TECHNOLOGIES*, [online] 15(1), pp.104-115.

Available at: https://sci-hub.tw/10.1515/cait-2015-0009

[Accessed 12 July 2020].

Lewis, J. and Hensly, T. (1998). *The May 4 Shootings at Kent State University: The Search For Historical Accuracy*. [online] Internet Archive Wayback Machine. Available at: https://web.archive.org/web/20080509131525/http://dept.kent.edu/sociology/lewis/LEWIHEN.htm#KENT

[Accessed 28 Jul. 2019].

Littlewood, C., 2018. *Harvard Business Review.* [Online] Available at: https://hbr.org/2018/10/prioritize-which-data-skills-your-company-needs-with-this-2x2-matrix
[Accessed 12 August 2019].

Matthews, K., 2018. *6 Ways big data is improving higher education.* [Online] Available at: https://www.technobuffalo.com/6-ways-big-data-is-improving-higher-education
[Accessed 30 July 2019].

Moreira, J., Carvalho, A. C. P. L. F. & Horváth, T., 2019. *A General Introduction to Data Analytics.* 1st ed. Hoboken(New Jersey): John Wiley & Sons, Inc.

Morris, T., 2018. *6 Competitive Advantages of Data-driven Organizations.* [Online] Available at: https://www.microstrategy.com/us/resources/blog/bi-trends/6-competitive-advantages-of-data-driven-organizati
[Accessed 1 August 2019].

Pernice, K. (2017). *F-Shaped Pattern of Reading on the Web: Misunderstood, But Still Relevant (Even on Mobile)*. [online] Nielsen Norman Group. Available at: https://www.nngroup.com/articles/f-shaped-pattern-reading-web-content/

[Accessed 29 Jul. 2019].

Pryke, B., 2018. *Jupyter Notebook for Beginners: A Tutorial.* [Online] Available at: https://www.dataquest.io/blog/jupyter-notebook-tutorial/
[Accessed 9 August 2019].

Rani, S. and Kumar, P., 2017. A Sentiment Analysis System to Improve Teaching and Learning. *Computer*, [online] 50(5), pp.36-43. Available at: <https://sci-hub.tw/10.1109/MC.2017.133> [Accessed 27 August 2020].

The University of Texas System. (2017). *Fall Enrollment | UT System Dashboard*. [online] Available at: https://data.utsystem.edu/data-index/fall-enrollment

[Accessed 29 Aug. 2019].

BIBLIOGRAPHY

Times Higher Education (THE). (n.d.). *Kent State University*. [online]

Available at: https://www.timeshighereducation.com/world-university-rankings/kent-state-university

[Accessed 28 Jul. 2019].

University of Wyoming Dashboard. (2016). [pdf] Laramie, Wyoming, pp.1-7. Available at: http://www.uwyo.edu/administration/_files/docs/DashBoard/uw_dashboard_April_2016%205.2.2016.pdf

[Accessed 29 Jul. 2019].

U.S. News & World Report. (2018). *Kent State University*. [online]

Available at: https://www.usnews.com/best-colleges/kent-state-university-3051

[Accessed 28 Jul. 2019].

U.S. News & World Report. (n.d.). *University of Wyoming*. [online]

Available at: https://www.usnews.com/best-colleges/university-of-wyoming-3932

[Accessed 29 Jul. 2019].

U.S. News & World Report. n.d. *University Of California--Los Angeles*. [online] Available at:

https://www.usnews.com/education/best-global-universities/university-of-california-los-angeles-110662

[Accessed 26 July 2020].

UTAR Research & Development Annual Report 2018. n.d. *Publication Report*. [online]

Available at: <http://research.utar.edu.my/rd/rnd2018/publication-report.html> [Accessed 27 July 2020].

BIBLIOGRAPHY

UTAR Research and Development Annual Report 2019. n.d. *Publication Report*. [online]

Available at: <http://research.utar.edu.my/rd/rnd2019/publication-report.html> [Accessed 25 July 2020].

UTAR Research & Development Annual Report 2018. n.d. *External Funding Report*. [online]

Available at: <http://research.utar.edu.my/rd/rnd2018/external-funding.html> [Accessed 11 August 2020].

