

Design and Development of a Learning Analytics Toolkit

Claris Kerlyn Ke Yue'e - 2109935K

School of Computing Science
University of Glasgow

Level 4 Project — March 26, 2015

Project Supervisor: Dr Andrew Ng

Education Use Consent

I hereby give my permission for this project to be shown to other University of Glasgow students and to be distributed in an electronic format. **Please note that you are under no obligation to sign this declaration, but doing so would help future students.**

Name: _____ Signature: _____

Abstract

Learning analytics is an emerging educational technology that involves measuring, collecting, analyzing and reporting learner-related data to monitor and predict learner's academic performance. Learning analytics provides early interventions to improve learners' learning and performance.

This project aims to design and develop a learning analysis toolkit that can analyze and report learner-related information. Regression and intervention methods such as sending email and SMS to respective students would be beneficial. The project outcomes not only benefits schools academic staffs, it can be used in organization to further monitor individuals and company performance.

Acknowledgements

I would like to thank Prof. Andrew Ng Keong for his invaluable support, advice and guidance throughout the year. I would also like to thank my family and friends I've come to know in Computing Science for helping keep me sane and on track throughout the past 2 years. Lastly, I would also like to thank the participants who took part in the evaluation.

Contents

Abstract	i
Acknowledgements	ii
List of Figures	viii
List of Tables	x
List of Listings	xi
1 Introduction	1
1.1 Learning Analytics Toolkit (LAT)	1
1.2 Aims and Objectives	1
1.3 Motivation	2
1.4 Outline	2
2 Review of Literature	3
2.1 Investigation of Existing Tools	3
2.1.1 Blackboard Analytics for Learn	3
2.1.2 Modular Object-Oriented Dynamic Learning Environment (Moodle)	4
2.2 Analytic Methods	4
2.3 Types of Data	5
2.3.1 Structured and Unstructured Data	6
2.3.2 Data Attributes	6
2.3.3 Processing Distorted Data	6

2.4	Exploratory Data Analysis	6
2.4.1	Histograms vs Bar Chart	7
2.4.2	Box Plots vs Notched Box Plots	7
2.4.3	Quantile-Normal Plots vs Normal Probability Plots	8
2.4.4	Scattered Plots vs Line Chart	8
2.5	Descriptive and Predictive Modelling Techniques	9
2.5.1	Descriptive Models Techniques for Educational Data	9
2.5.2	Predictive Models Techniques for Educational Data	10
2.6	Interventions	10
2.6.1	Effective Interventions Category	11
2.6.2	Peer Support	11
2.6.3	Address Poor Achievements	11
3	Requirement	12
3.1	Questionnaire	12
3.1.1	Data Analysis	12
3.2	Data Collection	13
3.2.1	Population Selected for Analysis	13
3.2.2	Data Sources	13
3.2.3	Data Gathering	13
3.2.4	Activity Tracking and Definitions	14
3.2.5	Activity Weighting	14
3.3	Functional Requirement	14
3.4	Non-functional Requirements	15
4	Design	16
4.1	User Experience	16
4.1.1	Preparing to Analyze Student Data	16
4.1.2	Preparing an Analyze	16
4.1.3	Wireframe Design	17

4.1.4	Early Intervention with Learning Analytics	17
4.1.5	Scenarios	18
4.1.6	User Feedback	18
4.2	User Interface	19
4.2.1	Initial Graphical User Interface	19
4.2.2	Product Wireframes	19
4.2.3	Data Representation	20
4.2.4	Product Logo Design	21
4.2.5	User Feedback	22
4.2.6	Application Final Design	23
4.3	Technologies Used	24
4.3.1	Microsoft Excel	24
4.3.2	SMS Server	24
4.3.3	SMTP Server	24
4.3.4	MATLAB Applications	25
5	Implementation	26
5.1	Initial Implementation	26
5.1.1	Creating Graphical User Interfaces in MATLAB	26
5.1.2	Home Screen	26
5.1.3	Summary Tab	27
5.1.4	Individual Tab	28
5.1.5	Risk Quadrant Tab	29
5.1.6	Intervention Tab - Send Email	30
5.2	User Feedback and Refinement	30
5.2.1	Evaluation Design	30
5.2.2	Results Gathered	30
5.2.3	Additional features to implement	31
5.3	Second Implementation Cycle	31

5.3.1	Addressing identified defects	31
5.3.2	Linear Regression	32
5.3.3	Risk Quadrant Tab - Success Index	34
5.3.4	Intervention Tab - SMS Notification	35
5.4	Summary	36
6	Evaluation	37
6.1	Usability Evaluations	37
6.1.1	Think Aloud Evaluation	37
6.1.2	SUS Evaluation	38
6.2	Dataset Evaluations	39
6.2.1	Application Accuracy	39
6.2.2	Efficiency of Sampling Dataset	39
7	Conclusion	40
7.1	Summary	40
7.2	Future Development	40
7.3	Lessons Learnt	40
	Bibliography	41
	Appendices	43
A	User Guide	A1
B	Data File Header	B1
C	Requirements Survey Results	C1
D	Usability Evaluation	D1
D.1	SUS Scoring	D1
D.2	Think Aloud Evaluation Task	D3
D.3	Evaluation Survey Results	D4

E Full Images

E1

F Ethics Checklist

F1

List of Figures

2.1	Monitoring performance against goal	3
2.2	Motivate students with comparisons	3
2.3	Engagement Analytics Report example	4
2.4	Engagement Analytics Block example	4
2.5	Different between Histograms and Bar Chart	7
2.6	Different between Box and Notched Box	7
2.7	Different between Quantile-Normal Plots and Normal Probability Plots	8
2.8	Different between Scattered Plots and Line Chart	9
4.1	Initial Planning Stage	17
4.2	Application Wireframe Planning	17
4.3	Application's Wireframes	18
4.4	Creating a new Matlab GUI	19
4.5	Creating a GUI with GUIDE	19
4.6	Application's Wireframes	20
4.7	Individual Tab's Data Representation	21
4.8	Regression Data Representation Tab	21
4.9	Application Logo Designs	22
4.10	Application's Final - Menu and sizing was enlarged and more inputs added	23
4.11	Application's Matlab	23
4.12	Architecture Diagram	24
5.1	Application main screen with custom tab menu bar	27
5.2	MATLAB Custom Views	27

5.3	Application Summary's Plot Screenshots	28
5.4	Application ranking Student's Position and Grade based on Total Marks	29
5.5	Application Risk Quadrant Screenshots	29
5.6	Graph Comparison between student grades and best fit line	33
5.7	Application Regression tab screenshots	34
5.8	Application's Risk Quadrant finalized screenshot	35
5.9	Application Early Intervention SMS Notification Screenshots	36
5.10	Application's Overall Diagram	36
6.1	SUS Score	38

List of Tables

2.2	Examples of Descriptive and Predictive Modelling Techniques	9
2.3	Improve engagement and reducing early school leaving	11
3.1	Some of the Student Data Capture in xSiTe	14
3.2	Some of the Student Activity Logs	14
5.1	Different regression's coefficient and norm of residuals comparison	32
6.1	The application received an SUS score of 88.96 denoting 'Excellent usability'	39
6.2	Time captured with different data sample size on the application	39

Listings

5.1	Begin initialization code	26
5.2	Create a uitabgroup and five uitabs in GUI.	27
5.3	Categorising of Engineering and Non-Engineering Students.	27
5.4	Primary code in used MATLAB.	28
5.5	Pseudo code to categorize student grade indicator.	28
5.6	Drawing Mean Lines for X and Y Axis.	29
5.7	Sending email using a gmail account in matlab.	30
5.8	Prompted with Message box when successful.	30
5.9	Curve Fitting MATLAB polyfit functions	32
5.10	Functions to calculate norm of residuals in MATLAB.	32
5.11	Curve Fitting MATLAB functional form.	33
5.12	Calculate marks based on GPA entered.	33
5.13	Changing colour format in Risk Quadrant	34
5.14	PHP Script to Communicate with server.	35
5.15	MATLAB function to Send SMS.	35

Chapter 1

Introduction

1.1 Learning Analytics Toolkit (LAT)

Learning analytics toolkit (LAT) is known as Virtual Learning Environments [11]. It is an emerging educational technology that is widely used which involves methods such as measuring, collecting, analyzing and reporting learner-related data to monitor and predict learners' academic performance. This is useful for targeted audience such as educators and administrators working in the education system. It provides early intervention to improve the learners' learning and performance.

The integration of classic teaching methods with technology enhancements will improve and enhance teaching many students in one module. In traditional teaching approach, the educator upload teaching contents such as lecture slides, tutorial and practical into the virtual learning environment. In addition, they could share video and audio recording, discussion and tests to enhance interaction with their students. If educators put in more effort in the design of such learning activities, they require tools that help them to observe and keep track of the behavior and performances of students. Educators will be able to measure the level of improvement that was established within the education environment.

Learning analytics offers personalized learning, which would enable students to have more efficient learning experiences. This personalized learning experience is essential in succeeding the theory and exercise of various program coordinators that learners begin the program at the same stage and progress through it at the same pace [9]. Learning analytics offers certainty for predicting and improving student success and recognition in part because it enables educators and schools to make data-driven decisions about student achievement and retention. Data decision-making involves making use of data, such as the sort provided in Learning Management Systems (LMS), to notify educator's predictions.

LAT looks beyond the data used in learning analytics that includes attendance monitoring, class participations, grades and more. It includes the regularity of LMS tool use, obtaining discussion board posts, and the number of times taking quizzes. If educators can build a model of successful learner behaviors, they could support learners to engage in these behaviors. Alternatively, they can also identify at-risk learners as ones who diverge from this model.

1.2 Aims and Objectives

This project aims to design and implement a Learning analytics toolkit to forecast early prediction and identify students that require extra help. Educators and administrators can maximize strategies and target recruitment on

efficiency and effectiveness measurements. It improves individual students' ability to reflect and keep track on their pattern of behavior's compared to others. Lastly, it helps to boost educators and administrator staffs to plan and create different interventions to target each and groups.

The intention of the project is to develop a predictive model and meaningful analysis plots based on the student activities and grades gathered. These data are collected from students after completing the assignments and quizzes as part of their learning work. Students will be clustered into groups based on their individual performance. Also, early interventions such as sending email and SMS notifications to targeted students should be deployed.

1.3 Motivation

LAT enable teachers and administrators to tailor educational opportunities to individual student's level of need and ability. LAT ensures to harness data mining, interpretation, and tailor education to individual students more effectively. Learning does not confine to only in the education system; it is also applicable to the different kind of organizations, which can use the information and learning process to drive learning and development of all employees. LAT will help organizations to focus on the positive predictive performance level and to achieve higher productivity.

1.4 Outline

The remainder of the report will cover the following:

- **Review of Literature** - This chapter of the report will include existing products, available tools and analysis methods available, different type of prediction methods will be discussed as well.
- **Requirements** - This chapter of the report will contain other analysis methods that are possible, questionnaire gathered and functional and non-functional requirements.
- **Design** - This chapter of the report will contain diagrams detailing how the framework works that includes initial planning stage of the prototype, wireframes and more.
- **Implementation** - This chapter of the report will discuss the implementation process, problems encountered, techniques used and a general outline of how the program runs.
- **Evaluation** - This chapter covers how the task was tackled, problems encountered at each stage and how they were resolved, and giving insight on how the project could have been improved and achieved.
- **Conclusion** - This chapter summarizes the points mentioned above, future enhancement, and challenges faced.

Chapter 2

Review of Literature

This chapter discusses several different existing artifacts, different type of analysis and prediction methods available.

2.1 Investigation of Existing Tools

There are a few similar tools and learning resources available for educational purpose. This section discusses some of the current tools available and used in majority of the education institutions.

2.1.1 Blackboard Analytics for Learn

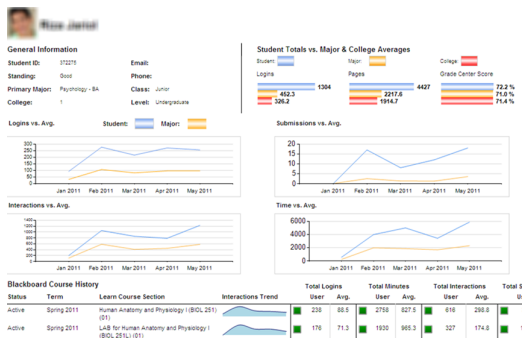


Figure 2.1: Monitoring performance against goal

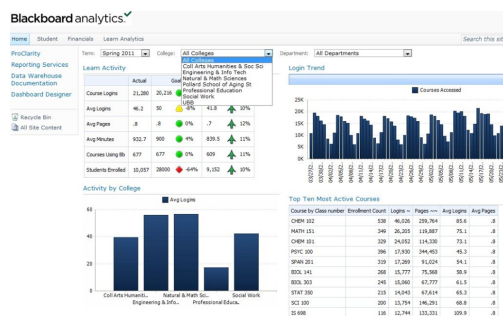


Figure 2.2: Motivate students with comparisons

Blackboard Analytics¹ offers a comprehensive suite of packaged analytics applications, each consisting of Pre-built data integration with leading student information and learning management systems, including PeopleSoft, Ellucian Banner and Ellucian Colleague [16]. Blackboard Analytics cover all areas in the Institution. Blackboard Analytics product modules cover Enrollment Management, Analytics for Learn, Finance and more [17].

Analytics for Learn was invented to support institutions enhance student success by establishing real-time data that helps distinguish patterns of student activity and achievement immediately into the hands of targeted audience to take action. It provides a wide variety of immediate and customizable way to data at the school, department, course, instructor and student level. This self-service system benefits decision-makers focus restricted resources on the policies and programs that promote convincing student practices [15].

¹Blackboard Analytics - <http://www.blackboard.com/Platforms/Analytics/Overview.aspx>

However, Analytics for learn is very costly, as there is license and yearly maintenance cost. It is also compulsory to join blackboard partnership, which includes four different pricing levels: Building Block, Signature, Premier, and Strategic [18]. It is also challenging to build partnership between IT and institutional research (IR) [28].

2.1.2 Modular Object-Oriented Dynamic Learning Environment (Moodle)

Moodle² is a free software e-learning platform, also known as Virtual Learning Environment. It serves to help educators create online courses with a focus on interaction and collaborative construction of content. Moodle also allows several plugins such as "Engagement Analytics" (see Figure 2.3) which is very similar to the project. The engagement analytics [24] block provides information about student progress against a range of indicators. This plugin also applies to activities, which have been distinguished by current study to have an impression on learner success in an ongoing course.

Engagement analytics for course: MBA - HRM

Scores are shown as "weighted score (raw score)"

Username	Assessment Activity	Forum Activity	Login Activity	Total ↑
Ken Student	25% (50%)	30% (100%)	20% (100%)	75%
Frances Smart	10% (20%)	30% (100%)	4% (20%)	44%
Dougal McFarlane	2% (4%)	30% (100%)	4% (20%)	36%
Angela Edgar	25% (50%)	30% (100%)	20% (100%)	75%

Figure 2.3: Engagement Analytics Report example

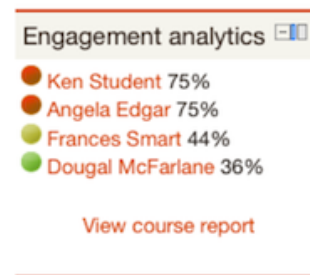


Figure 2.4: Engagement Analytics Block example

However, most of the data is generally held across different system and departments (such as Library, IT, Student Registry, Finance). The goal of the plugin is to produce education with information only, it does not automatically take any action based on the indicators automatically. Also, the plugin would be beneficial if all student data is deployed in Moodle. Installing and understanding the plugin is usually be vexing too.

Though Moodle tracking figures indicated that the majority of users were regularly engaging with the assistance, it was impracticable to observe the extent to which the e-portfolios were being practiced, which of the Mahara tools alumni were operating.

2.2 Analytic Methods

Learning Analytic utilizes different techniques to detect unusual designs isolated in educational data sets. A few different type of available methods are discussed as following:

Content analysis

Content analysis [32] is the analysis of texts of various types including writing such as essays, images, recordings and cultural artifacts. Content analysis includes both qualitative and quantitative approaches. Content analysis is used for a variety of purposes including attribution of texts to authors, testing of hypotheses, theory building, and evaluation research.

²Moodle - <https://moodle.org/>

Discourse Analytics

Discourse analytics aims to capture meaningful data on student interactions which (unlike 'social network analytics') aims to explore the properties of the language used, as opposed to just the network of interactions, or forum-post counts, etc.

Social Network Analytics

Social Network Analysis (SNA) is the checking of connections and flows between people, groups, organizations, computers, URLs, and other connected data entities. SNA produces both a visual and mathematical analysis of social relationships. Management experts use this methodology with their enterprise clients, giving perspicacity into the various roles and groupings in a network [19].

Social Learning Analytics

Social Learning Analytics (SNA) explores the role of social interaction in learning, the importance of learning networks, discourse used to sensemake [21]. As social networks [5] become necessary to support networked learning, enabling and analyze these systems are gaining popularity. By expressing a social network visually, attractive connections could be discussed and explored in a user-friendly manner. SNA is the quantitative research of the relations between individuals or organizations. One of the key features of networks is centrality, which correlates to the structural position of a node within a system and reveals the prominence of a node and the nature of its association to the remainder of the program. Three centrality measures called degree, closeness, and betweenness centrality are widely used in SNA.

Disposition Analytics

Disposition Analytics seeks to capture data regarding student's dispositions to their own learning, and the relationship of these to their learning [19]. For example, "curious" learners may be more inclined to ask questions - and this data can be captured and analysed for learning analytics.

2.3 Types of Data

There are boundless quantities of data that educational organizations have at their disposition to aid educators meet their intentions of enhancing students' achievement and increasing maintenance.

Table 2.1 illustrates the types of available data in the LMS. The first column includes data that is regularly produced by Learning Management System (LMS) [7]. The second column is an illustration of the types of data that is provided by the educator that can be stored in the LMS.

Data Generated by LMS	Data Generated by Instructor
Number of Times Spend	Grades on Exchange Forum
Date and Time of Access	Assignment Marks
Number of Discussion Posts Generated	Grades on Tests
Number of Discussion Posts Read	Final Grades
Types of Resource Accessed	Questions Asked in Discussion Forum
Topics Mastered	Emails notification to/for Instructor
	Attendance

Table 2.1: Types of Data Available for Learning Analytics

2.3.1 Structured and Unstructured Data

The complexity of structures depends on each individual representation. All data is expressed in the relational schema, and the data structure consists of the attribute-value pairs. In various data representations, the arrangements can be further involved such as characteristic hierarchies or aggregates of attributes. In educational domain, the data collected is usually structured, but seldom unstructured data is inevitable. For instance, when analyzing students' application codes or learning diaries, it is recommended to first extract meaningful features.

2.3.2 Data Attributes

Data types classify the attributes corresponding to their domains. The primary categories are to numerical (quantitative) and categorical (qualitative) data types. Numerical data has significance as numbers, and order, distance and equality of two numerical variables. In opposition, categorical data has no meaning as figures only correspondence within two definite variables.

Numeric data is organized as discrete or continuous. Discrete data countably many values, and if the domain is portrayed on the number line, it consists only of isolated cases. Discrete numbers are not always integers as it can be decimal numbers with a predefined precision. The categorical variables can be further classified as nominal or ordinal. The nominal values are only named for sections and have no other meaning than classifying.

2.3.3 Processing Distorted Data

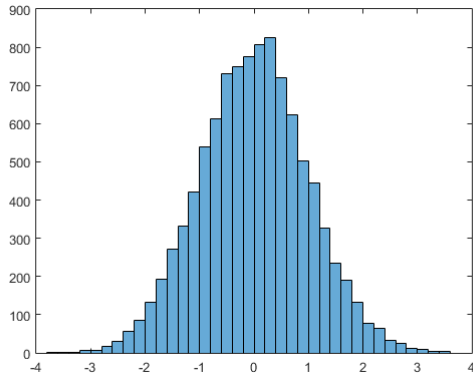
Processing distorted data is called data cleaning. The aim is to decrease the influence of different characteristics like missing or incorrect values, duplicates or inconsistencies [7]. During the process of consolidating data from different sources, it is common that some tuples are defined as missing attributes. There are two fundamental solutions: by either delete the whole tuple or repair the missing value with some null values or mean value in the entire population or some subclass.

The differences may be due to different formats, updates, or errors. Normally, it is best to use domain information or ask an expert (teacher) to adjust the values. If these inconsistencies could not fix, the only solution was to delete them. For instance, it is common that only the most active students participate in voluntary exercises. Presumably the students had engaged in several exams and the points were not updated to all records.

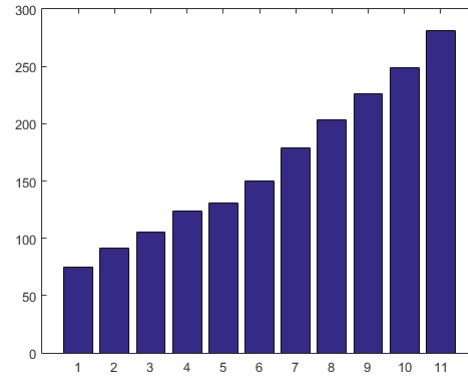
2.4 Exploratory Data Analysis

Univariate graphical Exploratory Data Analysis (EDA) is to determine correlations between the explanatory variables, and evaluating the regulation and estimated size of relationships among explanatory and an outcome variable [27].

The data that come from making a precise measurement on all of the subjects in a sample represent observations for characteristics such as marks, activities, or acknowledgment to a stimulus [31]. Graphical methods are more qualitative and involved a degree of subjective analysis.



(a) Histograms



(b) Bar Chart

Figure 2.5: Different between Histograms and Bar Chart

2.4.1 Histograms vs Bar Chart

Histogram is a bar plot that each bar represents the frequency or proportion for a range of utilities. Figure 2.5a explains the bars run vertically with the count axis running vertically. To manually construct a histogram, represent the series of data for each bar, count how many cases fall in each bin, and represent the bars high enough to indicate the number.

A bar chart is a data representation with rectangular blocks with dimensions proportional to the representing values. Figure 2.5b displayed is vertical or horizontal which shows comparisons among categories. Some bar graphs present bars clustered in groups of more than one and others show the bars divided into smaller parts to show the cumulative effect (stacked bar graphs).

2.4.2 Box Plots vs Notched Box Plots

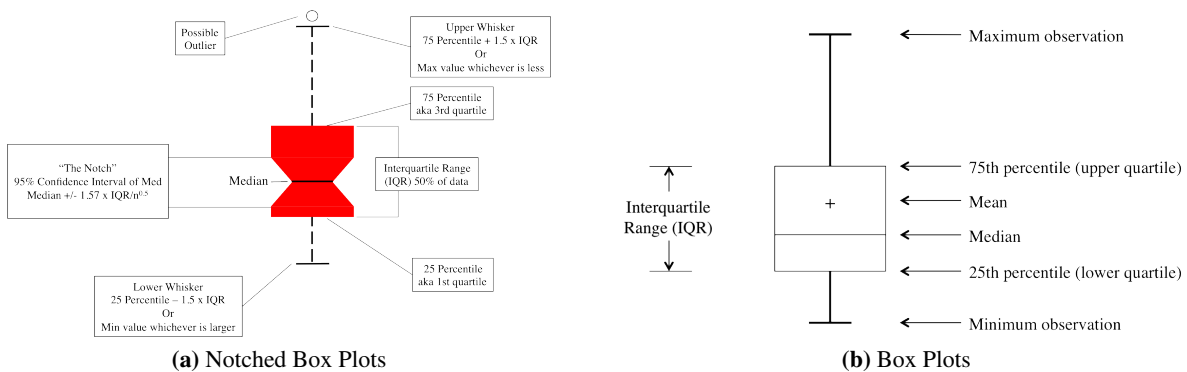


Figure 2.6: Different between Box and Notched Box

Box plots are comprehensive EDA plots because they rely on robust statistics like the median and interquartile range rather than as mean and standard deviation. The Notched box plot is useful and recognizable at a quick glance. Box plots it is easy to compare distributions because of their accurate and reliable statistics. Figure 2.6b consists of a "box", which goes from the first quartile to the third quartile. Within the box, a vertical line is painted at the median of the data set. Two horizontal lines, called whiskers, extend from the front and back of the box.

A Notched Box plot is a data representation of a box, a line in the box, and two whiskers. The box displays the interquartile range (IQR) from 25 to 75 percentiles as shown in Figure 2.6a. The IQR is where the centre 50 percent of the data points will fall. The whiskers add 1.5 times to the 75 percentile and subtract 1.5 times the IQR from the 25 percentile. The whiskers should include 99.3 percent of the data if from a normal distribution.

2.4.3 Quantile-Normal Plots vs Normal Probability Plots

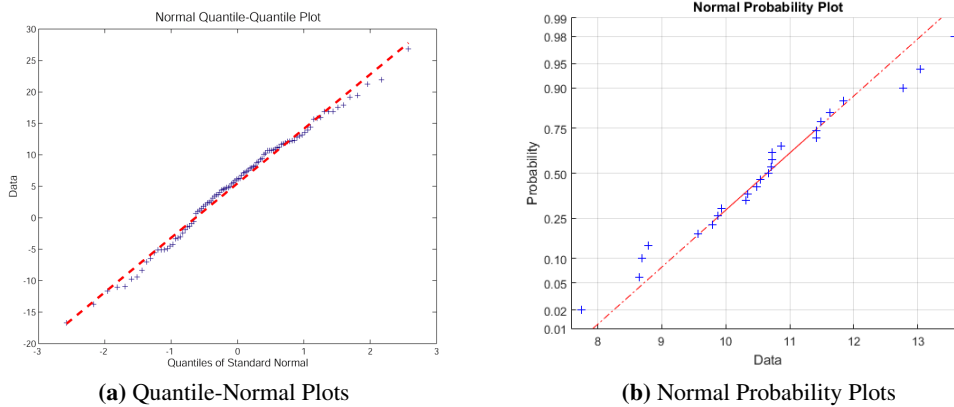


Figure 2.7: Different between Quantile-Normal Plots and Normal Probability Plots

Quantile-Normal Plots are used to analysis how a particular sample follows a particular theoretical distribution. Although it can be used for any theoretical distribution, it limits the attention to understanding how adequately a sample of data of size n resembles a Gaussian distribution with mean and variance resembling to the sample mean and variance. By examining the quantile-normal plot detects left or right skew, positive or negative kurtosis, and bimodality.

The normal probability plot is to identify and identifying substantive departures from normality. Normal probability plots are produced up of raw data, residuals from model fits and estimated parameters. The sorted data plotted vs. values elected to make the image look like a straight line if the data are approximately normally distributed.

2.4.4 Scattered Plots vs Line Chart

For two quantitative variables, the underlying graphical EDA technique is the scatterplot, which has one variable on the x axis, one on the y axis and a point for each case in dataset. It is a very strong convention to put the outcome on the y (vertical) axis. Figure 2.8a demonstrate how one variable is affected by another. The association between two variables is called correlation.

Figure 2.8b displays information as a set of data points called markers connected by a straight line segments. It is a fundamental type of chart standard in many platforms. It is similar to a scatter plot. However, line chart's points are ordered and joined with straight line segments. A line chart is often used to visualize a trend in data over intervals of time; thus the line is drawn chronologically.

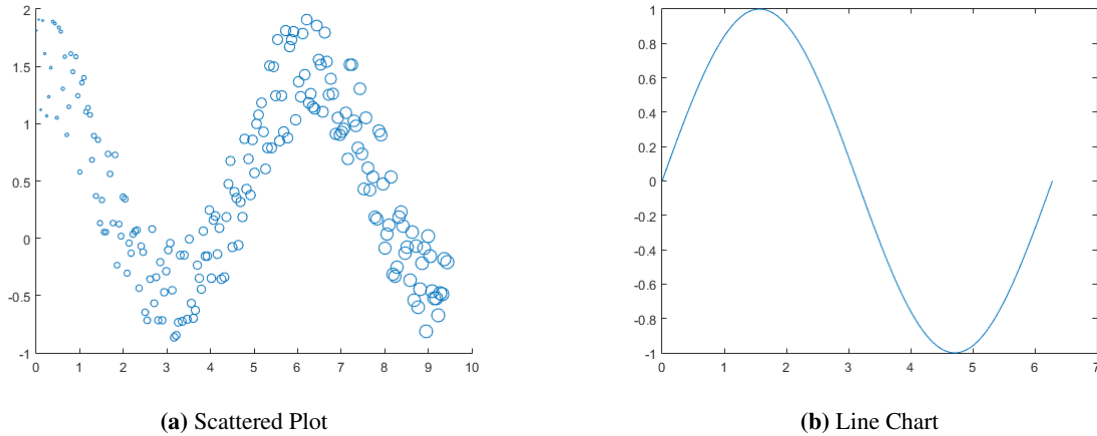


Figure 2.8: Different between Scattered Plots and Line Chart

2.5 Descriptive and Predictive Modelling Techniques

Descriptive modelling is the learning of new knowledge in educational data and predictive modelling predicts new learning results. The descriptive models show the suitability of the similar predictive model and guide the search. Table 2.2 illustrates the examples of descriptive and predictive modelling paradigm pairs.

Descriptive Techniques	Predictive Techniques
Correlation analysis	Linear regression
Associative rules	Probabilistic rules
Clustering	Classification
Episodes	Markov models

Table 2.2: Examples of Descriptive and Predictive Modelling Techniques

2.5.1 Descriptive Models Techniques for Educational Data

Correlation Analysis

Performing correlation analysis for numeric data is recommended as undesirable correlations will corrupt the model. It is the easiest technique that explains linear dependencies among two variables. In addition, if the data distributed, the lack of correlation means also statistical independence and no other tests are required.

Associative Rules

Association rule learning is to identifying unusual relationships between variables within extensive databases [1]. It is intended to identify active rules found in databases using different measures of interestingness. Based on the concept of positive rules, it proposed association rules for discovering regularities between goods in large-scale transaction data reported by point-of-sale (POS) policies in supermarkets.

Clustering

Clustering is the classifying a set of objects in such a way that objects in the same cluster are more similar to each other. It is a primary job of exploratory data mining and a common technique for statistical data analysis. It was practised in machine learning, pattern recognition, image analysis, information retrieval, and bioinformatics.

Episodes

Episode rules are a modification of the concept of association rules and frequent sets, applied to sequential data

[2]. Sequential data, such as text, is viewed as a sequence of pairs where feature vector made of a fixed set of features and index contains information about the position of the word in the series. It is common practice that the sequence represents in an increasing order of the indices. A feature can be a word (base form, inflected word form, stem), grammatical or other special character.

2.5.2 Predictive Models Techniques for Educational Data

Linear regression

Regression patterns are one of the most utilized statistical analysis methods in educational study [10]. It is implemented to examine and study the correlations among a set of independent variables based on a particular population and dependent variable. Regression models helps to find the best fit and analysis possible predictors for a provided set of data.

However, Linear Regression models are straightforward to produce a reasonable interpretable description regarding inputs affecting outputs [13]. Linear regression is the relationship between a scalar dependent variable y , and other explanatory variables (or independent variable) denoted X . For more than one explanatory variable, the method is called multiple linear regression.

In linear regression, data are modeled using the linear predictor functions, and unknown model parameters predicted from specific data. Linear regression refers to a model that the dependent mean of y given the value of X is an affine function of X . It is also a model in which the median and quantile of the dependent distribution of y given X represent as a linear function of X .

Probabilistic Rules

Probabilistic Rule learn consistent rules from deterministic patterns but do not take into account contingency [26]. The graphical model and statistical relational learning community reasons concerning uncertainty but have not yet contributed many approaches to learning logical rules. Logical rules are accomplished from probabilistic data and classifications can be probabilistic.

Classification

Classification is the primary obstacle of adaptive learning circumstances. Before choosing tasks, learning material or feedback for the distributed learner current situation, it is proposed to analyze the student's current situation. Usually, all information is presented in the student model conclude the appropriate action.

Markov models

Discovering a complex Markov model is very complicated as the educational datasets are typically small and sparse. In addition, it decreases the search space by analyzing the temporary patterns by identifying means. It is possibly that only some contexts, investigations, or acts are statistically dependent, and there is no model for all potential relations.

2.6 Interventions

Learners often find it challenging to acquire and use a new academic strategy in a short period. At such, Students might not be able to perform well for a certain test or module [29]. A number of obstacles can obstruct students from successfully utilizing tactics. For example, students may originally learn the steps of a approach inaccurately and become depressed when they realized that it did not privileges them. Even if learners enhance skills in using a strategy, they may fail to identify those academic situations. Alternatively, learners may know full well when they are supposed to use an approach but be unambitious to do so.

The most effective learning analytics was proactive during the approach to students, identifying obstacles at an early stage in their education and working to address them before learners become disengaged [20]. Intervention is a kind of systematic process of assessment and planning employed to remediate or prevent educational, or developmental problem for the students. Methods that should be carried out to identify effective interventions are below.

2.6.1 Effective Interventions Category

The interventions are classified into supportive school culture, school-wide strategies, and student-focused strategies. Supportive school culture and school-wide strategies relate to the academic context, to interventions or approaches that are school-wide. Supportive school culture is about the management, leadership, school culture and the commitment of the educational approach to students at risk.

School-wide strategies are about specific policies or initiatives that are adopted school-wide to all students, which help improve student engagement and reduce dropout rates. Student-focused strategies narrowed to the at-risk students than the whole cohort. Table 2.3 illustrates the successful interventions for improving engagement and reducing student drop-out rates.

Supportive school culture	School-wide strategies	Student case management
Continuous improvement Commitment to success for all	Broad curriculum provision Offering quality Education and Training	Mentoring Welfare support
Responsiveness to individual need High expectations	Community service Early intervention to support skill growth	Attendance policies and programs Tutoring and peer tutoring
Encouraging student responsibility	Programs to counter low achievement	Targeted financial support
Shared vision	Pathways planning, guidance and counselling Strategic use of teaching resources	Creative programs for at-risk students Pathways planning for at-risk students
	Smaller class sizes	

Table 2.3: Improve engagement and reducing early school leaving

2.6.2 Peer Support

Peers can assist strengthen confidence in other learners so take recognition of a learning peer. Many teachers implement on this approach. However, a student at risk may have a particular student to approach. Set this up for the learner would benefit him, so he knows whom to ask for clarification before going to the educator.

2.6.3 Address Poor Achievements

There are strong bonds between learners' levels of success and the possibility of excelling in school. Strategies involve objective aid for skill development between low achievers, strategic teacher establishing, and programs to showcase low achievement, discipline, priority professional development, attendance policies and programs. Likewise, educator could schedule a meeting with the targeted students to address on these poor achievements.

Chapter 3

Requirement

This chapter discusses the projects requirements and the variety of ways in which they were elicited.

3.1 Questionnaire

To gain user feedback towards existing learning analytics, an online questionnaire was created using Survey-Monkey¹. The questionnaire included specific questions targeted mostly teaching educators and learners. There are questions for participants who had used learning analytics before. The majority of the questions were closed format, with the exception of a few which elaborations verbally. There were 10 participants in total. 1 is educator and 9 are non-educators. A summary of the results is included below.

3.1.1 Data Analysis

Over half of the participants thinks both teacher and learner will benefit from a Learning Analytics Toolkit

For question 1, *Who will benefit the most from Learning Analytics?*, suggests that everyone is comfortable with the technology of Learning Analytics and willing to use it for his or her daily lives.

Make invisible information about learners visible is the main objective for Learning Analytics

For question 2, *What should the main objective for Learning Analytics be?*, most of the participants state that making invisible information about learners visible should be the most important factor.

Clustering of students can Learning Analytics bring innovations to educational technologies

For question 3, *Which added innovation can Learning Analytics bring to educational technologies?*, few of the participants stated that clustering of student tends to identify weaker students better to provide more assistance including early interventions.

Different frequently used Learning management system were identified

For question 4, *Which IT systems does your organization use for teaching and learning?*, The most popular toolkit among participants were identified as Moodle and Blackboard, which was discussed in Chapter 2.

Almost all participants stated they would only willing to share data to selected member of staffs

For question 5, *Should institutions internally share all the data they hold on students with all members of staff?*,

¹SurveyMonkey - <https://www.surveymonkey.com/>

the majority answered only selective related members. As the data are confidential and should be kept privately within a few relevant members.

Participants feels that teachers are to take action and determine next step for their students

For question 6, *How sure can we be that end users will draw the right conclusions from their data, and decide on the best course of action?*, the majority answered teachers should be the one to take action for their own students.

The full set of questions and results are included in Appendix C.

3.2 Data Collection

Every course offered by Singapore Institute of Technology (SIT) varies in terms of specifications, pace and complexity. For this reason, the author adopted the approach of building course-specific models capable of matching the divergence that exists among courses. The author focuses on year one undergraduate students from Sustainable Infrastructure Engineering (Land), MEngTech and BEng (Hons) undertaking C-Programming module course. The data was collected from September to December 2014 on the first trimester program.

Singapore Institute of Technology (SIT) offers a wide variety of degree programs in many different calendar formats. Some courses are of trimester while others are traditional semester format. For trimesters, one academic year would consist of three trimesters that takes around 4 months each. Each trimester has twelve study weeks which consist of one week team break, two exam weeks and two recess week. This calendar order will remain unchanged throughout the year.

3.2.1 Population Selected for Analysis

The data used for this study was collected from level one students undergoing C-Programming course on going one trimester, which occur between the months of September and December 2013. Special minority populations such as junior college students were eliminated. Likewise, only students who received a grade were included in the final dataset. Students who left the course during the drop/add timeframe were eliminated, as they did not receive a final grade. The total sample size was $N = 66$ students.

3.2.2 Data Sources

Information concerning to online student activity and assignment grades was collected from xSiTe². xSiTe is Singapore Institute of Technology's Learning Management System (LMS) which is powered by Desire2Learn's Integrated Learning Platform. This is equipped with innovative teaching and learning tools that support unique learning preferences. Information pertaining to enrollments, including final Polytechnic GPA and gender was obtained upon enrollment.

3.2.3 Data Gathering

Table 3.1 illustrates xSiTe database containing all relevant information related to student's activities. Information is stored showing when student navigating and accessing xSiTe existing documents such as lecture notes and

²xSiTe - <https://xsite.singaporetech.edu.sg/>

tutorials, number of logins and discussions captured. Other data such as student grades for each tests and Poly-technic Grade Point Average (GPA)³ were requested personally from the educator. Student’s same and matric numbers have been changed into randomized default name and matric numbers to protect the student’s identity.

Content	Discussion	Login history
#Visits	Post read	#Logins
Time Spent (h)	Threads created	
Total Time Spent (h)	Replies posted	
Topics visited (44 max)		

Table 3.1: Some of the Student Data Capture in xSiTe

3.2.4 Activity Tracking and Definitions

XSiTe controls an activity log to record student activity within the system. A list of the logged student activities relevant to this research is shown in Table 3.2. Activity two shows intercommunication with a particular course, as opposed to Activity one that shows intercommunication with the XSiTe system in common. Events three onwards are student activities or engagements as it proposes at least some engagement with a course.

Activity 1	Logged in to XSiTe system
Activity 2	Logged in to course section dashboard
Activity 3	Viewed the course syllabus
Activity 4	Opened an test
Activity 5	Completed an test
Activity 6	Viewed grade comments from educator
Activity 7	Opened a lesson including lecture slides or tutorials
Activity 8	Participants in a Discussion Forum

Table 3.2: Some of the Student Activity Logs

3.2.5 Activity Weighting

When examining LMS activity at multiple points in time throughout the duration of a course, the author proposes that recent behavior should be observed with more weightage than past behavior. For instance, when examining student activity in a trimester, information relevant to a student’s current achievement is more likely to be noticed over the last two or three weeks of logged activity compared to a month [7]. The student activities include the number of visits, hours spent on site and topics mastered. In keeping with this theory, student activities collected from the beginning till the end of the course is defined as the success index. Student background is added to this success index to determine stronger results.

3.3 Functional Requirement

From the various requirements capture activities; a list of functional requirements was formalized. Functional requirements capture the intended behavior of the system, it essentially specifies something the system should do. These requirements were structured in terms of priority, which splits requirements into two priorities: ‘primary’

³Grade Point Average (GPA) - <http://www.back2college.com/gpa.htm>

and ‘secondary’. Requirements are unordered within each priority, as it is likely that multiple requirements have the same importance.

The following **primary** requirements represent functionality the application to be considered a success.

FR	Requirement	Description
1	Observe and analyze data	Users must be able decipher the meaning behind each data and plots.
2	Select different students groups	Users must be able to select type of students they wish to compare.
3	Change graph plots	Users must be able to select different type of data they wish to compare.
4	Select different values	Users must be able to check or uncheck fields.
5	View student summary	Users must be able to view summary of good and poor students.
6	Send email notifications	Users must be able to send email notifications to targeted students.
7	Predict student results	Users must be able to inspect predictive lines generated by the system
8	Select specific files	Users must be able to select files to view upon entering the application
9	Send SMS notification	Users must be able to send SMS notifications to targeted students.
10	Group and cluster students	Users must be identify students based on each different success index.

The following **secondary** requirements are considered desirable but not necessary. They should be included in the final application if feasible within the time allocated to development.

FR	Requirement	Description
11	Import data format	Users should be able to import different format of files without learning the data.
12	Export desired data	Users should be able to export selected data and graphs they wish to for own reference.
13	Refresh search	Users should be able to click on a button to reselect new data they wish to view.
14	Validation user input	Users should be prompted if inputs entered are incorrect.
15	Grade calculator	Users should be estimate student grades based on GPA entered.
16	Zoom support	Users should be able to zoom in or enlarge the graph data.

3.4 Non-functional Requirements

Non-functional requirements describe how well a system performs its goals. They are fundamentally different to functional requirements, which describe the intended behaviour of the system.

NFR	Requirement	Description
1	Quick to add files	The application should be able to support large data files.
2	User-friendly interface	The application should be easy to navigate and use for all users.
3	Allow for future expansion	The application should be build to allow for additional features to be added.
4	Automatic search recovery	The application should automatically save history of each individual users.
5	Fit to page automatically	The application should be build to fit to any screen sizes.
6	Data integrity	Allow users to hide and protect the identify each student data.
7	System availability	Allow users to use the application as and when is needed.
8	User guide and support	The application should automatically provide assistance to users.

Chapter 4

Design

This chapter discusses the design of the application, presenting an overview of the overall system structure, as well as an in-depth account of the user interface design decisions.

4.1 User Experience

Education representatives such as educators and administrators are the targeted audiences for this project. They will be the main users to be using the application, learning analytics. Several storyboards were created to illustrate how the application will satisfy the primary functional requirements proposed on Chapter 3. The first four storyboards focus on the functionality accessible to the viewer. The remaining three storyboards illustrate the application in real-world scenarios.

Each segment of a storyboard focuses on one primary function, which maps to one tab of the application, allowing the flow of the application. The storyboards were roughly sketched with pen and paper and presented to potential users to explain how the application will work. Each storyboard is briefly described below stating the requirements it satisfies and a list of screens required is formalized.

4.1.1 Preparing to Analyze Student Data

Figure 4.1a illustrates the steps required to analyze students using learning analytics. Before using the application, the user (educators or administrators), will need to retrieve the relevant data from the existing student portal into a excel file. After which, the users will have to import the data into the application and wait for the application to be loaded successfully. From this storyboard, the following screens were identified:

- Student Portal (xSiTe)
- Export Data list
- Data viewer (Excel)
- Application MATLAB GUI

4.1.2 Preparing an Analyze

Figure 4.1b demonstrates the application flow when navigating the application. In order to use, a excel file (.xlsx) first need to be importing into the application. Once imported, a user can open the document and by selecting the file from the Drop Down Menu in the application GUI. This allows the users to view desired excel files to analyze

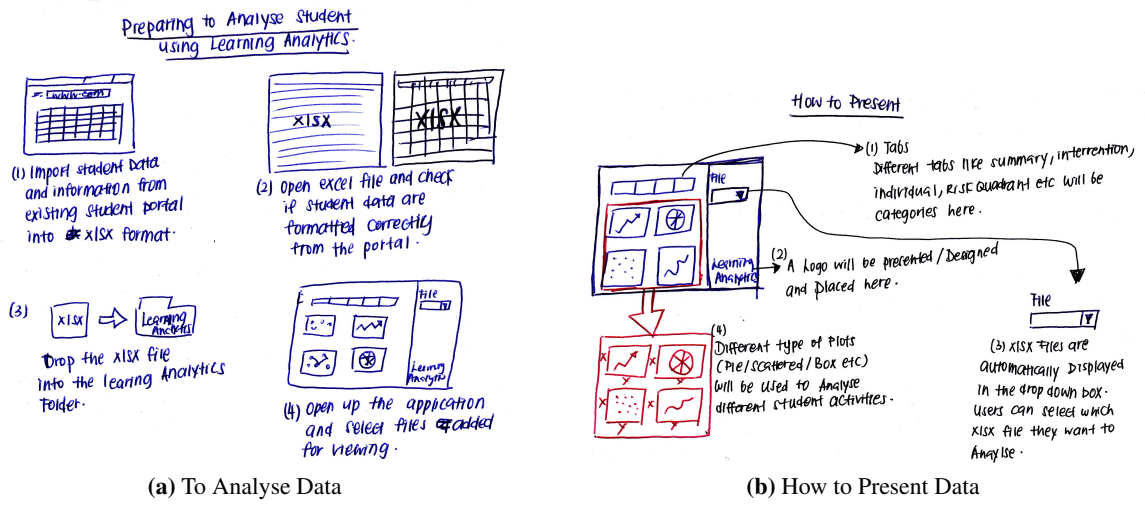


Figure 4.1: Initial Planning Stage

as and when required. The user can view different information panel by selecting the tabs. This storyboard uses many of the screens identified in the first storyboard and also requires the following screens:

- List of Student Data
- Different type of Analysing Plots
- GUI

4.1.3 Wireframe Design

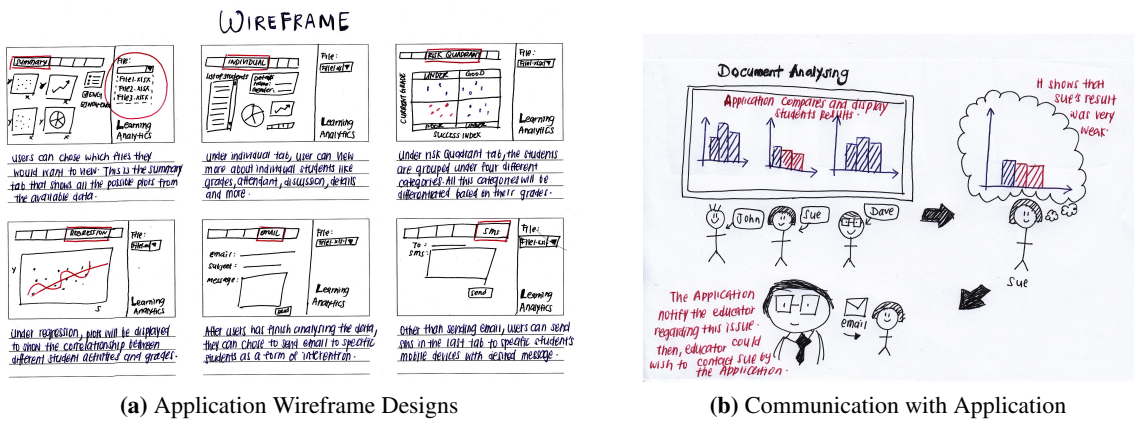


Figure 4.2: Application Wireframe Planning

Figure 4.2a shows the series of actions required to view a learning analytics. After user selects the desired excel file, the application will auto read in the excel file and reads the data and display the analyzes on application page. During analysis, the user can view different type of tabs that includes summary, individual, risk quadrant, regression and intervention. In addition, users are able to examine more about each individual students, and predict student grades with the regression tab. The risk quadrant tab is useful to observe different student clusters in one glance. Lastly, educators are capable to send email to selected students upon the intervention tabs.

4.1.4 Early Intervention with Learning Analytics

Figure 4.2b illustrates how users communicate with targeted students using Learning Analytics. After a user finish analyzing the student data, they might wish to express concern to a group of targeted students. Such early

intervention techniques can be applied to the application itself. The storyboard demonstrates a user sending an email-meeting schedule to a student. The student viewed the emails and responded to the user after class to conduct a meeting. Users can continue to monitor student’s progress and update them through this intervention tab as and when necessary.