UTAR Institute of Postgraduate Studies and Research. n.d. [online] Available at: <https://ipsr.utar.edu.my/Home.php>

[Accessed 16 August 2020].

Vasconcellos, P., 2018. *Top 5 Python IDEs For Data Science.* [Online] Available at: https://www.datacamp.com/community/tutorials/data-science-python-ide
[Accessed 6 August 2019].

Wexler, S., Shaffer, J. and Cotgreave, A. (2017). *The Big Book of Dashboards, visualizing your data using real-world business scenarios*. 1st ed. [ebook] Hoboken, New Jersey: John Wiley & Sons, Inc., p.19. Available at: http://93.174.95.29/main/2337000/0c8d12a9b107ddfc530e8d0e0c42c0e8/Steve%20Wexler_%20Jeffrey%20Shaffer_%20Andy%20Cotgreave%20-%20The%20Big%20Book%20of%20Dashboards_%20Visualizing%20Your%20Data%20Using%20Real-World%20Business%20Scenarios-Wiley%20%282017%29.pdf

[Accessed 23 Jul. 2019].

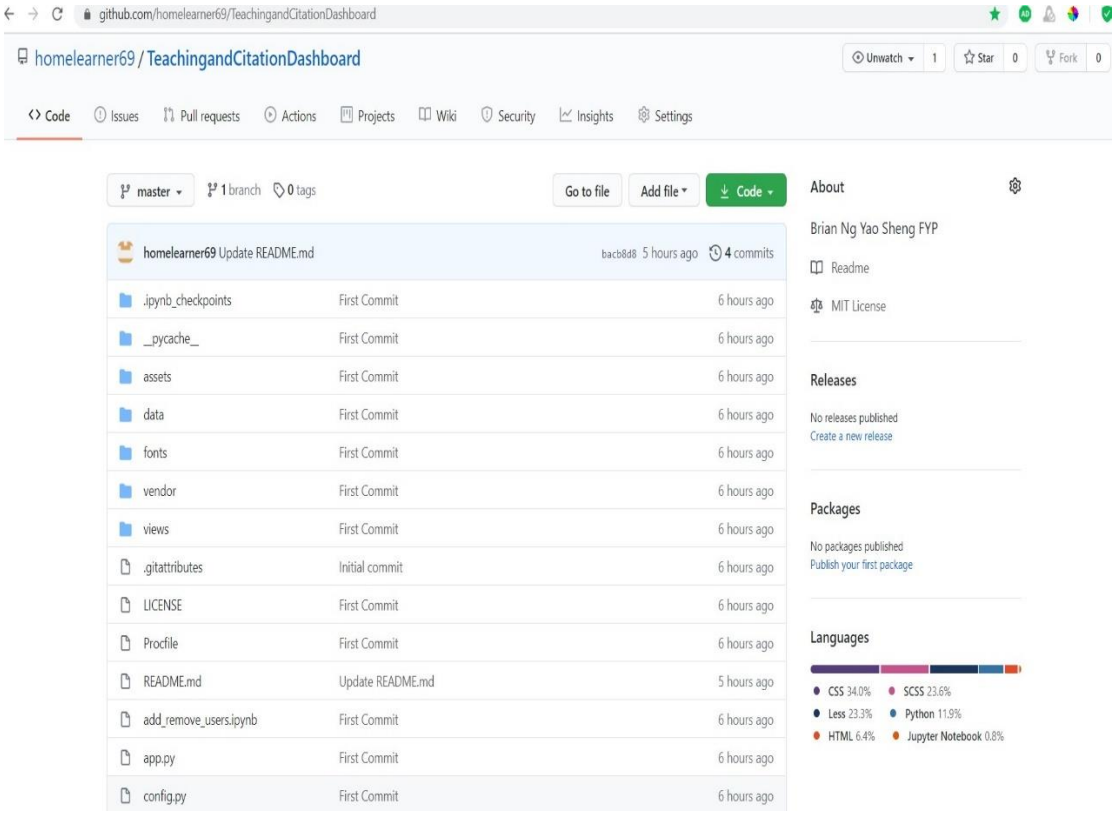
Wolf, M., Taimuty, M., Patel, M. and Meteyer, J., 2016. *The Dean's Information Challenge: From Data To Dashboard*. [online] EducauseReview. Available at: <https://er.educause.edu/articles/2016/11/the-deans-information-challenge-from-data-to-dashboard>

[Accessed 12 July 2020].

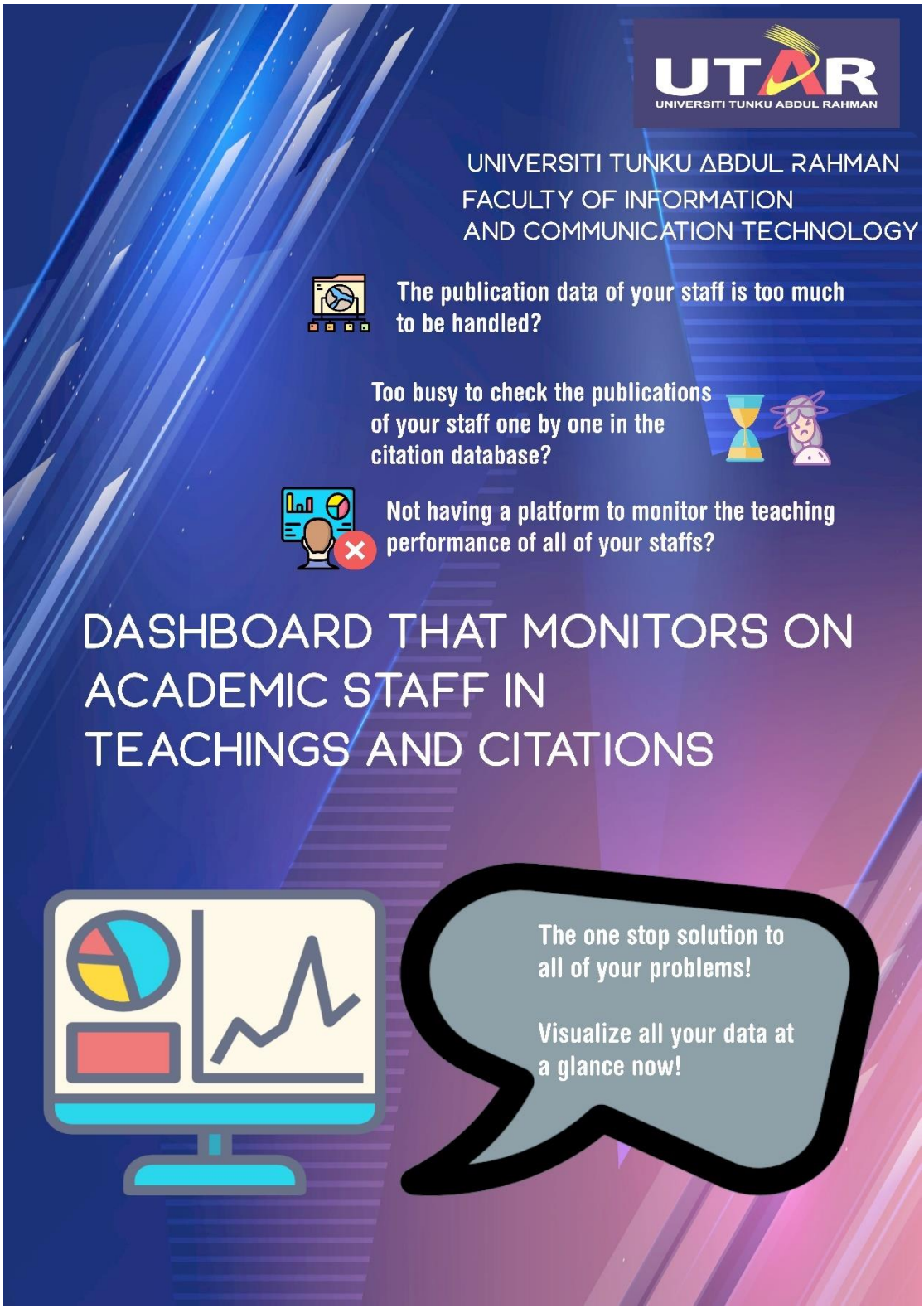
Zentner, A., Covit, R. & Guevarra, D., 2019. *Exploring Effective Data Visualization Strategies in Higher Education*, s.l.: SSR N.