4.1.5 Scenarios

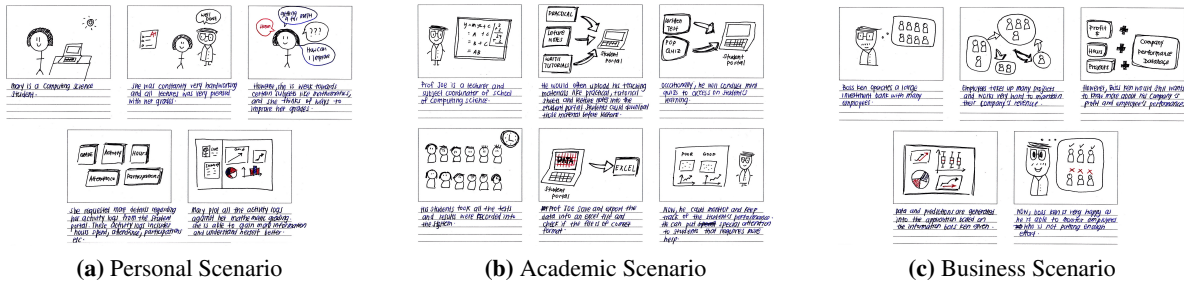


Figure 4.3: Application’s Wireframes

Personal scenario

Figure 4.3a gives an example of the application being used by two individuals: an educator and student. A student is very hardworking and would wish to know ways to improve himself better. He then requests his own activities and grades data from the school portal and analyzes the results. From there, he could wish to consult his educator or even self-monitor the results to strike harder for the next coming examinations.

Academic scenario

Figure 4.3b shows how the application could be used within academic circumstances. An educator downloaded all the student data from the learning portal and imports the file into the application by dragging the file into the folder. Then, he is able to open up the excel file in the application by selecting the excel file name through the drop down list. He then selects the excel file and analyze the student data. From the application, he is able to target the group of students that is performing badly both in their academic grades and activities. From the application, he is then able to select straight into this student’s email address and send them an email regarding this issue and even schedule a meeting to discuss further.

Business scenario

Figure 4.3c describe how the application could be used in a business environment, where it can often be difficult to monitor and analyze employee’s performance. The application is used to view company overall statistics, individual employee’s performance. For instances, these performances could be employee’s monthly profit, attendance, working hours and more. During the analysis, employer could target straight into the specific group of employee which done marvelously well, or relatively bad. After which, the employer can then decides to take whichever actions he wish to the group of employees.

4.1.6 User Feedback

The storyboards were presented to several potential users from different backgrounds. Users were asked to go through the storyboards and state if any parts were confusing all users were able to understand the storyboards and were able to follow the storyboards without asking much questions, which implies that the flow of the application is logical. They mentioned that the application would definitely save a lot of money and precious time. Students also commented that the application would be good to monitor their academic performance and motivate them to accomplish better results.

4.2 User Interface

This section illustrates the transition from requirements and key screens to detailed screen designs. Research regarding the available Graphical User Interface (GUI) platform was done to determine the platform conventions and constraints.

4.2.1 Initial Graphical User Interface

The graphical user interface enables users to click and navigates objects with a mouse instead of typing text at a command line [30]. The primary element of today's portable computing technologies is the defining technology. Wireless technology frees laptops to be mobile, granting an application from a particular platform at a particular location. In essence, portability and versatility designate access to knowledge and the ability to communicate from anywhere and anytime [12].

In order to create an application that will be valued by users, attention to minor details is very crucial. To help developers, MATLAB published many Apps with Graphical User Interfaces guides and tutorials, which helps to create a GUI interactively and programmatically. It is originated to reduce most of the programming hassle and to provide a friendly and regular access to the development of standalone MATLAB programs.

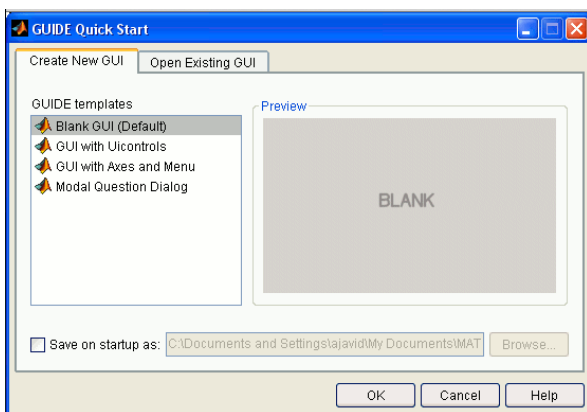


Figure 4.4: Creating a new Matlab GUI

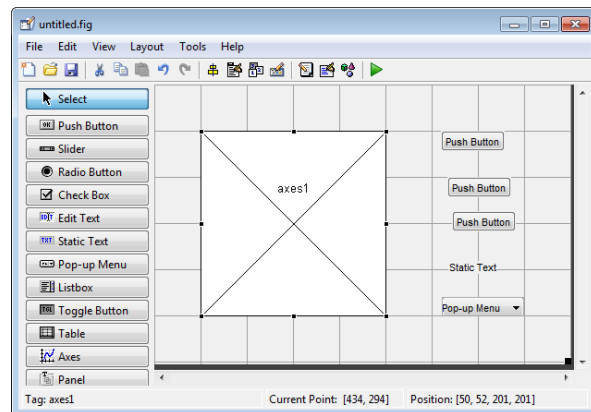


Figure 4.5: Creating a GUI with GUIDE

MATLAB GUI can be built either by using GUIDE (GUI Development Environment) or Coding from MATLAB editor. Figure 4.4 reveals the approach of creating GUI is straightforward. Once completed several examples of creating GUI, anyone can learn how to code from MATLAB editor.

Figure 4.5 illustrates that MATLAB provide a wide selection of Toolboxes including comprehensive collections of predefined functions for solving application [22]. Some basic GUI components are Push Button, Slider, and Radio Button. It is important to achieve the development of an interactive computer-based GUI for MATLAB that are applicable in several Pentium III graded PC.

4.2.2 Product Wireframes

After identifying the application process and platform constraints, wireframes were created for the primary screens. Wireframe is a visual guide that represents the skeletal framework. It is created for the purpose of arranging elements to best accomplish before actual application development. It involves planning the layout of the application, making it easier to visualize features to be accessible. Wireframes is beneficial to development

phases, as they are broken down into individual main screens, which helps to separate the development into distinct steps.

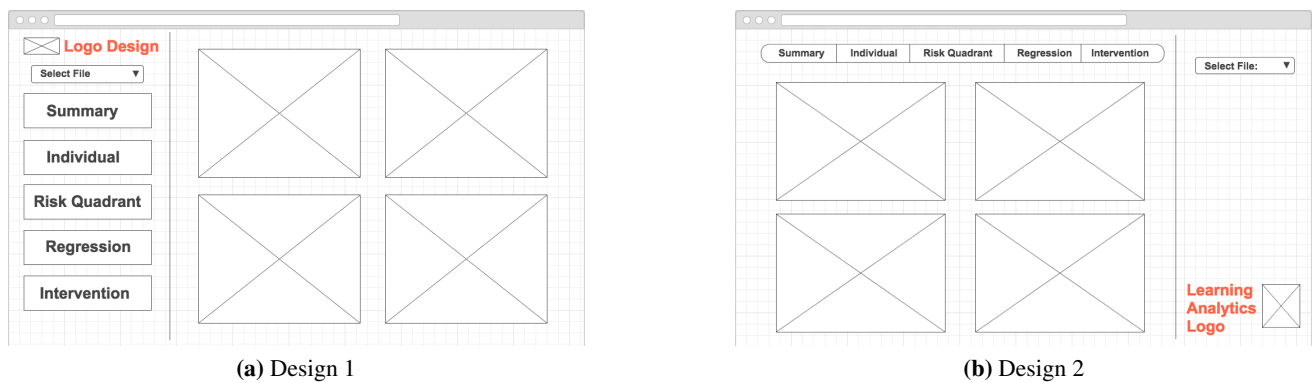


Figure 4.6: Application's Wireframes

Wireframes is an information design that involves communicating ideas to the user. It allows users to familiarize the interface through consistency throughout the whole application [14]. In order to gain feedback from potential users, a high-fidelity approach was taken that matches the original design of the application closely. Despite using this approach, the wireframes were purposely made to look like rough sketches, to encourage users to give honest feedback.

Design One (Main Screen)

This design focus on mainly accessibility, it aims to reduce the number of steps required to perform certain task. Figure 4.6a shows the main screen presented to the user when the application was launched. The screen was divided into two panels: a navigation panel on the left on the screen which provides access to other screens, and a large detailed panel on the right which shows all plots and details relating to the selected items in the navigation panel. The navigation panel includes five different pages, which includes summary, individual, risk quadrant, regression and intervention. In addition, it also includes a logo design and a dropdown list to select specific files to view. The size of the display is auto adjustable to different display screen to fit all kinds of computers.

Design Two (Main Screen)

In the second design Figure 4.6b, menus and other controls are kept smaller. This is to allow users to focus on the details and plots. The menu bar was changed to a tab located at the top of the display to minimize wastage of space. Each plot is represented by a larger cell to allow more space for title and other descriptions for better visualizing. As the tab within the display is located at the top of the display, users are able to comfortably switch tabs to enjoy each visualization process. The size of the display is auto adjustable to different display screen to fit all kinds of computers.

4.2.3 Data Representation

From the sketches of wireframes illustrated on Figure 4.6a, the author designed two set of detailed wireframe for the individual and regression tab. This is because these two tabs would require more information than the other tabs.

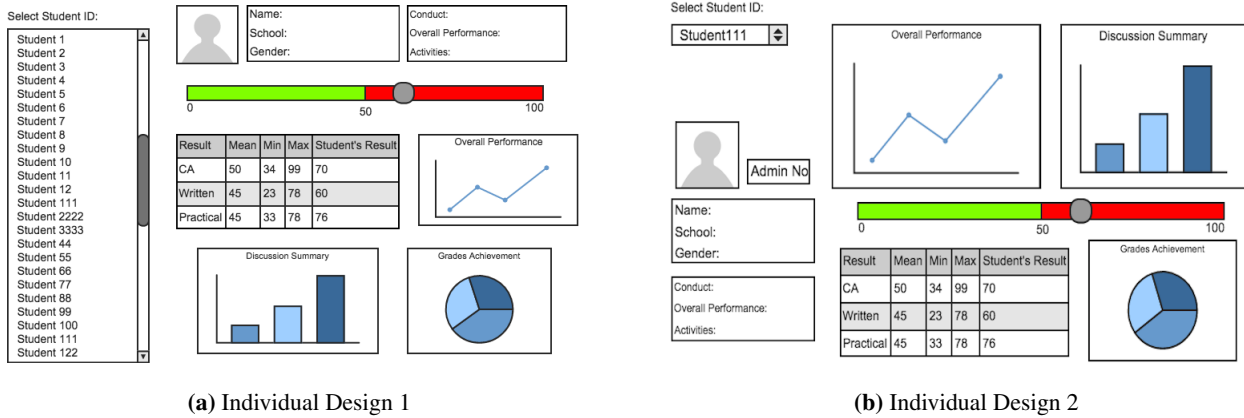


Figure 4.7: Individual Tab's Data Representation

Figure 4.7 shows two different designs for the individual tab. The list of student ID was presented as a list box and a drop down list respectively. These kinds of controls are common to organize large data sets. Student summary such as profile photo, name, id, gender and more are presented as well. Total marks are divided into continuous assessment (CA), written and practical test, and each of these tests is based on certain percentage. In addition, test marks are presented as a pie chart and table to improve the illustration purpose. Bar charts are implemented instead of a histogram, as the discussion summary's data are on a smaller scale. Thus, using a bar chart is adequate for this plot. Likewise, Pie charts are excellent for displaying data for around six categories or fewer.

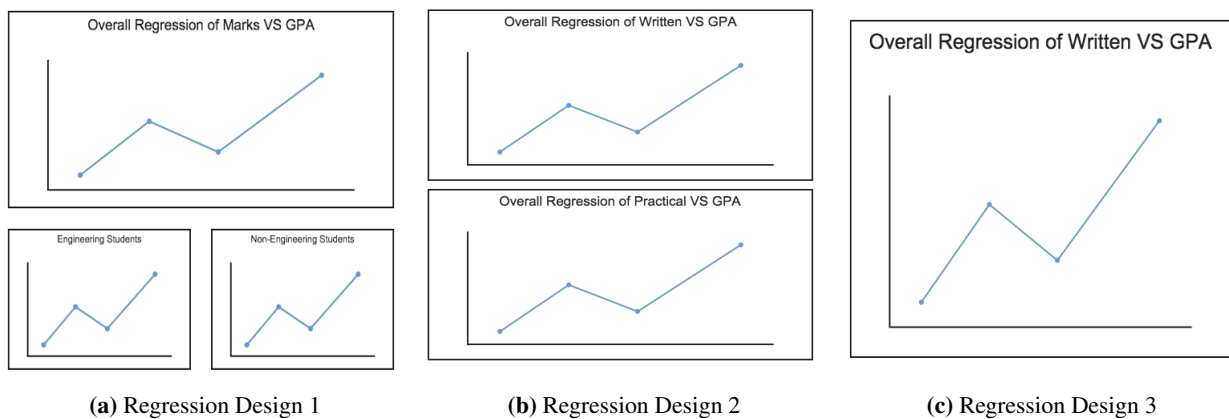


Figure 4.8: Regression Data Representation Tab

In addition, regression tabs are very important as it serve to illustrate meaning predictions and lines. It should be simple and easy to understand. From Figure 4.8a, the predictions are displayed in an overall view, followed by two different sub plots. Figure 4.8b displayed the plots are divided into two different views for different prediction purpose. Lastly, Figure 4.8c shows only a large plot to view and plot all the information into the space.

4.2.4 Product Logo Design

Logo is to aid and promote instant public recognition, to act as an identification for the application [8]. Two logo were designed to present to the users for their feedbacks. Figure 4.9a was designed to be soothing to the eyes with the colors used, while Figure 4.9b was a plain and simple logo with no images inserted.



(a) Logo Design 1



(b) Logo Design 2

Figure 4.9: Application Logo Designs

4.2.5 User Feedback

The design was introduced and explained to six potential participants to gain feedback, allowing a final design to be constructed based on the user feedbacks. The experiment was conducted in a quiet and friendly environment to enhance the effectiveness of the results.

Wireframe Design One

Participants like the idea of large selection panel for navigation purposing. However, many stated that this section would take up many redundant spaces. Participants also commented on the amount of wasted space on the application. Some participants from the technical sector mention that having large buttons to act as a navigation bar would be a redundant process. If the application were to be updated at a certain point of time, participants would have to edit the GUI layout manually to add more or delete the buttons. They noted that the image icon that represents each axes could be larger to add more information for each plot.

Wireframe Design Two

Participants preferred the tab menu bar at the top of the application, as it allows them to focus on each tab's information without being distracted by large menu bar and controls. The tab menu also helps to reduce wastage of space in the application. In addition, participants like how plots and information can be displayed in a larger space, allowing information to be spread across evenly. As compared to design one, participants commented on the amount of wasted space in the left navigation menu bar should be improved.

Data Representation

Two sets of detailed wireframes recommend to the participants for their feedback. For the individual tab, users preferred the first design better as most participants would prefer a list box in this case than a drop down box. Participants feel that during the section process upon, a drop-down box would block the information below. In addition, most participants like how the plots are depicted by each suitable exploratory data analysis plots. They feel that the plot representations are useful and appropriate for each different data sets.

In addition, participants mentioned that the three regression tabs designs are rather similar compared to one another. Most users feel that Figure 4.8a is a little unnecessary because of the two subplots representing similar ideas. This leads to a misuse of available spaces. Likewise, Figure 4.8c is not recommended as different regression lines will be confined together. This might not be feasible as it may be hard for users to visualise regression lines that are close to one another. Thus, Figure 4.8b would be the best-recommended design in this scenario.

Product Logo Design

Majority of Participants preferred Figure 4.9a than Figure 4.9b due to the refreshing colour chosen. The light-bulb icon that was inserted together with the logo represents cheerfulness and suits the project objective better. Participants feel that this logo would attract more users to use the application.

4.2.6 Application Final Design

A final design was developed, taking into account the user feedback and using concepts from the original designs. Figure 4.11 shows the final main screen design, which combines the accessibility of design one with the simplicity of design two.

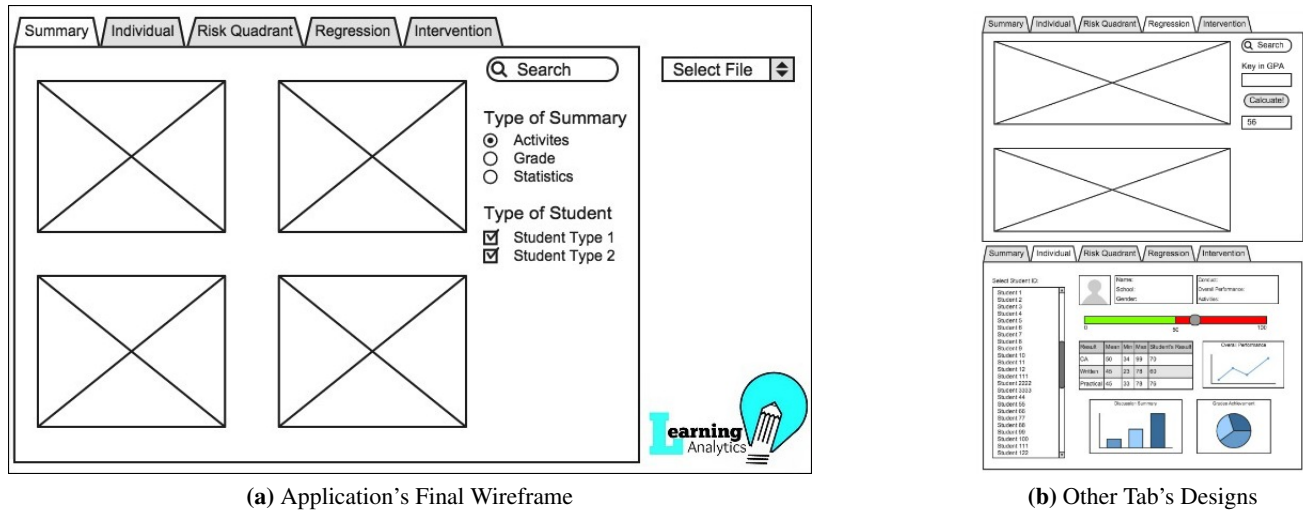


Figure 4.10: Application's Final - Menu and sizing was enlarged and more inputs added

The menu bar has seen the biggest change in the final design, addressing the wasted space issue (suggested by users). The height of all buttons on the menu have been increased to make use of all the available space. More user inputs such as radio buttons and checkboxes have been added to improve the user navigating experience. The button bar has also been removed, allowing information and plots to fill the entire screen. The logo designed was integrated into the wireframe too.

After finalizing the wireframes, work began on the visual aspects of the application. To experiment with different colours and textures, a vectorized version of main screen was designed in MATLAB GUI¹. It is beneficial to experiment with different design by applying colours and textures to individual vectors.

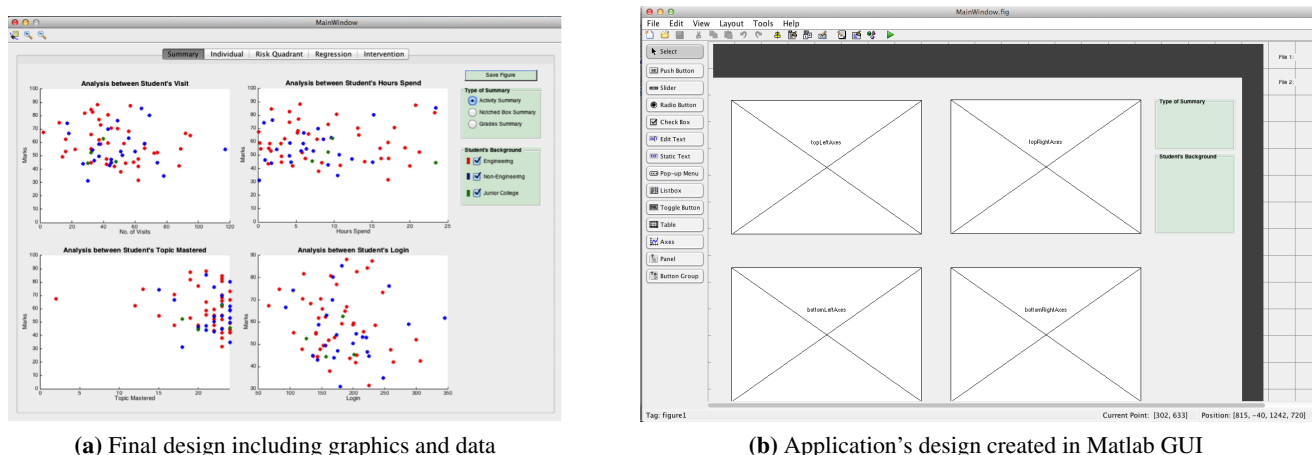


Figure 4.11: Application's Matlab

Each cell has a grey background with a slight drop shadow to add a sense of depth to the documents. Cell titles are styled with the same colour as the active tab, to link the documents grid with the bold colour palette. Cell

¹MATLAB GUI - www.mathworks.com/discovery/matlab-gui.html

titles are styled with the same colour as the active tab, to link the documents grid with the bold colour palette. Plots are distinguished by red and blue, representing a different type of students. Four plots are inserting on the same tab to ensure a proper comparison between the various student activities. Also, a standardised title font and size are inserted at the top of each plot for identification purposes.

4.3 Technologies Used

The user interface is the main component of the overall application. In order for the application to satisfy the entire requirements outlined Chapter 3: Microsoft Excel, SMS Server, SMTP Server and Applications from MATLAB are required. Each components is described below:

4.3.1 Microsoft Excel

For information and analysis to be available upon application launches, they need to be stored on a reliable platform. The actual student data will be stored directly into the Excel Spreadsheet, and to be read in the application. Such data consist of student ID, name, discussion, activities, grades, and more (Refer to Appendix B).

4.3.2 SMS Server

Figure 4.12a demonstrates Full-featured software gateway to mobile networks enabling users to broadcast and accept SMS notification, send bulk text messages and use SyncML Push technology. By using an internal protocol, it links text messages to send SMS alerts and notifications anywhere. Advanced features involve user authentication, gateway load balancing and maintenance for both locally connected and remote HTTP gateways [6]. The server can be reached via HTTP, producing unlimited possibilities for web applications.

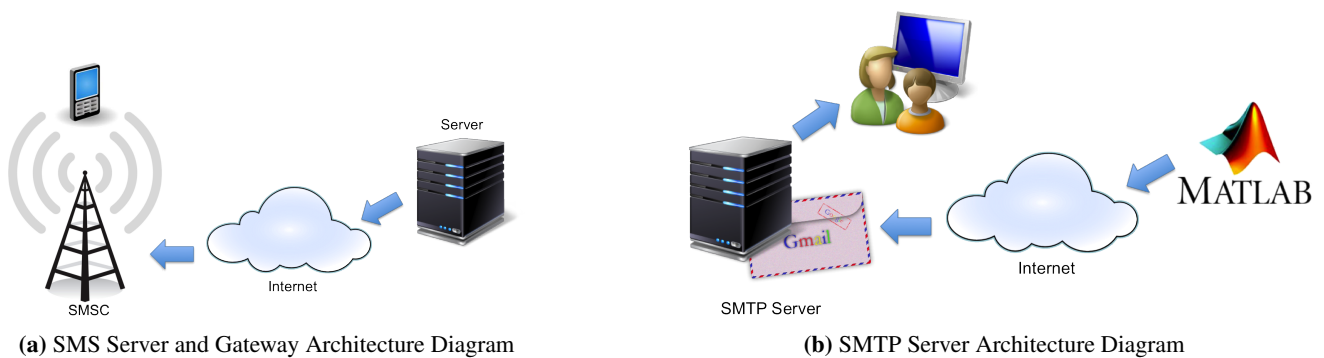


Figure 4.12: Architecture Diagram

4.3.3 SMTP Server

Figure 4.12b illustrates Simple Mail Transfer Protocol (SMTP) is an Internet standard for electronic mail (e-mail) transmission [25]. SMTP is individualistic transmission subsystem and expects only a reliable ordered data stream channel. The component should allow users to send the email to specific student’s email address during the intervention process. It should be able to sustain all different type of webmail systems such as Hotmail, Gmail and Yahoo! Mail.

4.3.4 MATLAB Applications

MATLAB Applications are interactive applications written to implement technical computing tasks. Curve Fitting Toolbox² are example of the various applications that are advantageous to this project. The toolbox allows experimenting with different exploratory data analysis, pre-process and post-process data, compare candidate models, and remove outliers. In addition, it can manage regression analysis using the library of linear and non-linear models given or specify own custom equations. The library offers optimized solver parameters and starting conditions to enhance the quality of individual fits.