APPENDIX A: FINAL YEAR PROJECT SOURCE CODE (GITHUB)

GitHub Link: https://github.com/homelearner69/TeachingandCitationDashboard



APPENDIX B: POSTER



APPENDIX C: FINAL YEAR PROJECT BIWEEKLY REPORT

FINAL YEAR PROJECT BIWEEKLY REPORT

(Project II)

Trimester, Year: Y3S3	**Study week no.: 1**
Student Name & ID: Brian Ng Yao Sheng 16ACB03597	
Supervisor: Dr. Pradeep Isawasan	
Project Title: A Dashboard for Monitoring Academic Staff Teachings & Citations	

1. WORK DONE

Developing callback function that returns faculty value and department value accordingly

2. WORK TO BE DONE

Developing filter to the dashboard

Redesign the UI of the dashboard

3. PROBLEMS ENCOUNTERED

The filter seems to having problems at the x-axis and still unable to enter UTAR for data collection

4. SELF EVALUATION OF THE PROGRESS

Trying different ways in Jupyter notebook to find possible solution to solve the filter problem

Supervisor's signature

Student's signature

FINAL YEAR PROJECT BIWEEKLY REPORT

(Project II)

Trimester, Year: Y3S3	**Study week no.: 3**
Student Name & ID: Brian Ng Yao Sheng 16ACB03597	
Supervisor: Dr. Pradeep Isawasan	
Project Title: A Dashboard for Monitoring Academic Staff Teachings & Citations	

1. WORK DONE

Developing filter to the dashboard

Redesign the UI of the dashboard

2. WORK TO BE DONE

Redesign the code for returning graph

Connect to scopus API for all utar data

3. PROBLEMS ENCOUNTERED

There are 3 different graphs to be used in plotly dash, it seems like only dcc.graph is working, unable to use plotly express and plotly graph objects

4. SELF EVALUATION OF THE PROGRESS

The data collection of teaching survey and publication data have to be done ASAP, time is short.

Pradeep

Ysmy

Supervisor's signature

Student's signature

FINAL YEAR PROJECT BIWEEKLY REPORT

(Project II)

Trimester, Year: Y3S3	**Study week no.: 5**
Student Name & ID: Brian Ng Yao Sheng 16ACB03597	
Supervisor: Dr. Pradeep Isawasan	
Project Title: A Dashboard for Monitoring Academic Staff Teachings & Citations	

1. WORK DONE Redesign the code for returning graph
2. WORK TO BE DONE Connect to scopus API for all utar data Creating login page and backend for the dashboard
3. PROBLEMS ENCOUNTERED Needs a database to store all of the username password and email
4. SELF EVALUATION OF THE PROGRESS Requires to focus more on how SQLAlchemy works

Supervisor's signature

Student's signature

FINAL YEAR PROJECT BIWEEKLY REPORT

(Project II)

Trimester, Year: Y3S3	**Study week no.: 7**
Student Name & ID: Brian Ng Yao Sheng 16ACB03597	
Supervisor: Dr. Pradeep Isawasan	
Project Title: A Dashboard for Monitoring Academic Staff Teachings & Citations	

1. WORK DONE

Creating login page and backend for the dashboard

2. WORK TO BE DONE

Data collection from the Scopus API

3. PROBLEMS ENCOUNTERED

The data have to be uploaded to github to avoid deployment issues

4. SELF EVALUATION OF THE PROGRESS

Start to study how deployment works and have to enter UTAR to collect data before its too late

Supervisor's signature

Student's signature

FINAL YEAR PROJECT BIWEEKLY REPORT

(Project II)

Trimester, Year: Y3S3	**Study week no.: 9**
Student Name & ID: Brian Ng Yao Sheng 16ACB03597	
Supervisor: Dr. Pradeep Isawasan	
Project Title: A Dashboard for Monitoring Academic Staff Teachings & Citations	