²Curve Fitting Toolbox - <http://www.mathworks.com/products/curvefitting/>

Chapter 5

Implementation

This chapter describes two implementation cycles and a user evaluation that took place after each cycle. The initial phase included implementing primary and secondary requirements partially as discussed in Chapter 4. A user evaluation was performed after the first development cycle to gain feedback and further improve the functional requirements. The final implementation phase based on the user feedback and fulfilled all the remaining of secondary requirements.

5.1 Initial Implementation

5.1.1 Creating Graphical User Interfaces in MATLAB

Graphical User Interfaces in MATLAB provide point-and-click control of software applications, reducing the demand to learn a language or type commands in order to run the application. As Graphical User Interfaces is created in MATLAB, the following initialization code is auto-generated by MATLAB.

```
1 function varargout = MainWindow(varargin)
2 % Begin initialization code
3 gui_Singleton = 1;
4 gui_State = struct('gui_Name',       mfilename, ...
5                   'gui_Singleton',  gui_Singleton, ...
6                   'gui_OpeningFcn', @MainWindow_OpeningFcn, ...
7                   'gui_OutputFcn',  @MainWindow_OutputFcn, ...
8                   'gui_LayoutFcn',  [], ...
9                   'gui_Callback',    []);
10 if nargin && ischar(varargin{1})
11     gui_State.gui_Callback = str2func(varargin{1});
12 end
13 % End initialization code
```

Listing 5.1: Begin initialization code

5.1.2 Home Screen

Implementation began with the creation of the home screen. In order to allow the users to operate the application, tab menu should be implemented. This is achieved by creating Uitable that can be tutorials can be found under Matlab Documentation. Uitable are tabbed panels for arranging UI components together. The code below illustrates how to build a Uitable in a figure and sets any required properties before displaying it. A total of five tabs: Summary, Individual, Risk Quadrant, Regression and Intervention has been outlined to build up the tab menu on the application. Each tabs will be further discussed in the sections below.

```

1 tabgrp = uitabgroup(handles.mainWindowPanel);
2 tab1 = uitab(tabgrp, 'Title', 'Summary');
3 tab2 = uitab(tabgrp, 'Title', 'Individual');
4 tab3 = uitab(tabgrp, 'Title', 'Risk Quadrant');
5 tab4 = uitab(tabgrp, 'Title', 'Regression');
6 tab5 = uitab(tabgrp, 'Title', 'Intervention');

```

Listing 5.2: Create a uitabgroup and five uitabs in GUI.

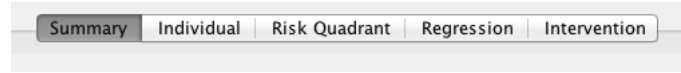
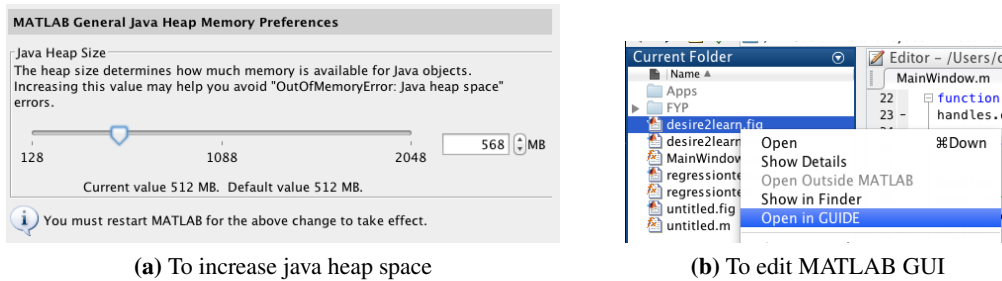


Figure 5.1: Application main screen with custom tab menu bar

Figure 5.1 illustrates the Summary tab selected upon the presentation in MATLAB GUI. After the tabs have been implemented, the author then imports the student data in (.xlsx) format to deploy the application further. The Java heap space is shared with MATLAB user interface components, such as figures, the Desktop, and the Editor. In order to prevent java heap space error is to change the maximum value by creating a java.opts file in the startup directory to set this value with a JVM flag as shown on Figure 5.2a.



(a) To increase java heap space

(b) To edit MATLAB GUI

Figure 5.2: MATLAB Custom Views

By default, GUIDE maintains two files on the first time-saving or running program. First, A FIG-file, with extension .fig, that holds a complete description of the UI layout and each UI component. The FIG-file is a binary file and modifiable without changing the design in GUIDE. Next, A code file .m which initially contains initialization code and templates for some callbacks that control UI behavior. It joins callbacks for the UI components to this file. As the Callbacks are functions, the UI code file will never be an MATLAB script. Figure 5.2b demonstrates the navigations to edit the MATLAB GUI components on the custom views by opening in GUIDE. The GUIDE will assist in adding more UI components such as axes, text labels, and tables into the application design.

5.1.3 Summary Tab

The Summary tab was implemented as a primary platform to view all the student data read from a .xlsx file. It is a platform to understand all student activities verse grades and the breakdown of student's grades. From the given data, students were grouped into two different backgrounds from their past diploma experiences: Engineering and non-engineering. Non-Engineering background consisted a broad range of exciting full-time diplomas courses to suit diverse interests from business, humanities and media. Listing 5.3 illustrates how students are categories in MATLAB.

```

1 nonEngine = handles.raw;
2 Engine = handles.raw;
3
4 nonEngine(cellfun(@(C) isnumeric(C) && isnan(C), nonEngine(:,3) ), :) = [];
5 Engine(cellfun(@(C) isnumeric(C) && isnan(C), Engine(:,4) ), :) = [];
6
7 handles.NEData = cell2mat(nonEngine(:,7:18));
8 handles.EData = cell2mat(Engine(:,7:18));

```

Listing 5.3: Categorising of Engineering and Non-Engineering Students.

Listing 5.4 emphasises the fundamental codes that were used to plot the data analysis on Summary tab. Data are categories into two major sections, engineering and non-engineering to represent the student type. Two scatter plots was achieved to identify the student with different colour representations. Notched boxes were executed to view the attributes such as mean, median and standard deviations of each student types. Furthermore, pie charts are to calculate the percentage of each unique grade across different student types. Figure 5.3 shows the plots created by the MATLAB codes.

```

1 % Scatter Plot
2 scatter(handles.topLeftAxes, handles.EData(:,2), handles.EData(:,9), 'MarkerEdgeColor', 'r', '
   MarkerFaceColor', 'r');
3 scatter(handles.topLeftAxes, handles.NEData(:,2), handles.NEData(:,9), 'MarkerEdgeColor', 'b', '
   MarkerFaceColor', 'b');
4
5 % Notched Box Plot
6 boxplot(handles.topLeftAxes, x1, 'notch', 'on', 'labels', {'Engineering Students'})
7
8 % Pie Chart to represent Grades
9 letter = ordinal(handles.EData(:,9), {'F', 'E', 'D', 'C', 'C+', 'B-', 'B', 'B+', 'A-', 'A', 'A+'
   }, [], [0, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 100]);
10 pie(handles.topLeftAxes, letter);

```

Listing 5.4: Primary code in used MATLAB.

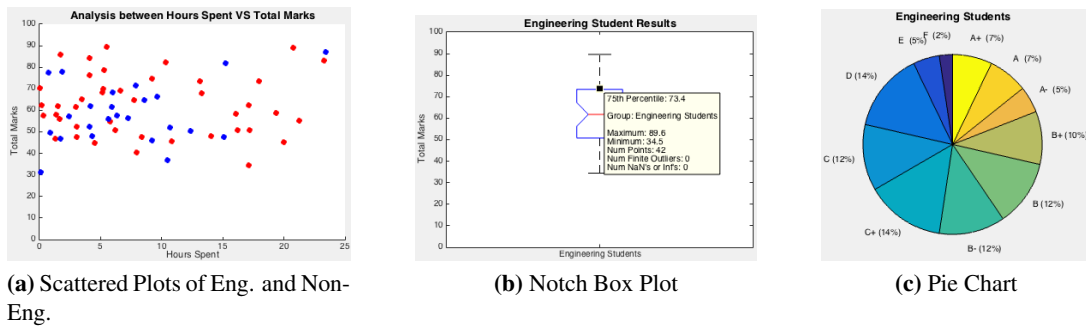


Figure 5.3: Application Summary's Plot Screenshots

5.1.4 Individual Tab

Individual tab is to represent each data from the excel sheet. Plots and Analysis shown in this panel will base on student individual data. These data describe student activities like hours spent; topics mastered, number of visits and Polytechnic GPA, which is implemented the same methods as described in the previous section. Furthermore, calculations were deployed to calculate each student's grade indicator and class position among the rest of their peers. The grade indicator was ranked into four categories: weak, average, good and excellent. The pseudo code below will illustrate the method used to deploy the grade indicator in MATLAB.

```

1 % Define percentile to identify students
2 weakPercentile = prctile(25)
3 averagePercentile = prctile(50)
4 goodPercentile = prctile(75)
5
6 %To identify Weak Students
7 if studentScore > weakPercentile
8     set studentCategory to 'Weak'
9
10 %To identify Average Students
11 if (studentScore < weakPercentile) and (studentScore > averagePercentile)
12     set studentCategory to 'Average'
13
14 %To identify Good Students
15 if (studentScore < averagePercentile) and (studentScore > goodPercentile)
16     set studentCategory to 'Good'
17
18 %To identify Excellent Students

```

```

19 if studentsScore < goodPercentile
20     set studentCategory to 'Excellent'

```

Listing 5.5: Pseudo code to categorize student grade indicator.

In order to calculate the class position, the application code group the student ID with their respective marks together to identify each student. These values will be stored as an array. Then, the marks in the array will be sorted in descending order. After which, `length()` will be used to get the position of each student and displayed in the application. Figure 5.4 illustrates how the explanations would look like in the application.

Overall Achievement	
Total Marks (%):	45.6
Grade Indicator:	Weak Class Position: 60 / 66

Figure 5.4: Application ranking Student’s Position and Grade based on Total Marks

5.1.5 Risk Quadrant Tab

The purpose of Risk Quadrant is to group students under four different categories: on track, under-engagement, dropout risk and academic risk. Each plot depicts a student in the course. As shown in Figure 5.5a, the top right quadrant comprises all students who are on-track and not at-risk. The bottom right quadrant includes students who are academically at risk, signifying that they are projected to score a D or F on the module. The bottom left quadrant receives students who expected to withdraw or Dropout. Lastly, the top left quadrant contains students who are under-engaged, implying that the students are predicted to succeed in the course, but their pattern of under-engagement might be a cause of another reason.

The MATLAB code below shows how the axis was plotted to divide each quadrant into four equal section. The line was define as the mean value of both X and Y Axis. The X axis was plotted as student’s hours spent as the success index, which Y Axis was represented by student’s total marks. Figure 5.7 illustrates the end product of Risk Quadrant page in the initial implementation.

```

1 hline = reffline([0 averageMarks]);
2 hline.Color = 'k';
3 line ([averageIndex, averageIndex], ylim, 'Color', 'k')

```

Listing 5.6: Drawing Mean Lines for X and Y Axis.

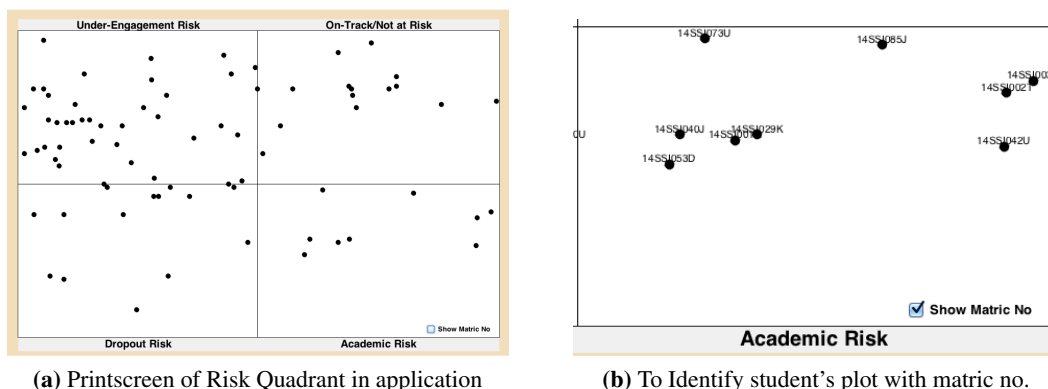


Figure 5.5: Application Risk Quadrant Screenshots

5.1.6 Intervention Tab - Send Email

Early intervention enables users to send an email to targeted students using a sendmail function. It can attach files to an email, which lets users mail files directly from MATLAB. Setting up an email address and SMTP server information with the setpref function enables users to use this feature. The code function is illustrated below:

```
1 function emailButton_Callback(hObject, eventdata, handles)
2 sender = 'dummyaddress@gmail.com';
3 psswd = 'password_of_dummy_address';
4
5 setpref('Internet','E_mail',sender);
6 setpref('Internet','SMTP_Server','smtp.gmail.com');
7 setpref('Internet','SMTP_Username',sender);
8 setpref('Internet','SMTP_Password',psswd);
9
10 props = java.lang.System.getProperties;
11 props.setProperty('mail.smtp.auth','true');
12 props.setProperty('mail.smtp.socketFactory.class',...
13                 'javax.net.ssl.SSLSocketFactory');
14 props.setProperty('mail.smtp.socketFactory.port','465');
15
16 sendmail(recipient, subject, message);
```

Listing 5.7: Sending email using a gmail account in matlab.

Furthermore, a message box is implemented to prompt users when sending email is completed and successful.

```
1 uiwait(msgbox('Successfully send email to students! Operation Completed','Success','modal'));
```

Listing 5.8: Prompted with Message box when successful.

5.2 User Feedback and Refinement

5.2.1 Evaluation Design

After implementing primary and secondary requirements partially during the initial development phase, a small-scale evaluation was performed. This assessment was conducted out to evaluate the interface and performance during the first approach to further prioritize the unimplemented requirements (outlined in Chapter 4). The 'think aloud' evaluation comprised of two parts:

1. Participants first introduced to the application by viewing the application in the author's personal laptop. It illustrates all the functionality currently implemented in the application.
2. Users then expressed their thoughts and feelings, which captured by the author and used to ranked additional feature requests.

5.2.2 Results Gathered

Five users took part in the 'think aloud' evaluation: three university students and two educators currently working in the educational sector. The application was presented on the author's Personal laptop (MacBook Pro).

User Friendliness

Most users find the system was easy to navigate, and the tab menu was rather pleasing in the GUI. However, users

mentioned that font sizes are not constant and tiny. Colours of each panel tabs and tables should be consistent and uniform throughout the whole application.

Usability issues

The waiting time for each tab to load was rather slow, and most users feel that the waiting time should be improved.

Risk Quadrant Definition

Users suggested that the Risk Quadrant definition seems to be incorrect as success index should not be based on one contributing factor. It should contain at least 2 student factors to be determined as a student success index. Also, the plots are a little hash to the eyes as the colours are too dull. The plots should be classified in various colours to fit different category.

Intervention (Email Notification)

Users suggested that the email notifications should be customizable according to student's grades. For instance, a weak student should be receiving an encouraging email and a excellent student should be receiving a email to congratulate on his achievements. Thus, the emails should be generated automatically based on the individual's achievement.

Overall Look and Feel

Overall, users feel that the application was useful and analytics are swiftly implemented. However, early interventions should include sending SMS to student's mobile devices to enhance the interaction between educator and learners. Furthermore, the system should have a reliable and uniform look to standardized the whole application.

5.2.3 Additional features to implement

Based on the results from the evaluation, the following features were prioritised for the second development cycle:

- Grade Calculator (Regression Tab)
- Linear Regression for predicting student results
- SMS Service for early intervention
- Group students category into specific colours

5.3 Second Implementation Cycle

The second implementation phase aimed to address the defects identified during the initial evaluation, and implement the features prioritized by users during the 'think aloud' evaluation. Regression tab including a grade calculator to estimate student future scores are also implemented in the Second Implementation cycle.

5.3.1 Addressing identified defects

Risk Quadrant

The success index should not be based on one contribution factor only. Thus, the author defined that student background, topics mastered, hours spend and number of visits to the student portal should be calculated into the success index. Furthermore, it will be useful if different quadrant can be classified into different colour groups. The calculation will be shown in the later part of the section.

Cosmetics Amendments

The application tabs should be uniform and standardized across the whole program. Thus, the application's GUI GUIDE should be changed accordingly. Also, a SMS server should be deployed in the early intervention tab as suggested during the 'think aloud' evaluation.

5.3.2 Linear Regression

To achieve the Regression tab's predictions, Curve Fitting Toolbox¹ is required. Curve fitting is a process of fitting a function to a set of data points, also known as regression analysis. Curve Fitting functions from MATLAB provided implements functions for fitting curves and surfaces to data. The function allows conducting exploratory data analysis, pre-process and post-process data, analyze candidate models, and remove outliers. It also allows conducting regression analysis applying the library of linear and nonlinear patterns provided or specify own custom equations. The library produces optimized solver parameters and starting conditions to improve the quality of each fit.

```

1 polyfit(x, y, 1);           % Linear Regression
2
3 polyfit(log(x), log(y), 1); % Power Function
4
5 polyfit(x, log(y), 1);     % Exponential Function
6 polyfit(x, log10(y), 1);
7
8 polyfit(log(x), y, 1);     % Logarithmic Function
9 polyfit(log10(x), y, 1);
10
11 polyfit(x, 1./y, 1);      % Reciprocal Function

```

Listing 5.9: Curve Fitting MATLAB polyfit functions

The norm of the residuals (R_2) is a model of the deviation within the correspondence fit and the data. A lower norm determine a better fit [4]. To compute R_2 , first compute a fit and then obtain the residuals from is. A residual is the approved difference between an observed dependent value and the value fit predicts for it. Listing 5.10 illustrates the codes required in MATLAB to calculate the norm of the residuals.

$$\text{Norm of Residuals}(R_2) = y_{\text{observed}} - y_{\text{fitted}}$$

```

1 %To calculate Linear's norm of residuals
2 p = polyfit(x, y, 1);
3 z = polyval(p, x);
4 resid = y - z;
5 norm(resid, 2)

```

Listing 5.10: Functions to calculate norm of residuals in MATLAB.

A comparison of coefficient and norm of residuals shown in Table 5.1 to prove the correctness when its typical size is sufficiency to ease any worries that conclusions might be incorrect. The smaller the value of residual, it proves the better the curve fit it is.

Curve Fitting Function	Coefficient	Norm of Residuals
Linear	p1 = 3.9938, p2 = 48.587	112.8746
Power	p1 = 0.2033, p2 = 3.8541	472.8884
Exponential	p1 = 0.0790, p2 = 3.8385	476.0660
Logarithmic	p1 = 9.6256, p2 = 50.1148	191.8222
Reciprocal	p1 = -0.0017, p2 = 0.0225	508.2212

Table 5.1: Different regression's coefficient and norm of residuals comparison

¹Curve Fitting Toolbox - <http://www.mathworks.com/products/curvefitting/>

As recorded on Table 5.1 and Figure 5.6, Linear Regression would be the most suitable procedure to accomplish the goal for the Regression tab. The least value as shown on norm of residuals and best fitting line as shown with the scattered plots proves that it is the best suggestion to use in this application.

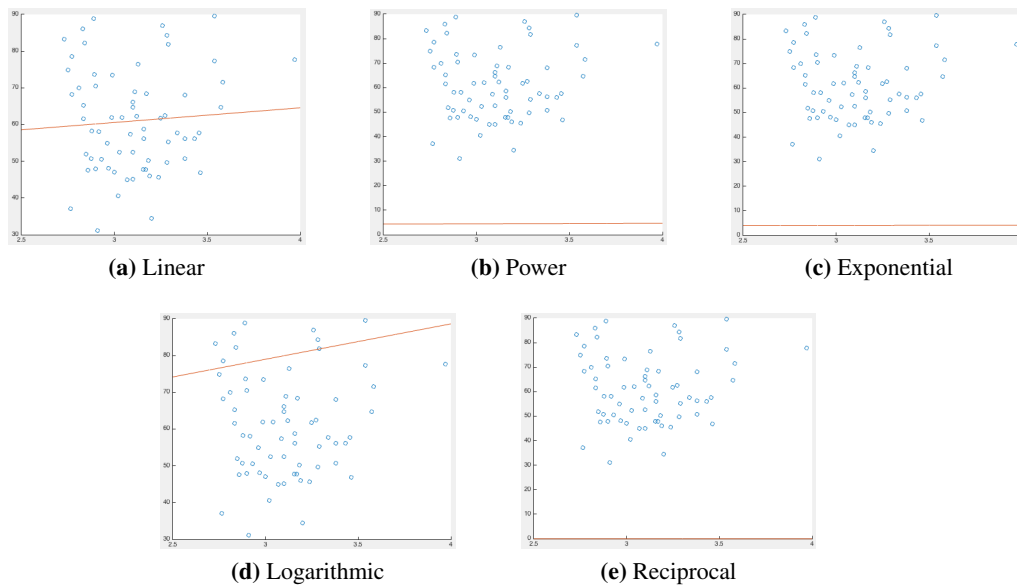


Figure 5.6: Graph Comparison between student grades and best fit line

```

1  y = mx + c           % Linear Regression
2
3  ln(y) = m ln(x) + ln(c)   % Power
4
5  ln(y) = mx + ln(c)       % Exponential
6  log(y) = mx + log(c)
7
8  y = m ln(x) + c         % Logarithmic
9  y = m log(x) + c
10
11 (1/y) = mx + c         % Reciprocal

```

Listing 5.11: Curve Fitting MATLAB functional form.

Next, a calculator was implemented to calculate the estimated both expected total marks and written and practical marks based on student’s GPA entered. These marks’ equations are coded on the Listing 5.11. The simplified codes of the calculator without any validations are presented on Listing 5.12. The Calculator’s GUI will be displayed in Figure 5.7b. Likewise, the screenshot for regression tab with linear regression based on GPA and marks is displayed on Figure 5.7b.

The grade calculator is suitable to estimate the student’s grade based on their past GPA achieved. It is beneficial during interview and student intake process, where educators can predict the student’s grades based on the student’s background that is reflected in the overall GPA.

```

1  function checkButton_Callback(hObject, eventdata, handles)
2  xE = str2double(get(handles.regMarkText, 'String'));
3  equE = pE(1,1) * xE + pE(1,2);
4  equEText = sprintf('%1.2f', equE);

```

Listing 5.12: Calculate marks based on GPA entered.

Figure 5.7a illustrates the screen shot of the final regression tab. The regression tab consists of two plots that consisted of total marks and a average of practical and written marks. Continuous assessment (CA) marks was eliminated for the second plot as the author believed that it would be unreasonable to estimate students based on team assignment and other class participations. Therefore, the second plot is the average written and practical marks that students scored based on self-effort.

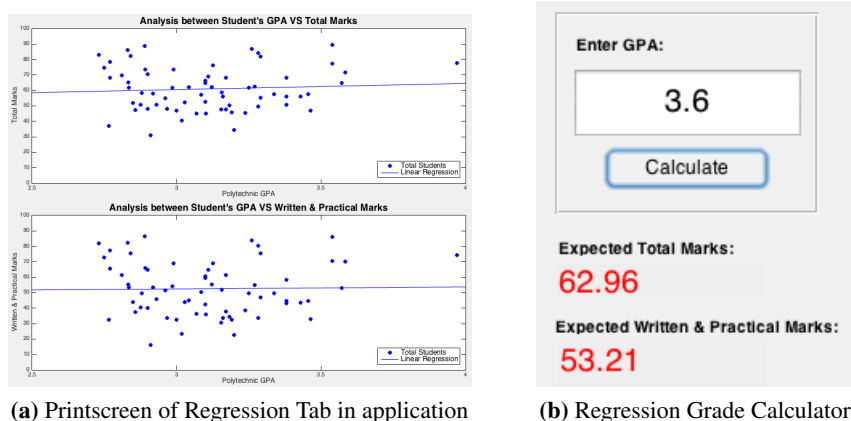


Figure 5.7: Application Regression tab screenshots

5.3.3 Risk Quadrant Tab - Success Index

The Success Index was revised to base on four contributing factors: student's background, hours spent, number of visits and topics mastered on that module [23]. The success index appears to be a more accurate one, as it helps to minimize the number of students at risk as compared to the first initial implementation. The risk quadrant was well spread in a way that most students end up at the 'on track' quadrant, which is a good representation for the analysis.