1. WORK DONE

Data collection of publication data and teaching survey data is done

The filter function is done, waiting to add in new data to give it a try

2. WORK TO BE DONE

Data cleaning on both publication data and teaching survey data

Plotting teaching survey data into graphs

3. PROBLEMS ENCOUNTERED

Wordcloud is not a component of dash

4. SELF EVALUATION OF THE PROGRESS

Data collection is quite late, have to hasten up the speed of data cleaning and integrate into the dashboard

Supervisor's signature

Student's signature

10/09/2020

FINAL YEAR PROJECT BIWEEKLY REPORT

(Project II)

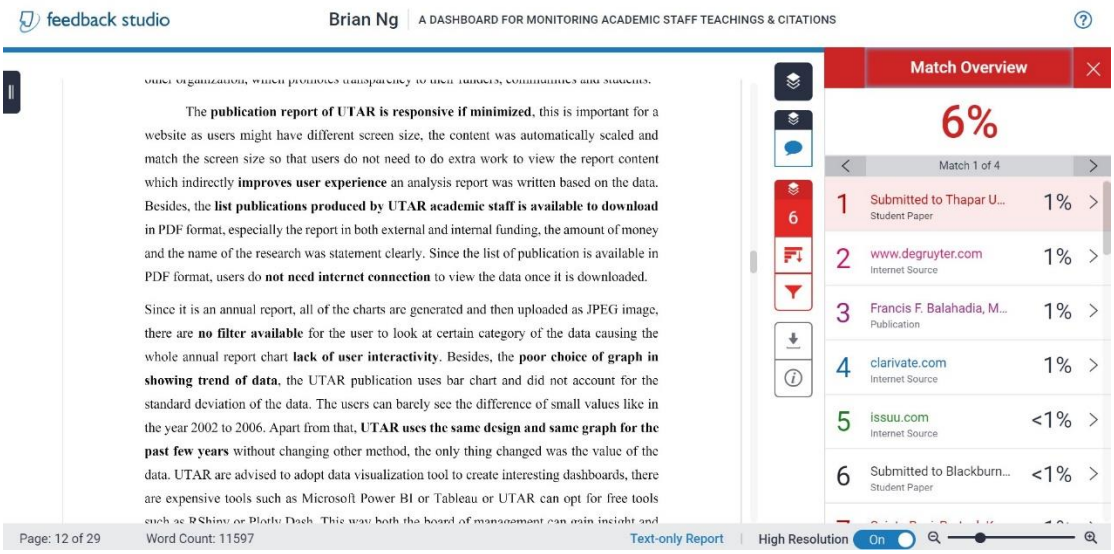
Trimester, Year: Y3S3	**Study week no.: 12**
Student Name & ID: Brian Ng Yao Sheng 16ACB03597	
Supervisor: Dr. Pradeep Isawasan	
Project Title: A Dashboard for Monitoring Academic Staff Teachings & Citations	

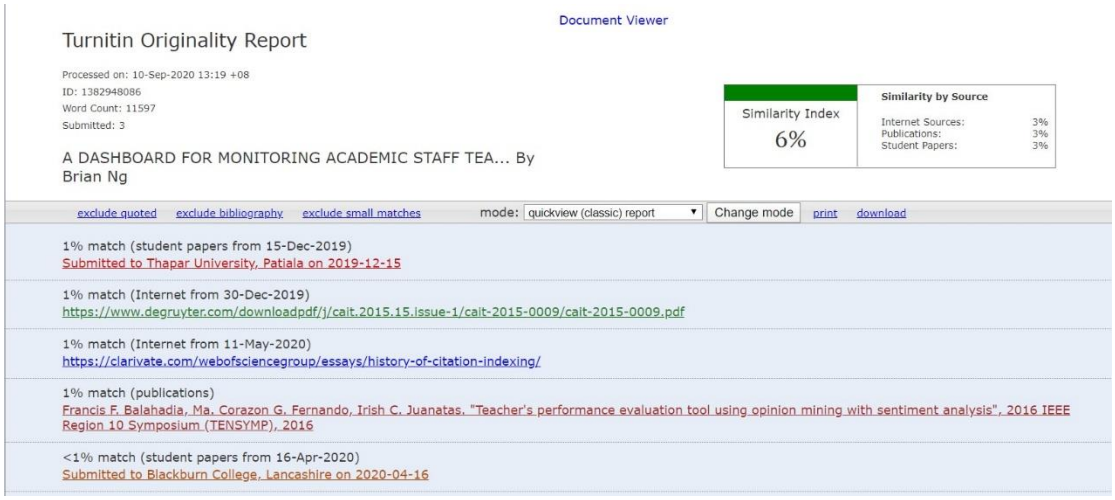
1. WORK DONE The dashboard is now fully functional
2. WORK TO BE DONE
3. PROBLEMS ENCOUNTERED
4. SELF EVALUATION OF THE PROGRESS The development of dashboard is done and deployed