The following formula was identified:

$$\text{Success Index (\%)} = \left(\frac{GPA}{4} \times 25\%\right) + \left(\frac{hours(h)}{max(hours)} \times 25\%\right) + \left(\frac{visits}{max(visits)} \times 25\%\right) + \left(\frac{topics}{max(topics)} \times 25\%\right)$$

where,

GPA representing student's background,

the number 4 is the maximum GPA score,

and the total student activities added up to 100% to represent the success index.

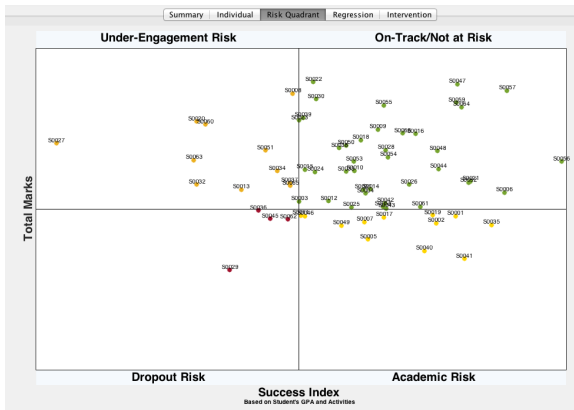
Furthermore, the simplified code below illustrates a how is the plots being classified into four different colour: green, yellow, orange and red. These colour are arranged based on student on track to student at risk respectively to the colour spectrum. Figure 5.8 shows the end product of the risk quadrant.

```

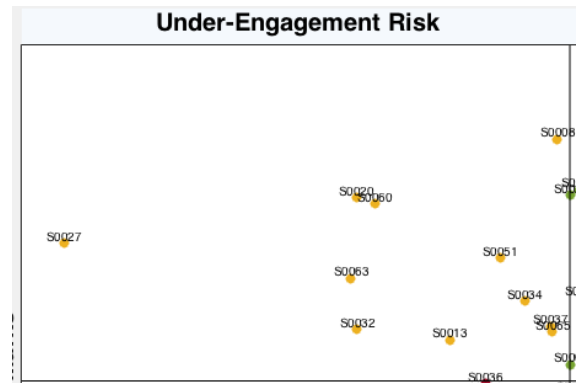
1 A = logical((meanOverallmarks < averageOverallmarks) & (meanIndex < averageIndex));
2 B = logical((meanOverallmarks > averageOverallmarks) & (meanIndex < averageIndex));
3 C = logical((meanOverallmarks < averageOverallmarks) & (meanIndex > averageIndex));
4 D = logical((meanOverallmarks > averageOverallmarks) & (meanIndex > averageIndex));
5
6 %Plotting
7 scatterplot = gscatter(meanIndex, meanOverallmarks, ca(:,1));
8 set (scatterplot(1), 'MarkerFaceColor', [0.635 0.078 0.184], 'color', [0.635 0.078 0.184], 'MarkerSize'
9     ,28)
...

```

Listing 5.13: Changing colour format in Risk Quadrant



(a) Risk Quadrant with coloured plots



(b) Risk Quadrant's plot identified by matrix no.

Figure 5.8: Application's Risk Quadrant finalized screenshot

5.3.4 Intervention Tab - SMS Notification

As suggested by users earlier, the capability to send a SMS notification to respective students would be beneficial to the application. To achieve this, a daemon service (Gammu) was deployed. Gammu is a cross-platform software for mass processing SMS messages. Gammu is running on a linux server which equipment necessitated dougle 3G and a Sim Card. After which, the dougle will be attached to the server. To ensure that the dougle has been attached, the command `lsusb` is called

Next, it is necessary to configure the Gammu setup file by the command `sudo nano /.gammurc`. This configuring file will tell you which port the dougle is connected to and the transfer rate. Following, the author switched the dougle to swap them into modern mode, ensuring Gammu is detecting the dougle `gammu identify`.

After the set up is completed, a php script is written to communicate with a linux server due to MATLAB limitations. Thus the PHP script that enables sending of gammu request is illustrated by Listing 5.14.

```

1 <?php
2 $number = $_POST[ 'number' ];
3 $message = $_POST[ 'message' ];
4
5 $command = "echo $message | gammu sendsms text +65$number ";
6 $output = shell_exec($command);
7 echo "<pre>$output $number $message</pre >";
8 ?>
```

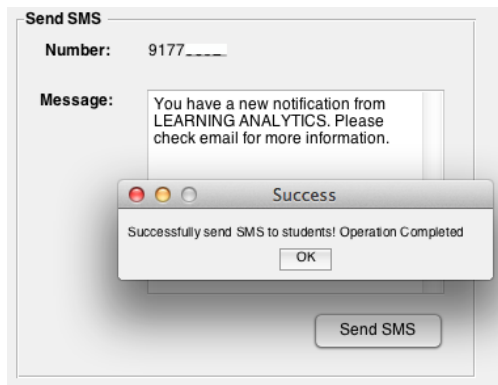
Listing 5.14: PHP Script to Communicate with server.

```

1 function smsButton_Callback(hObject, eventdata, handles)
2 number = get(handles.numberText, 'String');
3 message = get(handles.msgText, 'String');
4
5 URL = 'http://fyp-uog.me/ClarisMatlab/sendSMS.php';
6 str = urlread(URL, 'POST', { 'number', number, 'message', message });
7 uiwait(msgbox('Successfully send SMS to students! Operation Completed', 'Success', 'modal'));
```

Listing 5.15: MATLAB function to Send SMS.

Therefore, the Listing 5.14 PHP script will be called by matlab. The following Figure 5.9 illustrates the screenshots in MATLAB GUI upon sending SMS successful and a SMS notification on the mobile device respectively. A message box will prompt the users when sending of SMS is successful.



(a) Sending SMS Success



(b) View SMS on Mobile Devices

Figure 5.9: Application Early Intervention SMS Notification Screenshots

5.4 Summary

Many techniques and predictive models have been employed in the application, encountered and learnt about during the implementation of this application. During the implementation phase a large proportion of time was spent developing a best-fit regression line for prediction. Figure 5.10 illustrates the final tree diagram of the complete application.

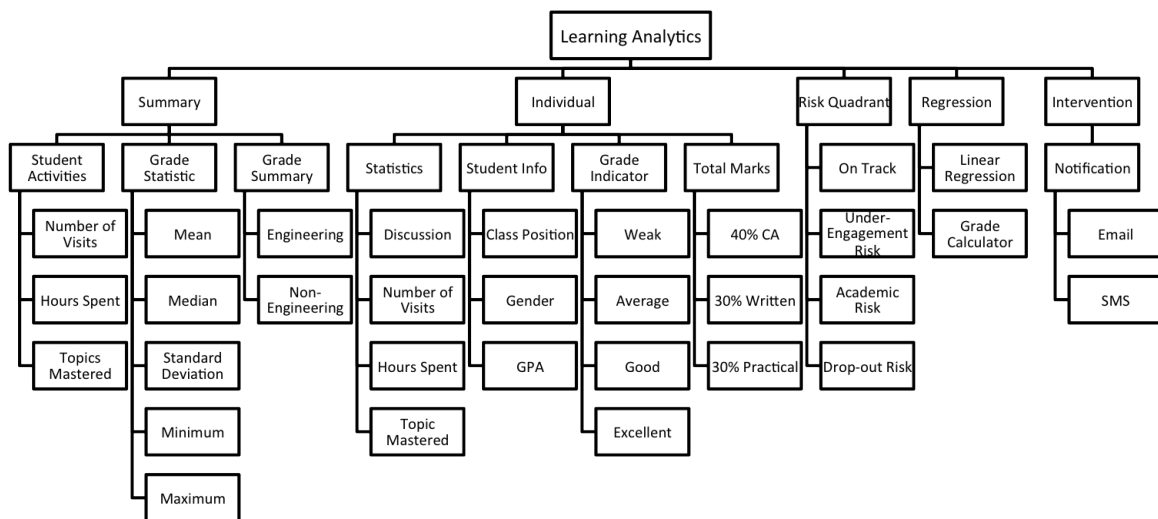


Figure 5.10: Application's Overall Diagram

Chapter 6

Evaluation

6.1 Usability Evaluations

A usability evaluation focused on an individual user's experience when interacting with the system was used to evaluate the application of the implementation cycle. Observations and results gathered will be represented at the later part of this section.

The think aloud evaluation was designed to test the usability of the applications key features in a short period of time. The evaluation aimed to expose users to both versions of the application whilst testing the presenter and viewer functionally.

Participants were then issued a task sheet containing six short tasks (task sheet is included in Appendix D.2). The set of tasks required each participant to select the excel files for analysing. The section set of task is to navigate and identify the few different major sections of the project. Then, participants go on to calculate the estimated scores. Lastly, participants would have to send an email and SMS to the respective students.

The System Usability Scale (SUS) is a simple, ten-item scale giving a global view of subjective assessments of usability. It was developed by John Brooke in 1986 as a tool to be used in usability evaluations of electronic office systems [3]. Each item in the SUS has five response options, labeled 1 to 5, which represent Strongly Disagree and Strongly Agree, respectively. Results from the SUS are combined to yield a single number, which represents a composite measure of the overall usability of the system being studied (full details of how the overall score is calculated is included in Appendix D.1) The study also showed that an SUS evaluation with a sample size of twelve, would lead to the same conclusion of a bigger sample size.

6.1.1 Think Aloud Evaluation

Feedback towards the application was overwhelmingly positive. Many users commented that the application was extremely polished and specifically praised the ability to send email and SMS notifications. A few users had some difficulties selecting a excel file due to the uninformative text. Two participants found it difficult to identify particular plots on the individual tab. This is due to too many information are presented on one screen. Based on this feedback, some of the plots were modified and represented by different colours to identify each plot better. All participants were able to successfully complete the five tasks mentioned on Appendix D.2.

Experiment Scenario

All participants walks into the experience area seated in a comfortable position. The author starts explaining on

the description of application. Then, the participants are presented with the task sheets and each experience was about five to ten minutes long.

Results Gathered

Twelve participants took part in the short-term evaluation, ten of which were student peers. The feedback gained through the ‘Think Aloud’ evaluations and online questionnaires is summarised below. Additionally, all feedback gained from the online questionnaire is included in Appendix D.3.

What data would you add to Learning Analytics?

Participants identified eight unique data to be analysed by the application. Six of the proposed features had previously been identified during the requirements capture.

- Student Grades
- Student past achievement (GPA Scoring, Background)
- Class Attendance
- Discussion forums
- Knowledges
- Individual class position
- Hours spent
- Class Participation

Multiple participants requested the ability to generate student data automatically into the application directly. These data should be automatically collected and read either from the school-learning portal or directly from the educator. As such, data such as class attendance and participations should be synced into the application everyday to view the most up to date details. Unfortunately, this was considered out of scope for the project due to the time constraints. However, auto syncing of data and all other features will be considered for future versions of the application.

What did you find difficult when using Learning Analytics?

Half of the participants stated they had no difficulties using the application. Minority participant declared that they were confused with the flow of the individual tab at time due to the grade chart of student vs the class. They state that they were unable to identify the class average, minimum and maximum score at first. However, these users also noted that the legend guide was informative and felt that it would only need to be referred to a few times.

When would you imagine using Learning Analytics?

Participants were given three environments choices where they felt the application could be used. As expected, all participants stated the application could be used in a teaching environment. Two participants stated the application would be ideal for social media analysis, as it would allow them to study user activities and statistics without the need of existing tools that might be very costly.

Would you recommend Learning Analytics?

All participants stated they would recommend the application. This statistic is extremely rewarding as it proven the success of the application.

6.1.2 SUS Evaluation

The results of the System Usability Scale are shown in Table 6.1. The application scored 88.96 out of a possible 100. This implies the application has ‘Excellent’ usability, based on the adjective rating scale. The score is illustrated on this scale in figure 6.1.

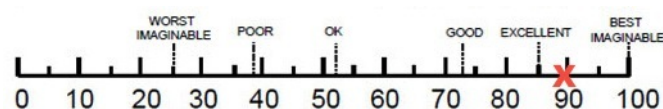


Figure 6.1: SUS Score

#	Question	Rating Avg	Score Contribution
1	I think that I would like to use this application frequently.	4.83	3.83
2	I found this application unnecessarily complex.	1.83	3.17
3	I thought this application was easy to use.	4.83	3.83
4	I think that I would need assistance to be able to use this application.	1.00	4.00
5	I found the various functions in this application were well integrated.	4.50	3.50
6	I thought there was too much inconsistency in this application.	1.17	3.83
7	I would imagine that most people would learn to use this application very quickly.	4.58	3.58
8	I found this application very cumbersome/awkward to use.	1.75	3.25
9	I felt very confident using this application.	4.08	3.08
10	I needed to learn a lot of things before I could get going with this application.	1.50	3.50
		SUS Score:	88.96

Table 6.1: The application received an SUS score of 88.96 denoting ‘Excellent usability’

6.2 Dataset Evaluations

The experiment was designed to simulate the non-functional requirements defined on Chapter 3. The objective of the evaluation is to test the robustness and the capability to support varies dataset in the application. This section evaluates the accuracy and efficiency of the application responding to various dataset factors.

6.2.1 Application Accuracy

The author testes the application with various dataset collected on the initial implementation stage of the project. The testing dataset collected has originated as the final implementation dataset, which are authentic student data from Singapore Institute of Technology (SIT). However, the testing dataset are targeting approximately 70 Sustainable Infrastructure Engineering students taking the Mathematic module.

The dataset was collected and categorized according to the application’s philosophy. After plotting the dataset into the application, dataset was able to adapt and make sense in the application positively. This experiment prove that the application was able to make sense any sets of data as long as the data is structured precisely.

6.2.2 Efficiency of Sampling Dataset

Furthermore, the author measured efficiency of the application by recording the time differences corresponding to different data size. Two sample sizes 66 and 198 have been applied for this experiment whereby 66 is the initial dataset and 198 is the multiply of three datasets ($N = 66$). Table 6.2 shows the time taken for both data sample size. The timing needed for both sample sizes did not reveal sufficient significant contrast. This timing proves the robustness of the application to support any massive data sample size without much time delays.

Time (sec)	Data Sample Size (N)
3.057	66
3.159	198

Table 6.2: Time captured with different data sample size on the application

Chapter 7

Conclusion

7.1 Summary

The aim of this project is to design and develop a toolkit that ables to analyze and report learner-related information. After cycles of implementation and usability evaluations, the author is confident that the finished product has met all the high priority requirements identified. Furthermore, the author was very pleased with the overall look and feel of the application. User feedback was extremely positive during the usability tests described in Chapter 6. Several individuals have shown keen interest in the product and are eager for more details of the application.

7.2 Future Development

This section describes some of the possible future work for developing this application further. These features were not in the scope of the original project, however; they are suitable to be incorporated in future versions of learning analytics application. Participants identified several additional features during different stages of evaluation. Some of the future features of this project are as follows.

- Database management system support
- File synchronisation across applications
- More comprehensive data for better analysing
- Automatic generate student data
- Ability to read all data file format
- Other platform support

7.3 Lessons Learnt

Throughout the project the author has become increasingly competent in designing and developing a learning analytics. The author learnt that good time management is the key to success. In addition, the author worked closely with project supervisor to ensure effectiveness and efficiency of the project. Creating a learning analytic application that provides a universal educational platform proved an exciting challenge, resulting in a deep understanding of Matlab application design. During the implementation phase, the author has become a proficient Matlab developer with a solid understanding of the varieties of plots and statistics. As the application involved a substantial amount of data predictions, a solid understanding of the fundamental concepts in regression analysis has also been achieved.

Bibliography

- [1] Agrawal, R., Imielinski, T. and Swami, A. (1993) Mining Association Rules between Sets of Items in Large Databases. *Department of Computer Science, Katholieke Universiteit Leuven, Belgium* pp. 2–3.
- [2] Ahonen, H., Heinonen, O. and Klemettinen, M. (1998) Applying Data Mining Techniques for Descriptive Phrase Extraction in Digital Document Collections. *Department of Computer Science* pp. 2–3.
- [3] Bangor, A., Kortum, P. and Miller, J. (2009) Determining what individual sus scores mean: Adding an adjective rating scale. *Journal of Usability Studie* pp. 1–4.
- [4] BETZLER, K., PHYSIK, F. and OSNABRUCK, U. (2014) FITTING IN MATLAB. *FACHBEREICH PHYSIK* pp. 3–4.
- [5] Chatti, M. A., Dyckhoff, A. L., Schroeder, U., and Thus, H. (2012) A Reference Model for Learning Analytics . *International Journal of Technology Enhanced Learning (IJTEL)* pp. 12–22.
- [6] Communications, D. (2015) SMS Server and GatewaY. [Online] Available from: <http://www.deerfield.net/products/icewarp/sms/>. [Accessed 3 March 2015].
- [7] D, V. C. S. P. (2013) Predictive modeling to forecast student outcomes and drive effective interventions in online community college courses. *Journal of Asynchronous Learning Networks, Volume 16* pp. 4–6.
- [8] dictionaries, O. (2015) Logo Definition. [Online] Available from: <http://www.oxforddictionaries.com/definition/english/logo>. [Accessed 2 March 2015].
- [9] Dietz-Uhler, B. and Hurn, J. E. (2013) Using Learning Analytics to Predict (and Improve) Student Success: A Faculty Perspective. *Journal of Interactive Online Learning* pp. 1–4.
- [10] Ding, C. S. (2006) Using Regression Mixture Analysis in Educational Research. *Practical Assessment Research and Evaluation Vol 11 No 11* pp. 5–8.
- [11] Dyckhoff, A. L., Zielke, D., Bultmann, M., Chatti, M. A. and Schroeder, U. (2012) Design and Implementation of a Learning Analytics Toolkit for Teachers. *Educational Technology and Society* pp. 1–19.
- [12] Ezziane, Z. (2007) Information Technology Literacy: Implications on Teaching and Learning . *Information Technology Literacy: Implications on Teaching and Learning. Educational Technology and Society, 10 (3)* pp. 185–186.
- [13] Filzmoser, P. (2008) Linear and nonlinear methods for regression and classification and applications in R. *Department of Statistics and Probability Theory Vienna University of Technology Wiedner Hauptstr* pp. 9–14.
- [14] Garrett, J. J. (2011) The Elements of User Experience. *The Elements of User Experience* pp. 55–60.
- [15] Inc, B. (1997-2011) Analytics for Learn DS. pp. 01–02.
- [16] Inc, B. (2014a) Blackboard Analytics Overview. [Online] Available from: <http://www.blackboard.com/Platforms/Analytics/Overview.aspx>. [Accessed 26 October 2014].

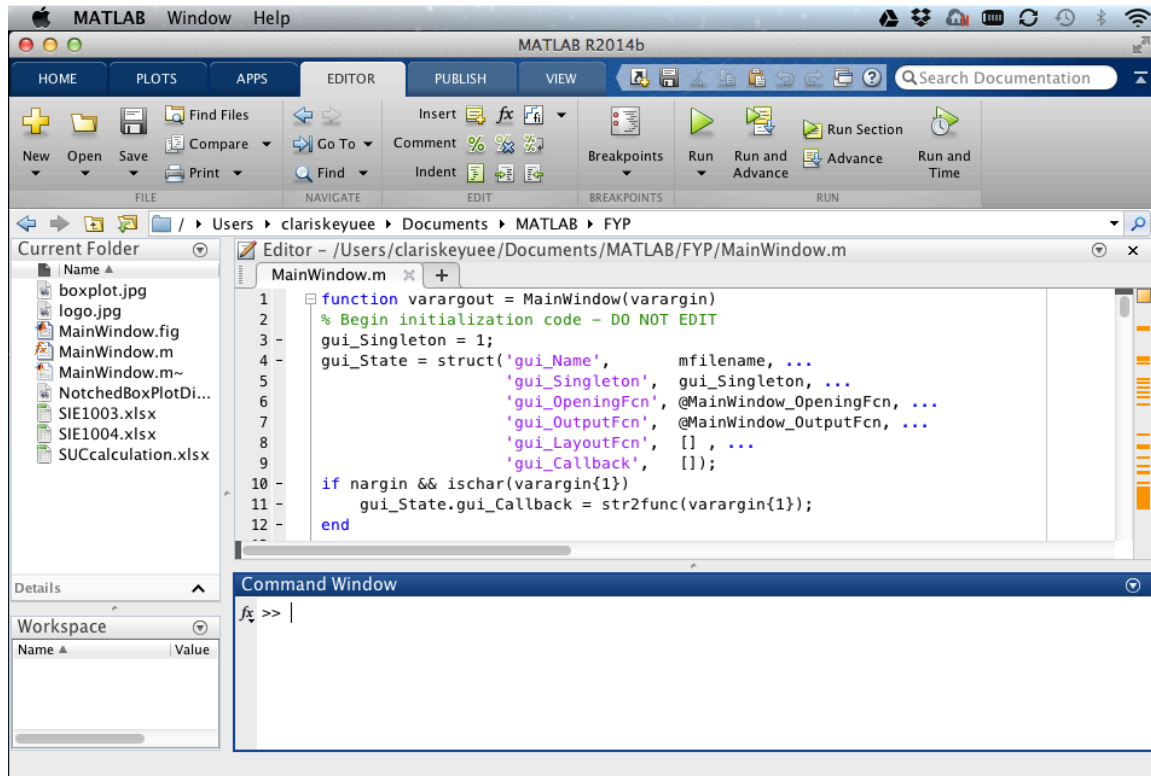
- [17] Inc, B. (2014b) Blackboard Analytics Overview. [Online] Available from: <http://www.blackboard.com/Platforms/Analytics/Products/Overview.aspx>. [Accessed 26 October 2014].
- [18] Inc, B. (2014c) Blackboard Partnership Program FAQs. [Online] Available from: <http://www.blackboard.com/Partnerships/Partnerships-Program/FAQs.aspx>. [Accessed 26 October 2014].
- [19] Krebs, V. (2013) Social Network Analysis A Brief Introduction. [Online] Available from: <http://www.orgnet.com/sna.html>. [Accessed: 03 November 2014].
- [20] Lamb, S. and Rice, S. (2008) Effective Intervention Strategies for Students at Risk of Early Leaving. *Centre for Post-compulsory Education and Lifelong Learning* pp. 1–10.
- [21] M.C.A., P. K. V. and M.Phil. (2013) Survey Of Data Mining In Socio-Academic Perspective. *International Journal of Scientific and Technology Research Volume Two* pp. 02–07.
- [22] Nasiruzzaman, A. B. M. (2010) IUsing MATLAB to develop standalone graphical user interface (GUI) software packages for educational purposes. *Matlab - Modelling, Programming and Simulations* pp. 3–20.
- [23] Nespereira, C. G., Dai, K., Vilas, A. F. and Redondo, R. P. D. (2014) Is the LMS Access Frequency a Sign of Students' Success in Face-to-Face Higher Education?. *Second International Conference on Technological Ecosystems for Enhancing Multiculturality* pp. 3–8.
- [24] Olley, A. and Holman, A. (2014) Engagement analytics. [Online] Available from: https://moodle.org/plugins/view.php?plugin=report_engagement. [Accessed: 28 October 2014].
- [25] Postel, J. B. (1982) Simple Mail Transfer Protocol. *Network Working Group* pp. 1–3.
- [26] Raedt, L. D. and Thon, I. (2010) Probabilistic Rule Learning. *Mining Association Rules between Sets of Items in Large Databases* pp. 1–3.
- [27] Seltman, H. (2014) Experimental Design and Analysis. *36-309/749 Experimental Design for Behavioural and Social Sciences* pp. 61–83.
- [28] Suess, J., Dillon, M. and Ross, Y. M. (2012) Analytics for Institutional Transformation. [Online] Available from: <http://www.educause.edu/ero/article/using-analytics-institutional-transformation>. [Accessed 17 October 2014].
- [29] Teater, D. B. (2012) Curriculum guide - Social work intervention methods. *Produced in partnership to support curriculum development in initial social work education* pp. 1–3.
- [30] TechTerms.com (2015) GUI. [Online] Available from: <http://techterms.com/definition/gui>. [Accessed 1 March 2015].
- [31] Waltenburg, E. and McLauchlan, W. (2012) Exploratory Data Analysis A Primer for Undergraduates. *Exploratory Data Analysis A Primer for Undergraduates* pp. 4–6.
- [32] Wikipedia (2012) Content analysis. [Online] Available from: http://en.wikipedia.org/wiki/Content_analysis. [Accessed 17 October 2014].

Appendices

Appendix A

User Guide

1. Run the MainWindow.m on MATLAB R2014b



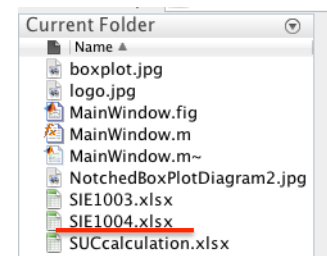
2. Change the directory name accordingly.

```

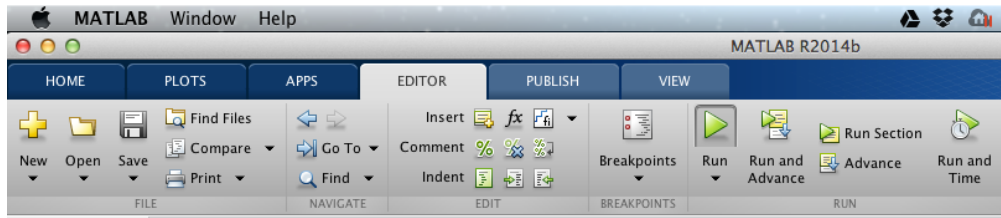
79 -   dirName =  '/Users/clariskeyuee/Documents/MATLAB/FYP' ;
80 -   files = dir( fullfile(dirName, '*.xlsx') );
81 -   handles.files = {files.name}';
82 -   set(hObject, 'string', handles.files, 'FontSize', 12);

```

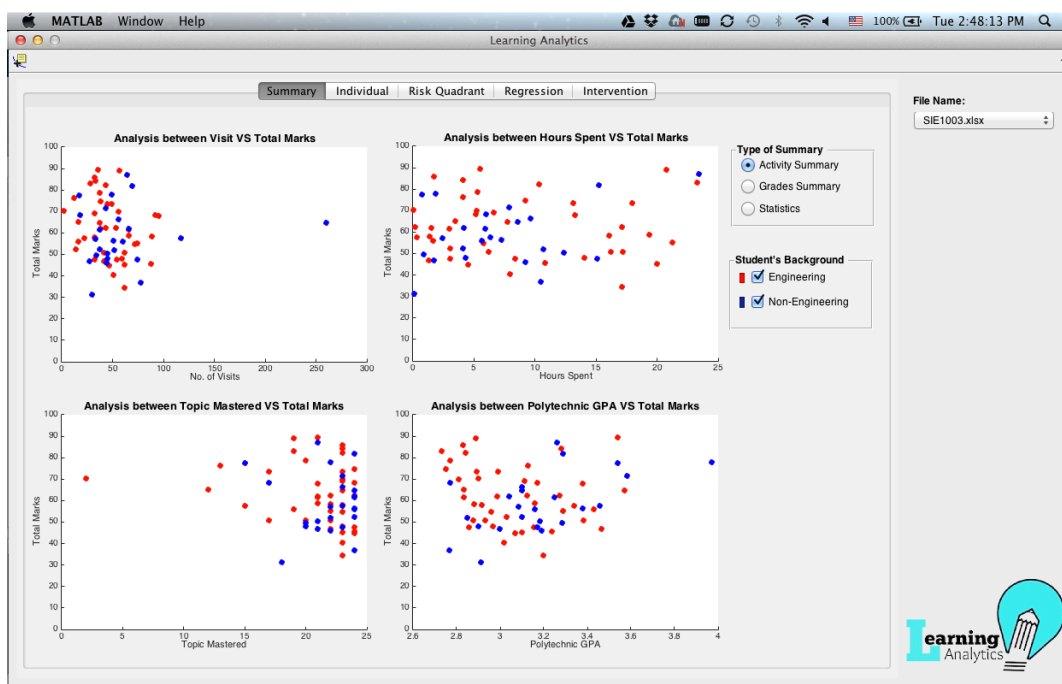
3. Add in the .xlsx file by pasting into the Current Folder.



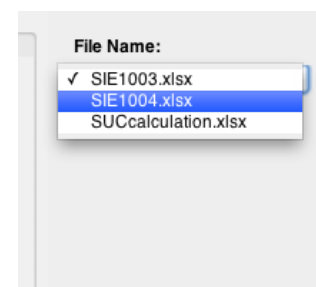
4. Click on the 'Run' button



5. Application launched from MATLAB



6. Now you are able select the file to view from the application



Appendix B

Data File Header

No.	Field
1	Student ID
2	Name
3	Non-engineering
4	Engineering
5	Email
6	Gender
7	Poly GPA
8	#Visits
9	Total Time Spent (h)
10	Topics visited (24 max)
11	Success Index
12	Post read
13	Threads created
14	Replies posted
15	Overall Total
16	Total CAs (100%)
17	Written exam (100%)
18	Practical exam (100%)
19	Handphone Number

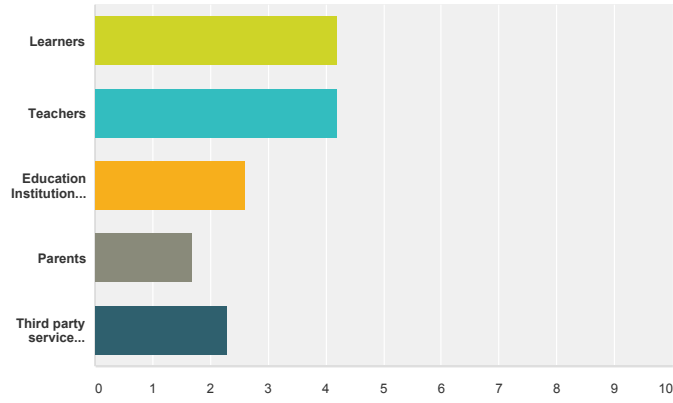
Appendix C

Requirements Survey Results

Q1

In your opinion, who will benefit the most from Learning Analytics? Please rank the following answers according to your opinion (drag&drop).

Answered: 10 Skipped: 0

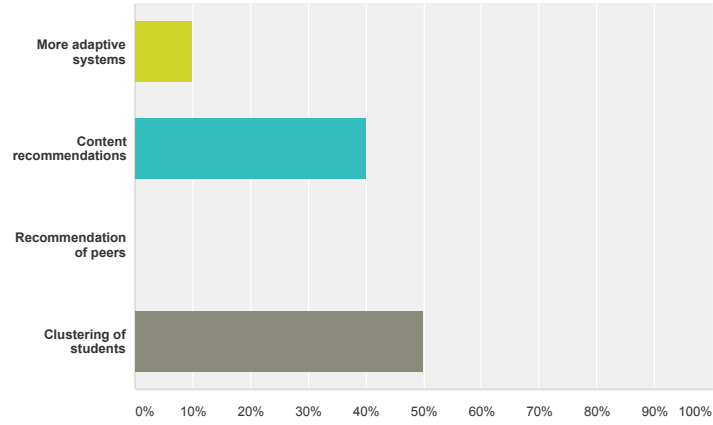


	1	2	3	4	5	Total	Score
▼ Learners	20.00% 2	80.00% 8	0.00% 0	0.00% 0	0.00% 0	10	4.20
▼ Teachers	50.00% 5	20.00% 2	30.00% 3	0.00% 0	0.00% 0	10	4.20
▼ Education Institution (School, College, University, etc.)	0.00% 0	0.00% 0	60.00% 6	40.00% 4	0.00% 0	10	2.60
▼ Parents	0.00% 0	0.00% 0	10.00% 1	50.00% 5	40.00% 4	10	1.70
▼ Third party service providers (companies)	30.00% 3	0.00% 0	0.00% 0	10.00% 1	60.00% 6	10	2.30

Q2

Which added innovation can Learning Analytics bring to educational technologies?

Answered: 10 Skipped: 0

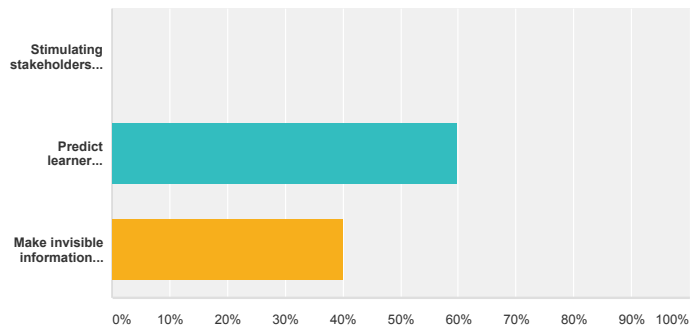


Answer Choices	Responses
More adaptive systems	10.00% 1
Content recommendations	40.00% 4
Recommendation of peers	0.00% 0
Clustering of students	50.00% 5
Total	10

Q3

What should the main objective for Learning Analytics be?

Answered: 10 Skipped: 0

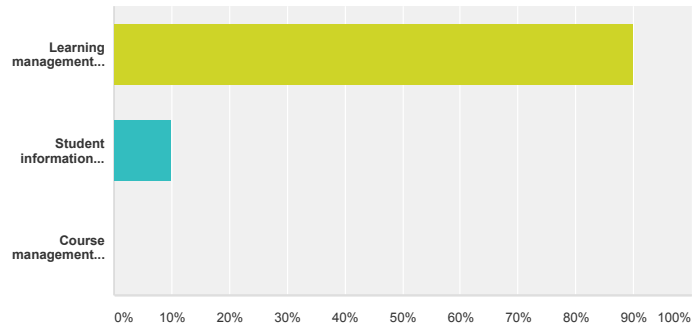


Answer Choices	Responses
Stimulating stakeholders to reflect on their performance and learning behaviour	0.00% 0
Predict learner performance and lead to automatic system adaptation	60.00% 6
Make invisible information about learners visible	40.00% 4
Total	10

Q4

Which IT systems does your organization use for teaching and learning?

Answered: 10 Skipped: 0

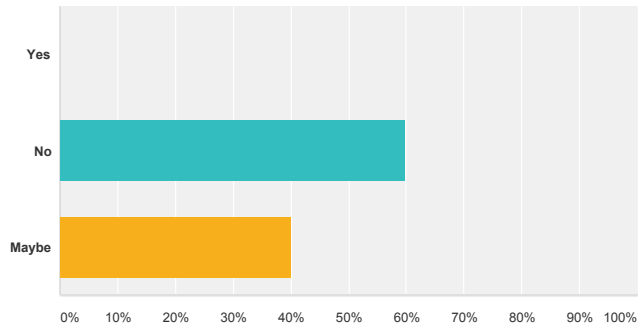


Answer Choices	Responses
Learning management system (e.g., Blackboard, Moodle)	90.00% 9
Student information system	10.00% 1
Course management system	0.00% 0
Total	10

Q5

Should institutions internally share all the data they hold on students with all members of staff (academic, technical, administrative, research)?

Answered: 10 Skipped: 0

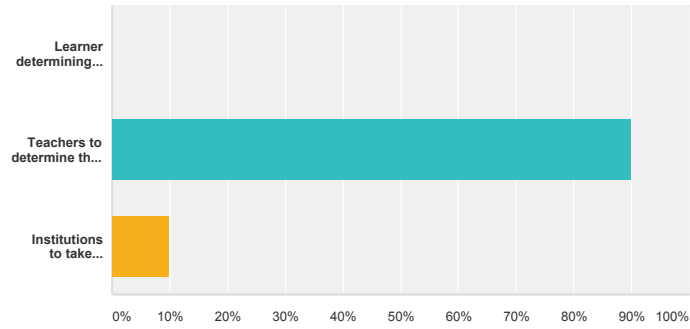


Answer Choices	Responses
Yes	0.00% 0
No	60.00% 6
Maybe	40.00% 4
Total	10

Q6

How sure can we be that end users will draw the right conclusions from their data, and decide on the best course of action?

Answered: 10 Skipped: 0



Answer Choices	Responses
▼ Learner determining their next step	0.00% 0
▼ Teachers to determine the next step for their students	90.00% 9
▼ Institutions to take positive policy decisions	10.00% 1
Total	10

Appendix D

Usability Evaluation

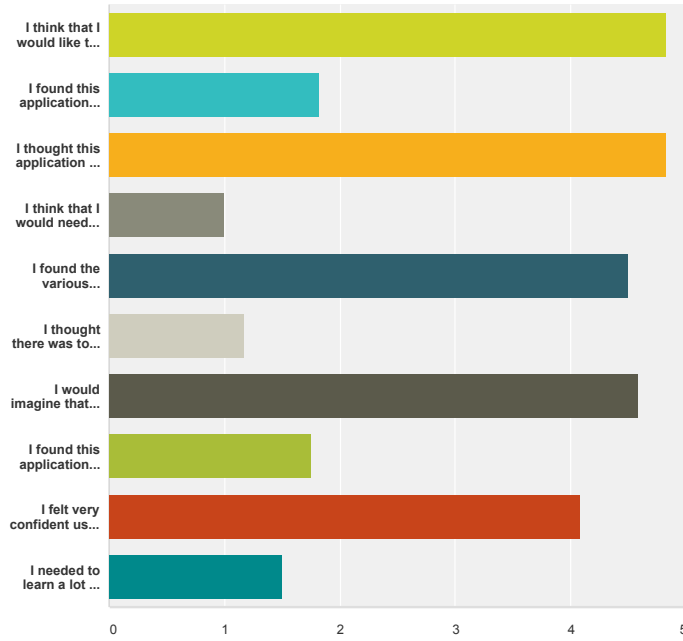
D.1 SUS Scoring

SUS yields a single number representing a composite measure of the overall usability of the system being studied. Note that scores for individual items are not meaningful on their own. To calculate the SUS score, first sum the score contributions from each item. Each item's score contribution will range from 0 to 4. For items 1, 3, 5, 7 and 9 the score contribution is the scale position minus 1. For items 2, 4, 6, 8 and 10, the contribution is 5 minus the scale position. Multiply the sum of the scores by 2.5 to obtain the overall value of SU. SUS scores have a range of 0 to 100.

Q1

SUS Evaluation

Answered: 12 Skipped: 0



	1 Strongly disagree	2	3	4	5 Strongly agree	Total	Weighted Average
▼ I think that I would like to use this application frequently.	0.00% 0	0.00% 0	0.00% 0	16.67% 2	83.33% 10	12	4.83
▼ I found this application unnecessarily complex.	16.67% 2	83.33% 10	0.00% 0	0.00% 0	0.00% 0	12	1.83
▼ I thought this application was easy to use.	0.00% 0	0.00% 0	0.00% 0	16.67% 2	83.33% 10	12	4.83
▼ I think that I would need assistance to be able to use this application.	100.00% 12	0.00% 0	0.00% 0	0.00% 0	0.00% 0	12	1.00
▼ I found the various functions in this application were well integrated.	0.00% 0	0.00% 0	0.00% 0	50.00% 6	50.00% 6	12	4.50
▼ I thought there was too much inconsistency in this application.	83.33% 10	16.67% 2	0.00% 0	0.00% 0	0.00% 0	12	1.17
▼ I would imagine that most people would learn to use this application very quickly.	0.00% 0	0.00% 0	0.00% 0	41.67% 5	58.33% 7	12	4.58
▼ I found this application very cumbersome/awkward to use.	50.00% 6	25.00% 3	25.00% 3	0.00% 0	0.00% 0	12	1.75
▼ I felt very confident using this application.	0.00% 0	0.00% 0	16.67% 2	58.33% 7	25.00% 3	12	4.08
▼ I needed to learn a lot of things before I could get going with this application.	66.67% 8	16.67% 2	16.67% 2	0.00% 0	0.00% 0	12	1.50

D.2 Think Aloud Evaluation Task

Interactive ‘Think Aloud’ Evaluation

This sheet lists the tasks involved in this evaluation. Please note that you may not ask for help completing the tasks, but you may ask for clarification on what each task is asking you to do. You are encouraged to talk through your thought process as you are carrying out these tasks, the evaluator will take notes based on this.

Remember it is the application, not you, that is being tested.

Please read through the in app User Guide. This should take 1-2 minutes.

For these tasks you will take on the role of an educator. The educator has collected all the data of students and wish to analyze these student data using the application – *Learning Analytics*.

1. Copy and paste the excel file into MATLAB platform
2. Run the application
3. Select file to analyze from the application
4. Navigate through the application
Please inform the demonstrator what do you think each tab would represent
5. Go to the ‘Regression’ Tab and try on the Grade calculator
6. Try and send a email and SMS notification under ‘Intervention’ Tab

D.3 Evaluation Survey Results

Q1

What data would you add to Learning Analytics?

Answered: 12 Skipped: 0

Responses (12)

Text Analysis

My Categories

Categorize as...

Filter by Category

Search responses



Showing 12 responses

grades, student portal activities, class activities

3/21/2015 7:31 PM

[View respondent's answers](#)

Student grades, discussion participations , number of visits

3/21/2015 7:30 PM

[View respondent's answers](#)

student number, email, grades

3/21/2015 6:44 PM

[View respondent's answers](#)

Student grades, knowledge, number of logins

3/21/2015 6:30 PM

[View respondent's answers](#)

topics, class position

3/21/2015 6:00 PM

[View respondent's answers](#)

student number, email, grades

3/21/2015 5:18 PM

[View respondent's answers](#)

topics learn, GPA, discussion forums

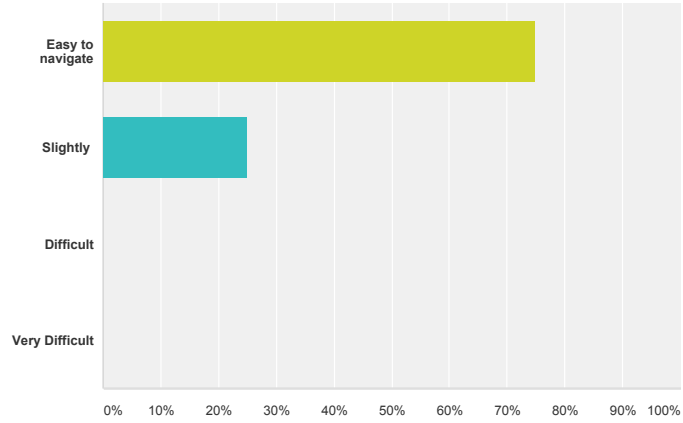
3/21/2015 4:57 PM

[View respondent's answers](#)

Q2

What did you find difficult when using Learning Analytics?

Answered: 12 Skipped: 0

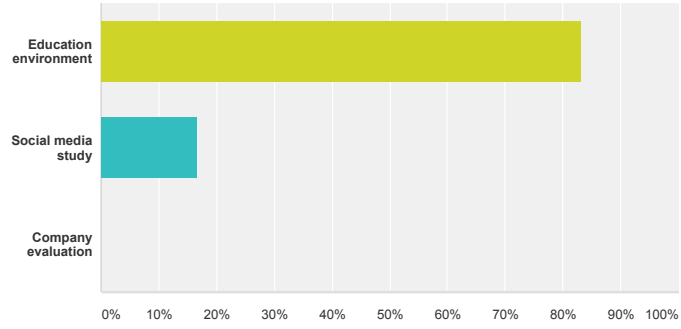


Answer Choices	Responses
▼ Easy to navigate	75.00% 9
▼ Slightly	25.00% 3
▼ Difficult	0.00% 0
▼ Very Difficult	0.00% 0
Total	12

Q3

When would you imagine using Learning Analytics?

Answered: 12 Skipped: 0

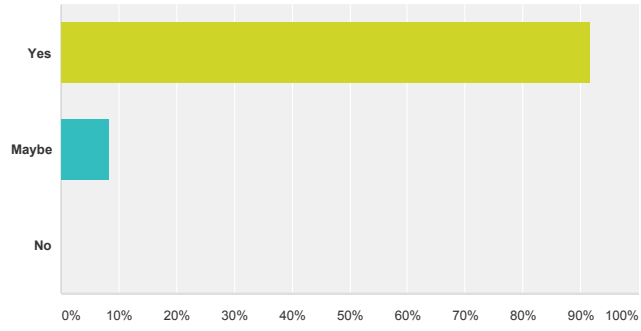


Answer Choices	Responses
Education environment	83.33% 10
Social media study	16.67% 2
Company evaluation	0.00% 0
Total	12

Q4

Would you recommend Learning Analytics?

Answered: 12 Skipped: 0



Answer Choices	Responses	
▼ Yes	91.67%	11
▼ Maybe	8.33%	1
▼ No	0.00%	0
Total		12

Appendix E

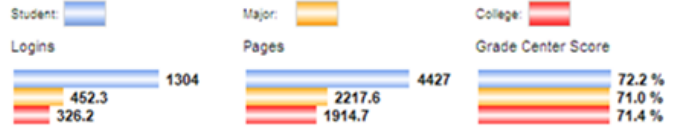
Full Images



General Information

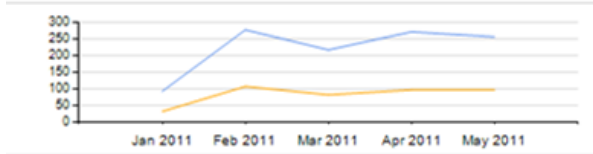
Student ID: 372275 **Email:**
Standing: Good **Phone:**
Primary Major: Psychology - BA **Class:** Junior
College: 1 **Level:** Undergraduate

Student Totals vs. Major & College Averages

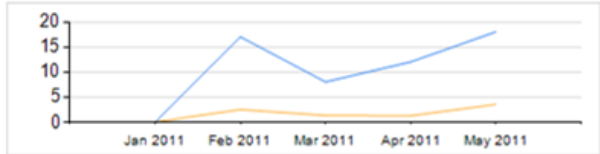


Logins vs. Avg.

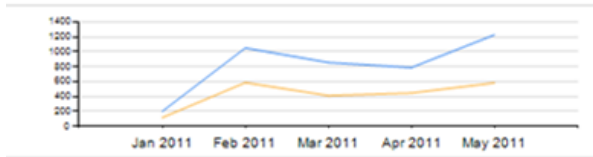
Student: Major:



Submissions vs. Avg.



Interactions vs. Avg.



Time vs. Avg.