Supervisor's signature

Student's signature

APPENDIX D: PLAGIARISM CHECK RESULT





Document Viewer

Turnitin Originality Report

Processed on: 10-Sep-2020 13:19 +08
ID: 1382948086
Word Count: 11597
Submitted: 3

A DASHBOARD FOR MONITORING ACADEMIC STAFF TEA... By Brian Ng

Similarity Index 6%

Similarity by Source
Internet Sources: 3%
Publications: 3%
Student Papers: 3%

exclude quoted | exclude bibliography | exclude small matches | mode: quickview (classic) report | Change mode | print | download

1% match (student papers from 15-Dec-2019)
Submitted to Thapar University, Patiala on 2019-12-15

1% match (Internet from 30-Dec-2019)
https://www.degruyter.com/downloadpdf/j/cait.2015.15.issue-1/cait-2015-0009/cait-2015-0009.pdf

1% match (Internet from 11-May-2020)
https://clarivate.com/webofsciencegroup/essays/history-of-citation-indexing/

1% match (publications)
Francis F. Balahadia, Ma. Corazon G. Fernando, Irish C. Juanatas, "Teacher's performance evaluation tool using opinion mining with sentiment analysis", 2016 IEEE Region 10 Symposium (TENSYMP), 2016

<1% match (student papers from 16-Apr-2020)
Submitted to Blackburn College, Lancashire on 2020-04-16

APPENDICES

Universiti Tunku Abdul Rahman			
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.: 1of 1

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY

Full Name(s) of Candidate	Brian Ng Yao Sheng
ID Number(s)	16ACB03597
Programme / Course	Bachelor of Computer Science (HONS)
Title of Final Year Project	A Dashboard for Monitoring Academic Staff Teachings & Citations

Similarity	**Supervisor's Comments (Compulsory if parameters of originality exceeds the limits approved by UTAR)**
Overall similarity index: 6 % **Similarity by source** Internet Sources: 3 % Publications: 3 % Student Papers: 3 %	
Number of individual sources listed of more than 3% similarity: 0	

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

Note: Parameters (i) – (ii) shall exclude quotes, bibliography and text matches which are less than 8 words.

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.

Pradeep

Signature of Supervisor

Name: PRADEEP ISAWASAN

Date: 10/09/2020

Signature of Co-Supervisor

Name: _____

Date: _____

APPENDIX E: FYP2 CHECKLIST

UNIVERSITI TUNKU ABDUL RAHMAN

FACULTY OF INFORMATION & COMMUNICATION TECHNOLOGY (KAMPAR CAMPUS)

CHECKLIST FOR FYP2 THESIS SUBMISSION

Student Id	16ACB03597
Student Name	BRIAN NG YAO SHENG
Supervisor Name	DR. PRADEEP ISAWASAN

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.
✓	Front Cover
✓	Signed Report Status Declaration Form
✓	Title Page
✓	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)
✓	Poster
✓	Signed Turnitin Report (Plagiarism Check Result – Form Number: FM-IAD-005)

*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: 10/09/2020	Supervisor verification. Report with incorrect format can get 5 mark (1 grade) reduction. (Signature of Supervisor) Date:10/09/2020