Blackboard Course History

Status	Term	Learn Course Section	Interactions Trend	Total Logins		Total Minutes		Total Interactions		Total S
				User	Avg.	User	Avg.	User	Avg.	
Active	Spring 2011	Human Anatomy and Physiology I (BIOL 251) (01)		238	88.5	2758	827.5	616	298.8	U:
Active	Spring 2011	LAB for Human Anatomy and Physiology I (BIOL 251L) (01)		176	71.3	1930	965.3	327	174.8	U:

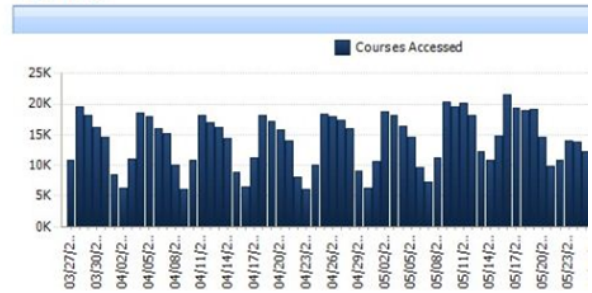
- ProClarity
- Reporting Services
- Data Warehouse Documentation
- Dashboard Designer
- Recycle Bin
- All Site Content

Term: **Spring 2011** College: **All Colleges** Department: **All Departments**

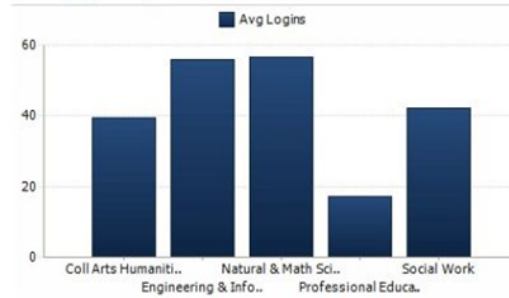
Learn Activity

	Actual	Goal			
Course Logins	21,280	20,216	●		
Avg Logins	46.2	50	▲ -8%	41.8	▲ 10%
Avg Pages	.8	.8	● 0%	.7	▲ 12%
Avg Minutes	932.7	900	● 4%	839.5	▲ 11%
Courses Using Bb	677	677	● 0%	609	▲ 11%
Students Enrolled	10,057	28000	▼ -64%	9,152	▲ 10%

Login Trend



Activity by College



Top Ten Most Active Courses





Course by Class number	Enrollment Count	Logins ~	Pages ~	Avg Logins	Avg Pages
CHEM 102	538	46,026	259,764	85.6	.8
MATH 151	349	26,205	119,887	75.1	.8
CHEM 101	329	24,052	114,330	73.1	.8
PSYC 100	396	17,930	344,453	45.3	.8
SPAN 201	319	17,269	91,024	54.1	.8
BIOL 141	268	15,777	75,568	58.9	.8
BIOL 303	245	15,060	67,777	61.5	.8
STAT 350	215	14,043	67,614	65.3	.8
SCI 100	200	13,754	146,291	68.8	.8
IS 698	116	12,744	133,331	109.9	.8

Engagement analytics for course: MBA - HRM

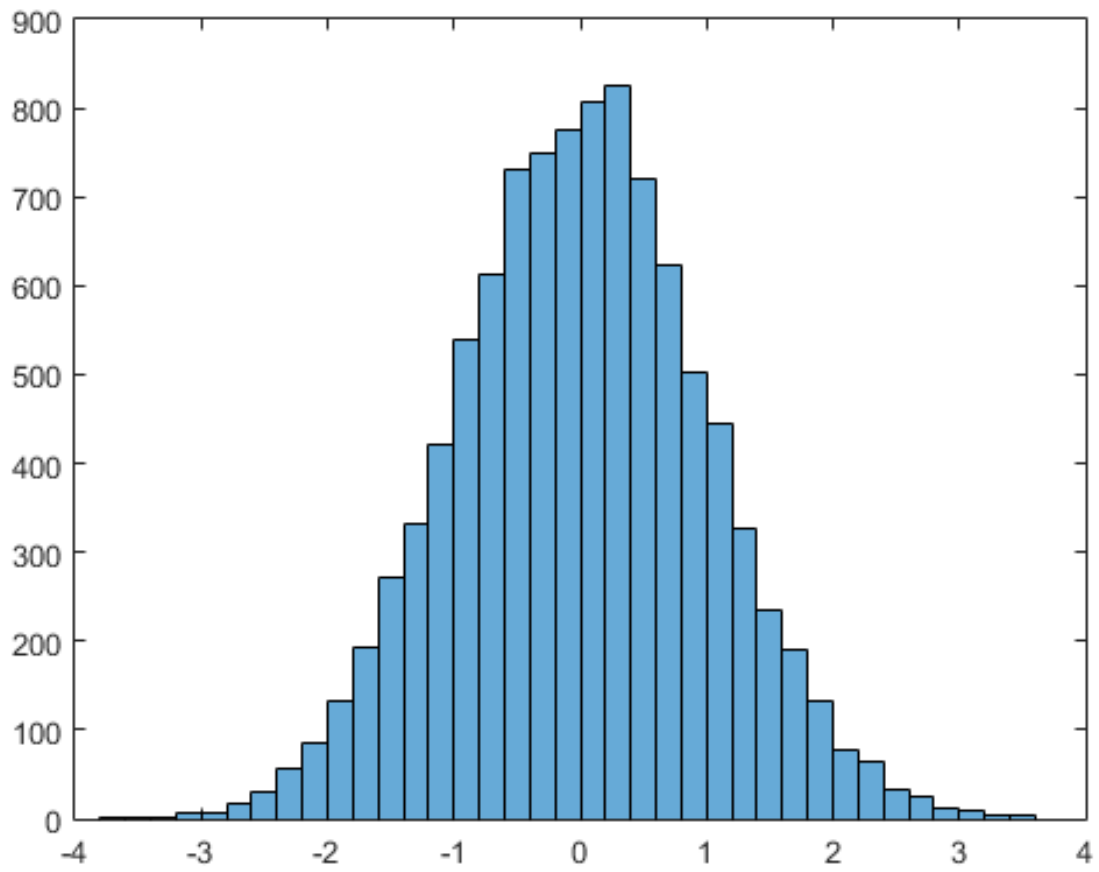
Scores are shown as *weighted score (raw score)*

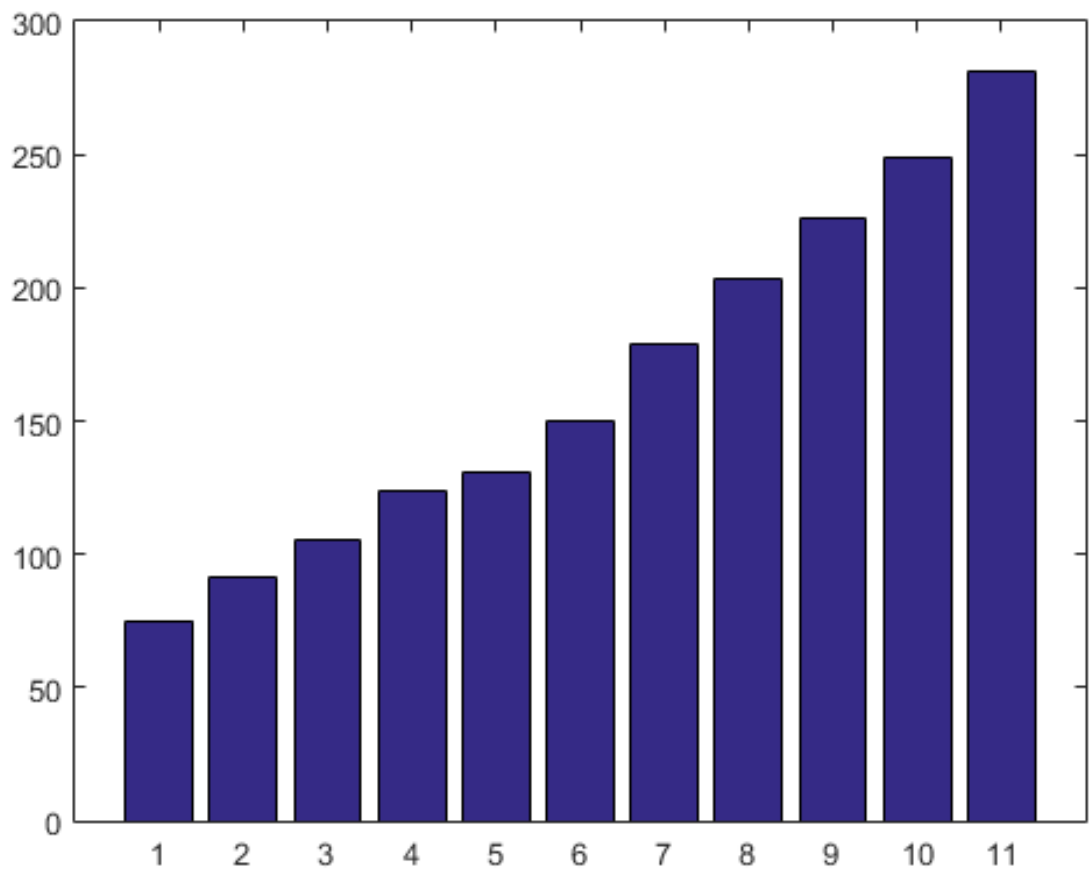
Username	Assessment Activity	Forum Activity	Login Activity	Total ↑
Ken Student	25% (50%)	30% (100%)	20% (100%)	75%
Frances Smart	10% (20%)	30% (100%)	4% (20%)	44%
Dougal McFarlane	2% (4%)	30% (100%)	4% (20%)	36%
Angela Edgar	25% (50%)	30% (100%)	20% (100%)	75%

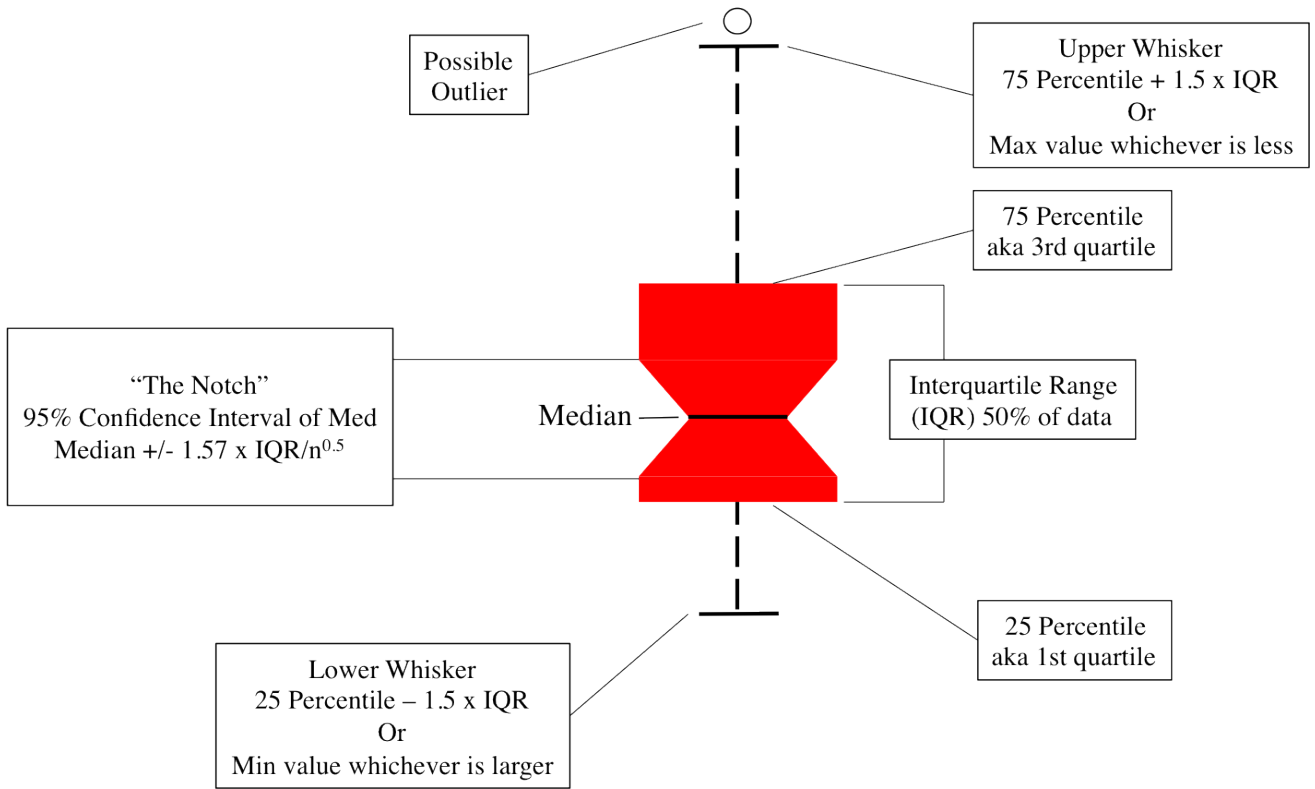
Engagement analytics

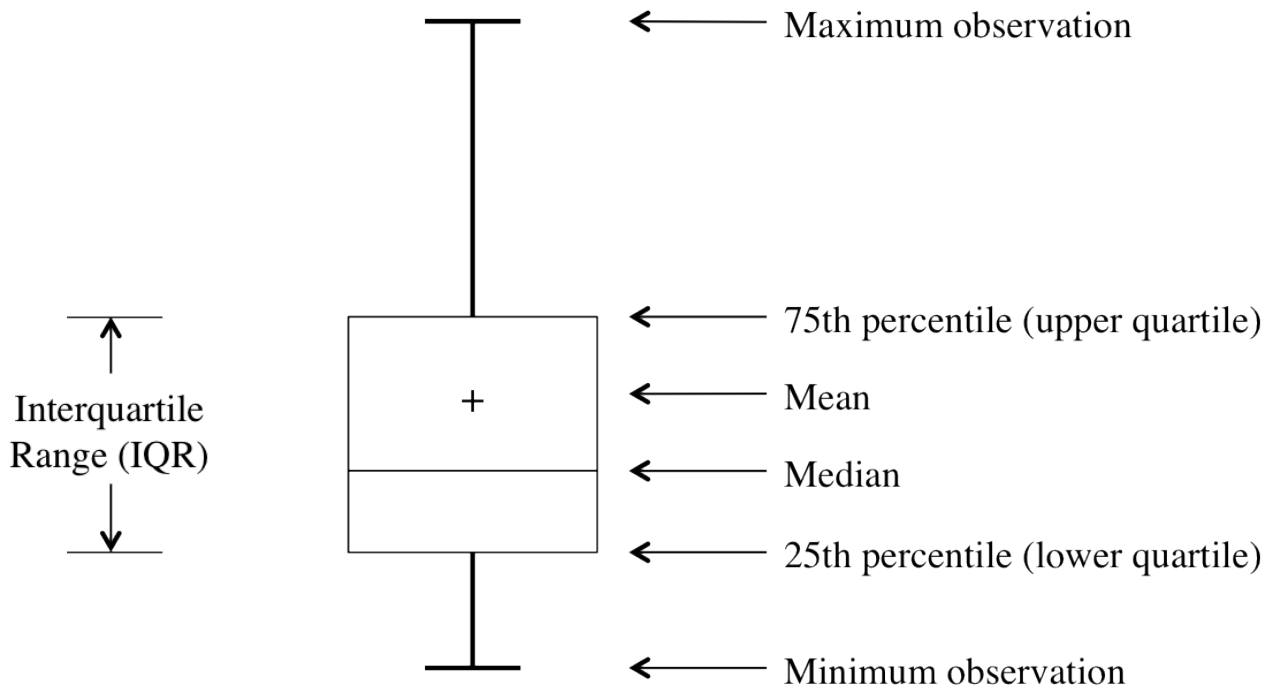
-  **Ken Student** 75%
-  **Angela Edgar** 75%
-  **Frances Smart** 44%
-  **Dougal McFarlane** 36%

[View course report](#)

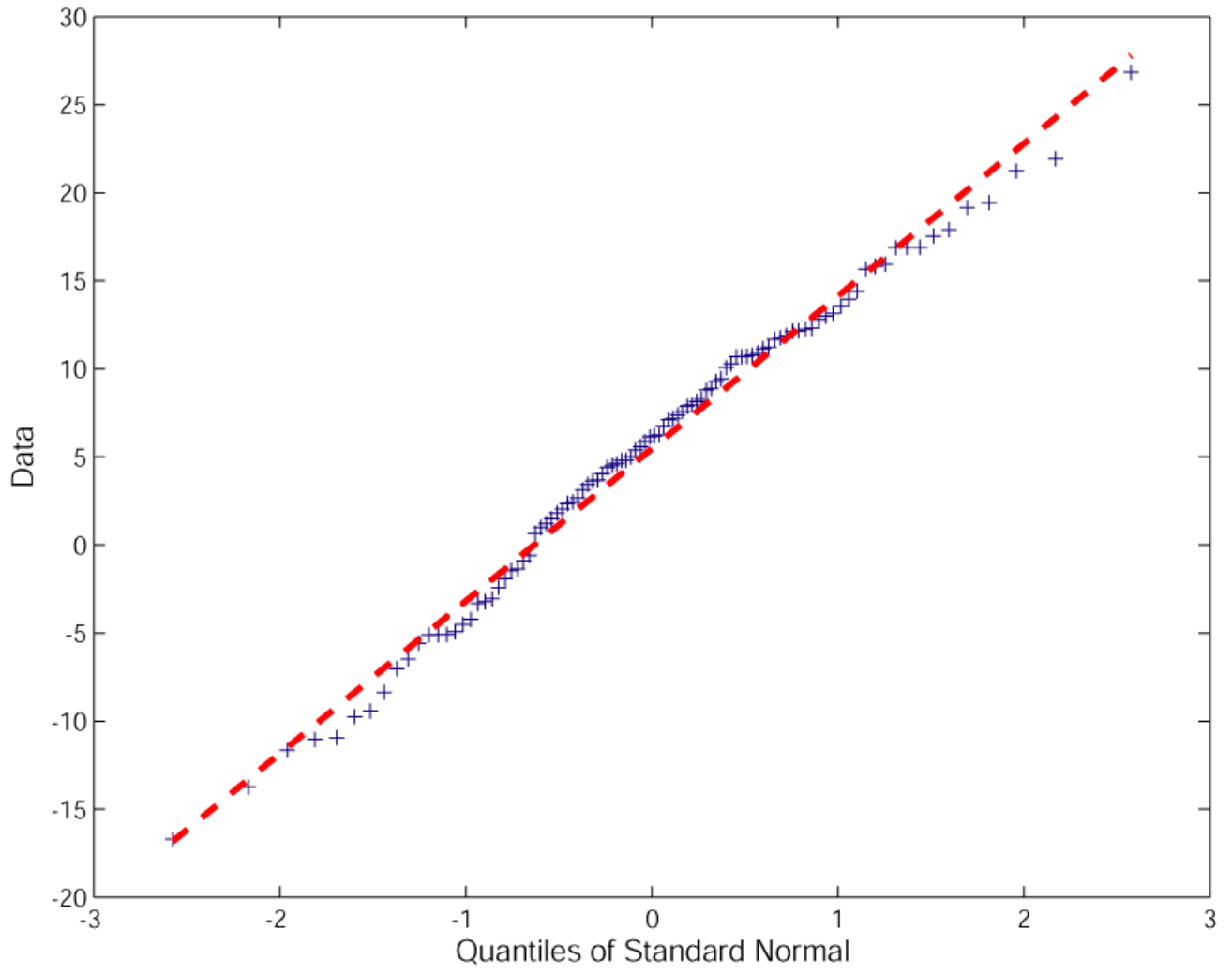


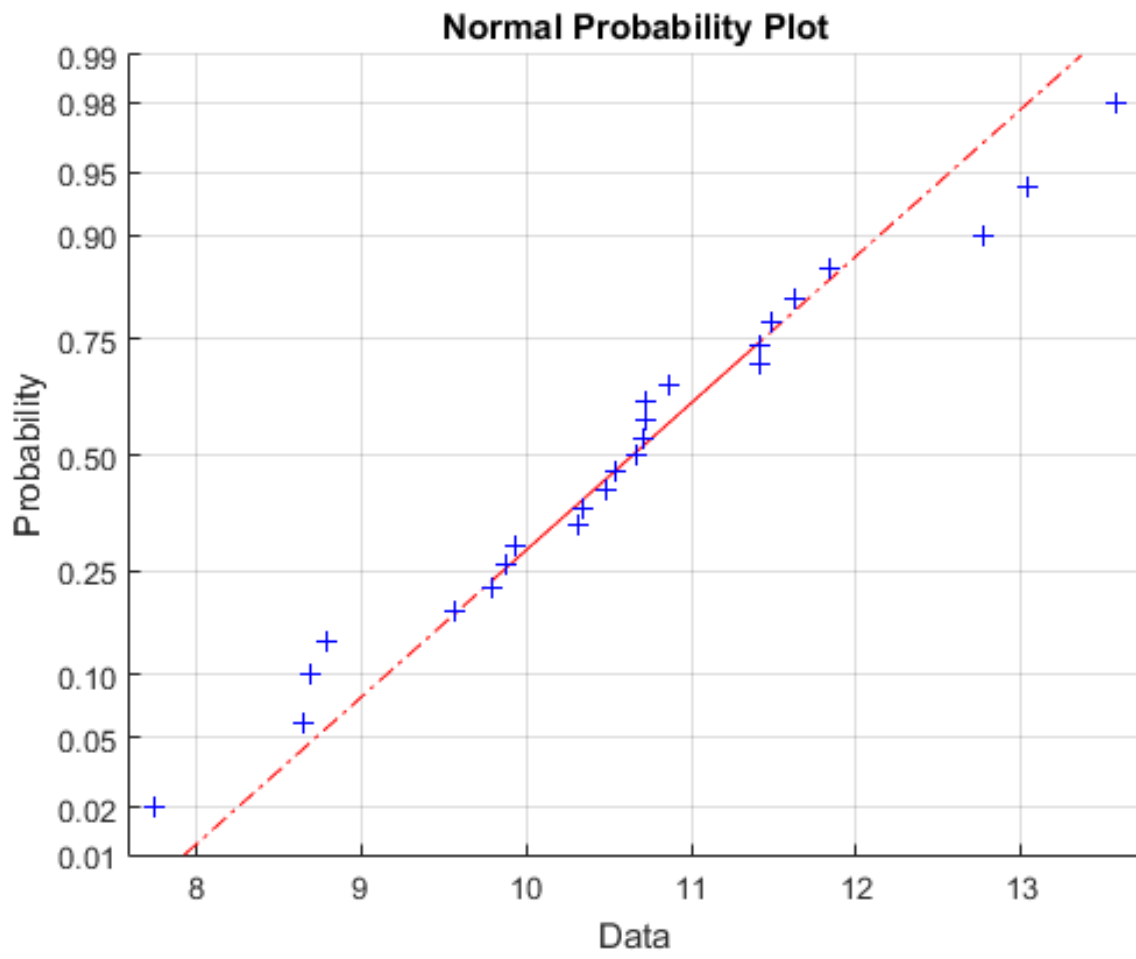


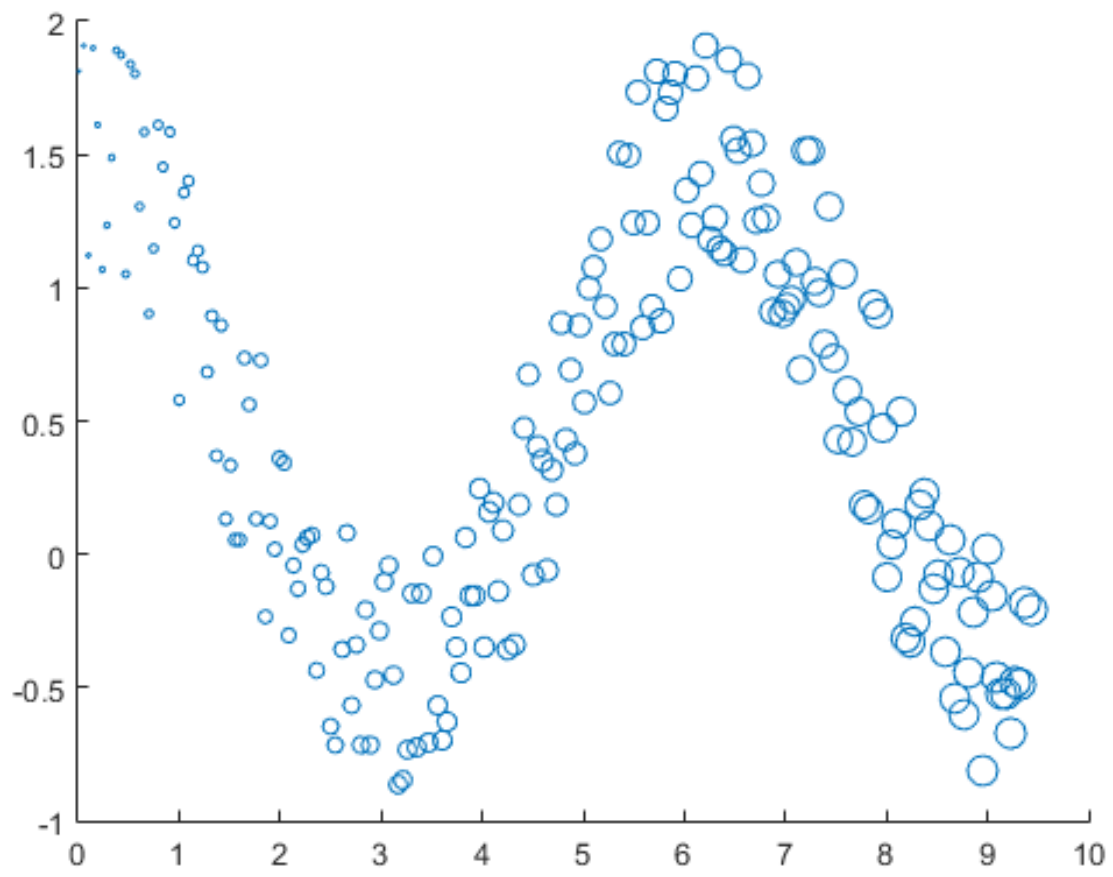


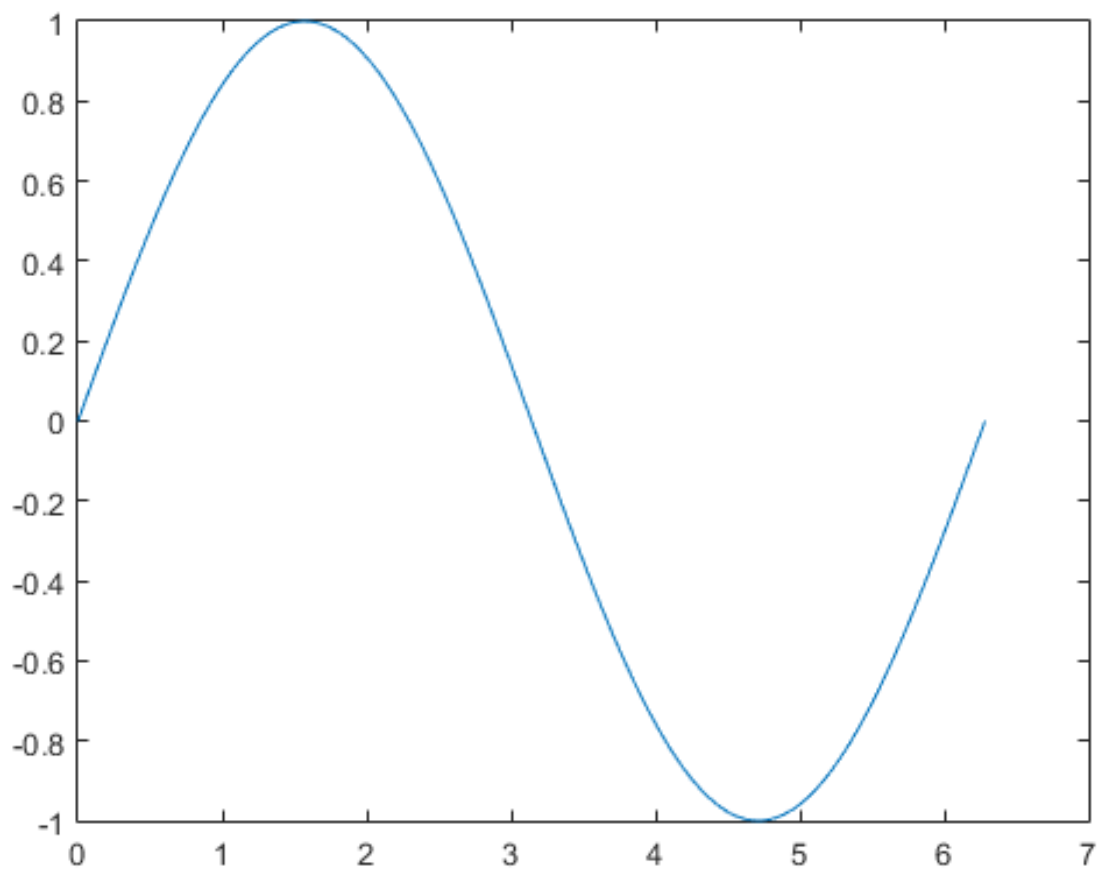


Normal Quantile-Quantile Plot

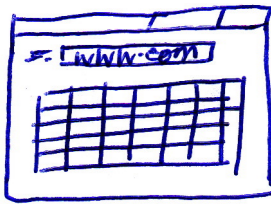




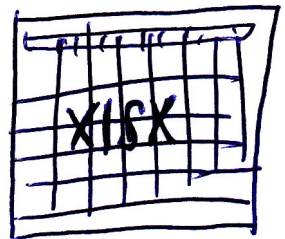
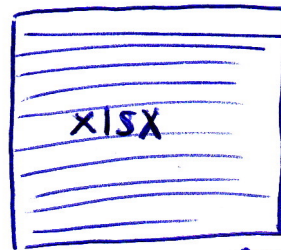




Preparing to Analyse Student using Learning Analytics.



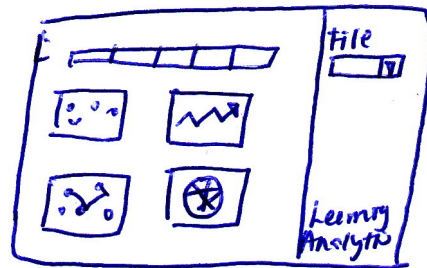
(1) Import student data and information from existing student portal into ~~the~~ xlsx format.



(2) Open excel file and check if student data are formatted correctly from the portal.

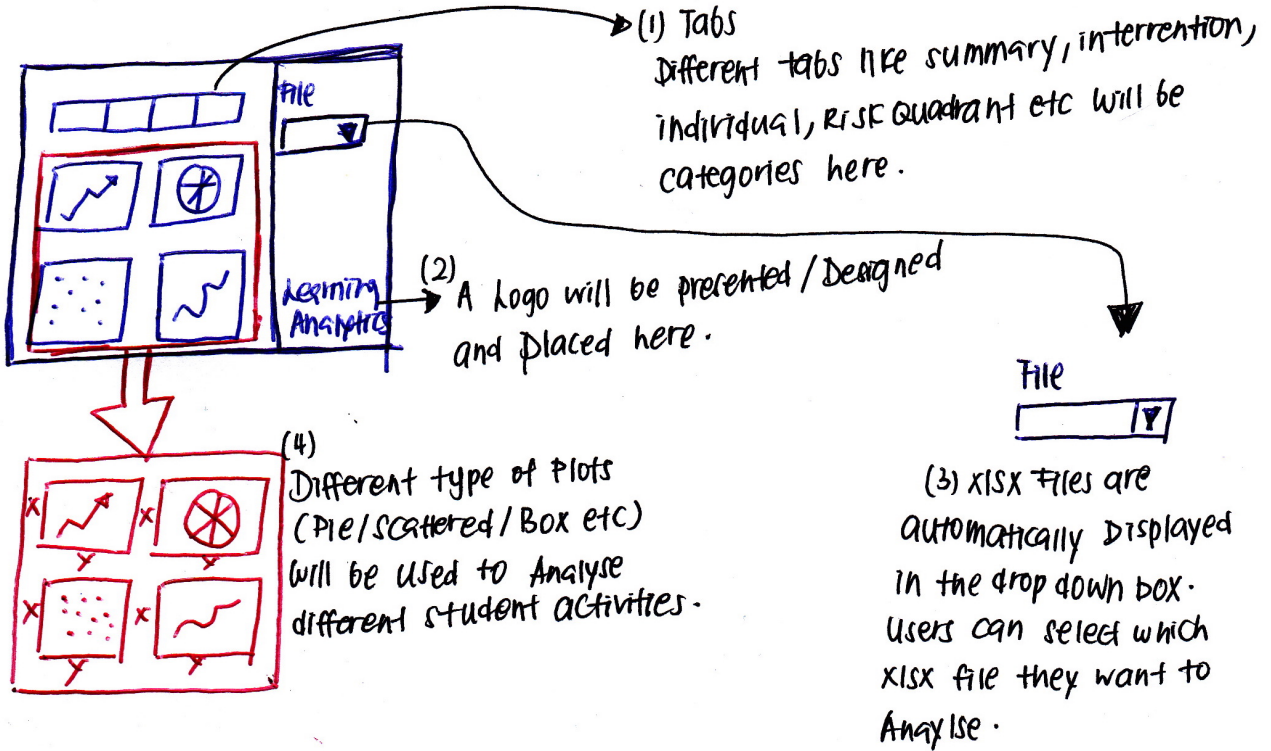


Drop the xlsx file into the Learning Analytics folder.

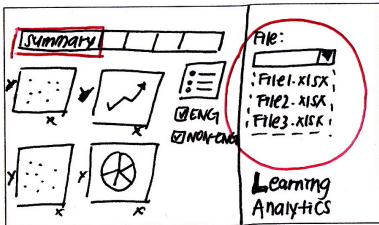


(4) open up the application and select files ~~at~~ added for viewing.

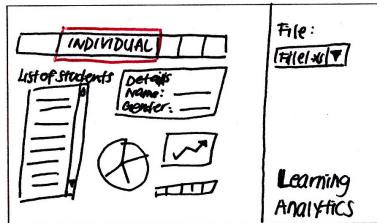
How to Present



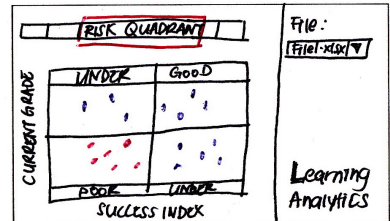
WIREFRAME



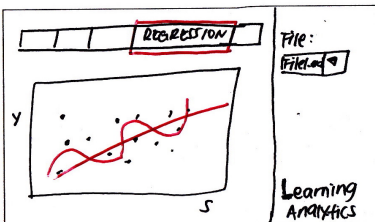
Users can chose which files they would want to view. This is the summary tab that shows all the possible plots from the available data.



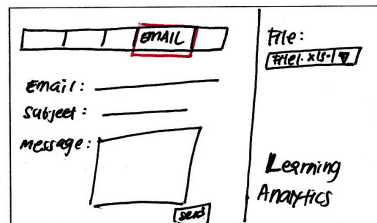
Under individual tab, user can view more about individual students like grades, attendant, discussion, details and more.



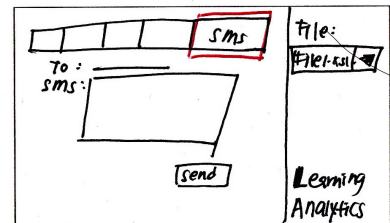
Under risk Quadrant tab, the students are grouped under four different categories. All this categories will be differentiated based on their grades.



Under regression, plots will be displayed to show the correlationship between different student activities and grades.

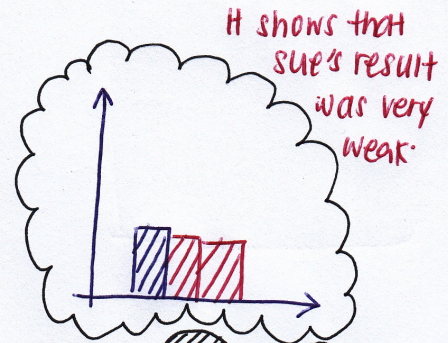
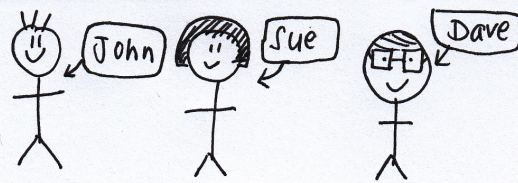
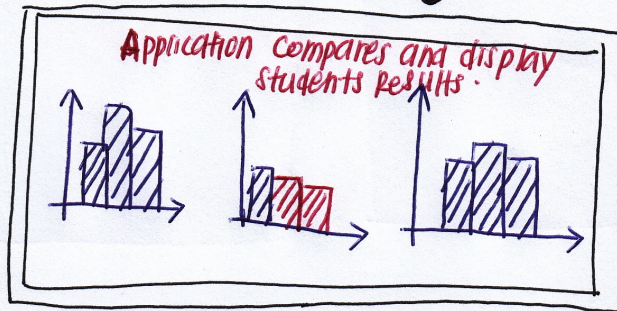


After users has finish analysing the data, they can chose to send email to specific students as a form of interrention.



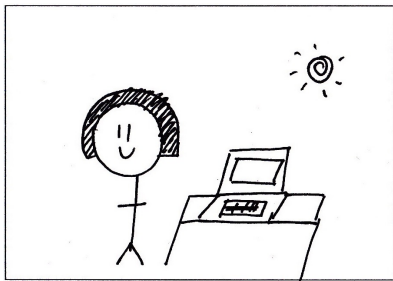
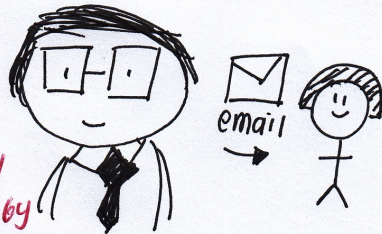
Other than sending email, users can send sms in the last tab to specific student's mobile devices with desired message.

Document Analysing

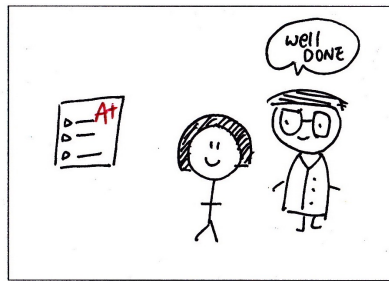


Sue

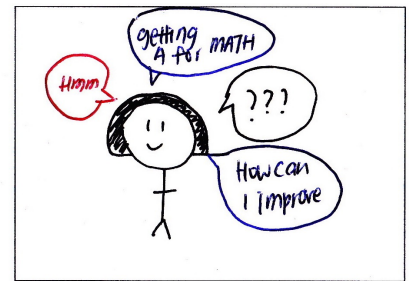
The Application notify the educator regarding this issue. then, educator could wish to contact sue by the Application.



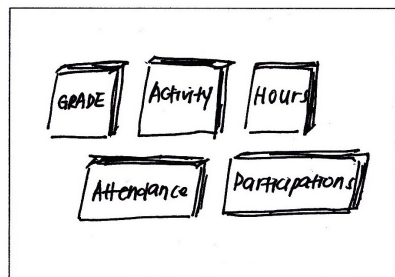
Mary is a Computing science student.



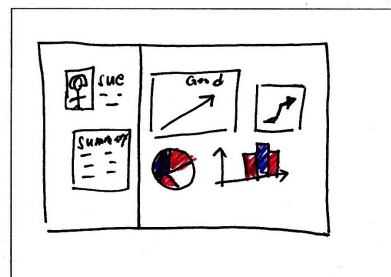
she was constantly very hardworking and all lectures was very pleased with her grades.



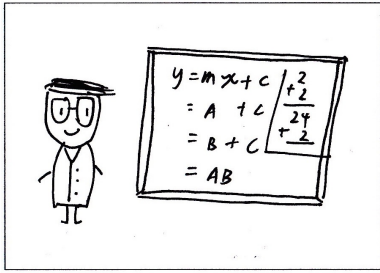
However, she is weak towards certain subjects like mathematics, and she thinks of ways to improve her grades.



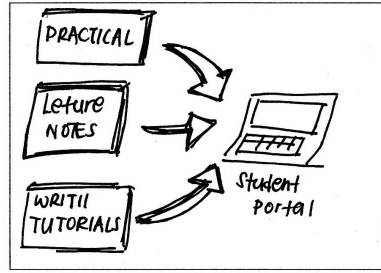
she requested more details regarding her activity logs from the student portal. These activity logs includes hours spend, attendance, participations etc.



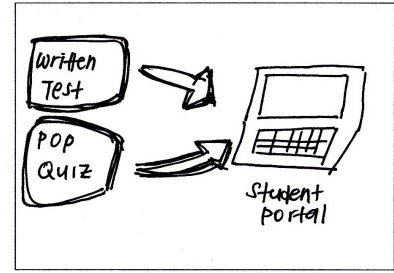
Mary plot all the activity logs against her mathe matic grading. she is able to gain more information and understand herself better.



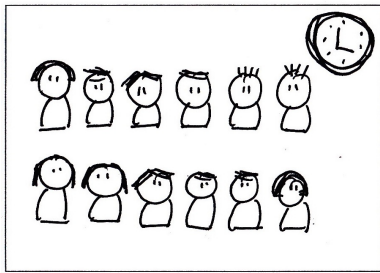
Prof Joe is a lecturer and subject coordinator of school of computing science.



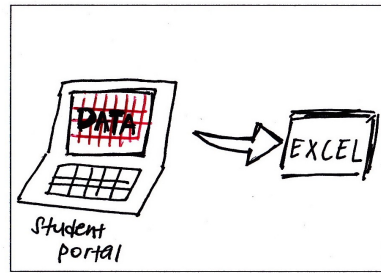
He would often upload his teaching materials like practical, tutorial sheets and lecture notes into the student portal. Students could download those material before lecture.



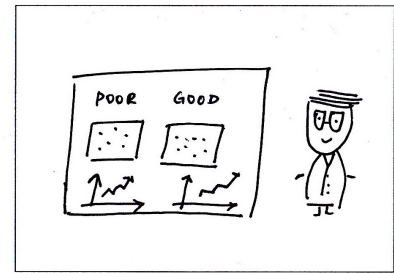
Occasionally, he will conduct mini quizzes to assess on students learning.



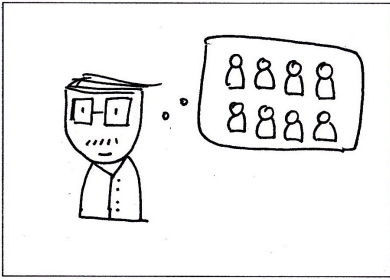
His students took all the tests and results were recorded into the system.



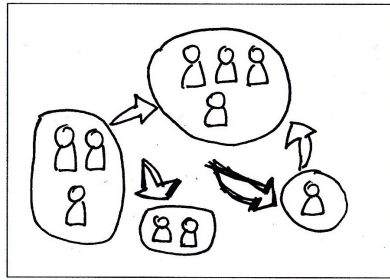
Prof Joe save and export the data into an excel file and check if the file is of correct format.



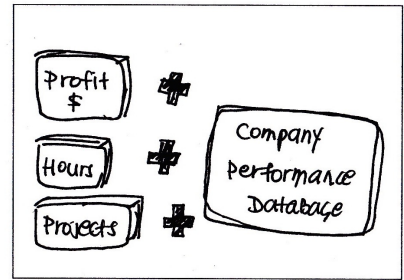
Now, he could monitor and keep track of the student's performance. He can put special attention to students that requires more help.



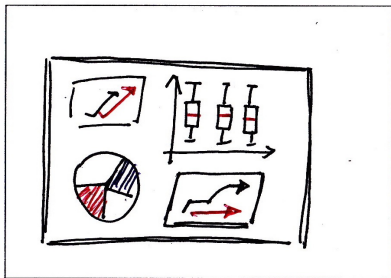
Boss Ken operates a large investment bank with many employees.



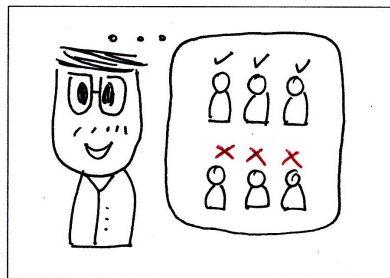
Employees takes up many projects and works very hard to maintain their company's revenue.



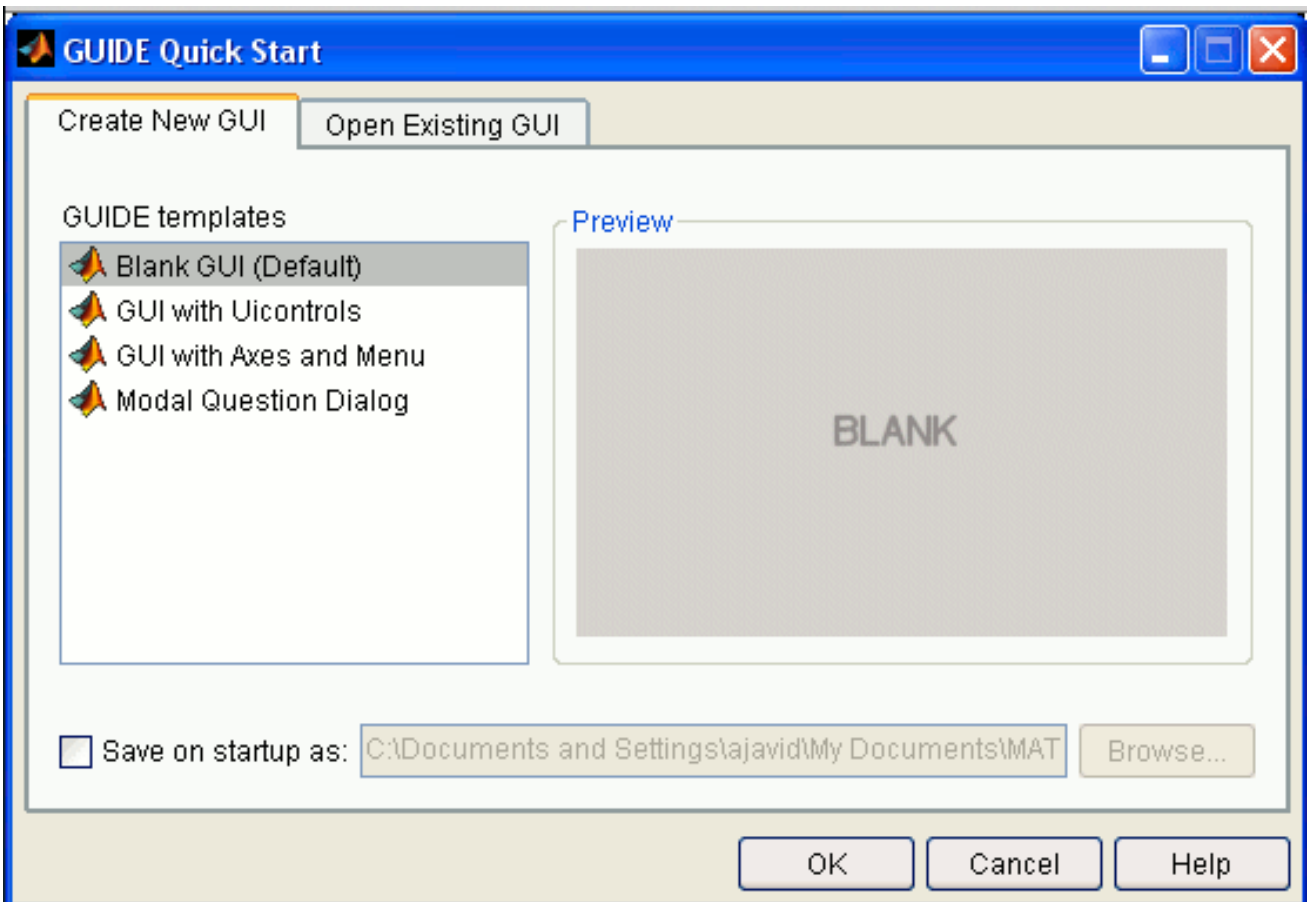
However, Boss Ken would still wants to know more about his company's profit and employee's performances

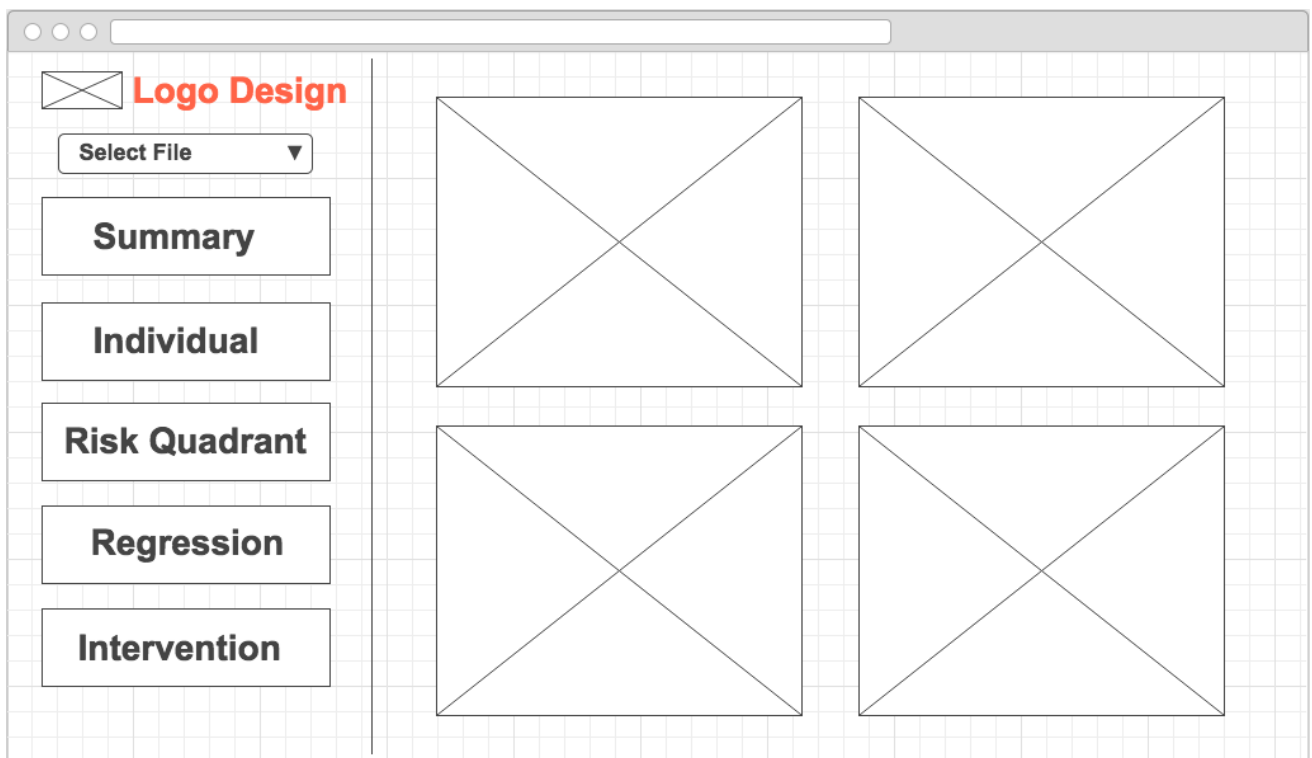
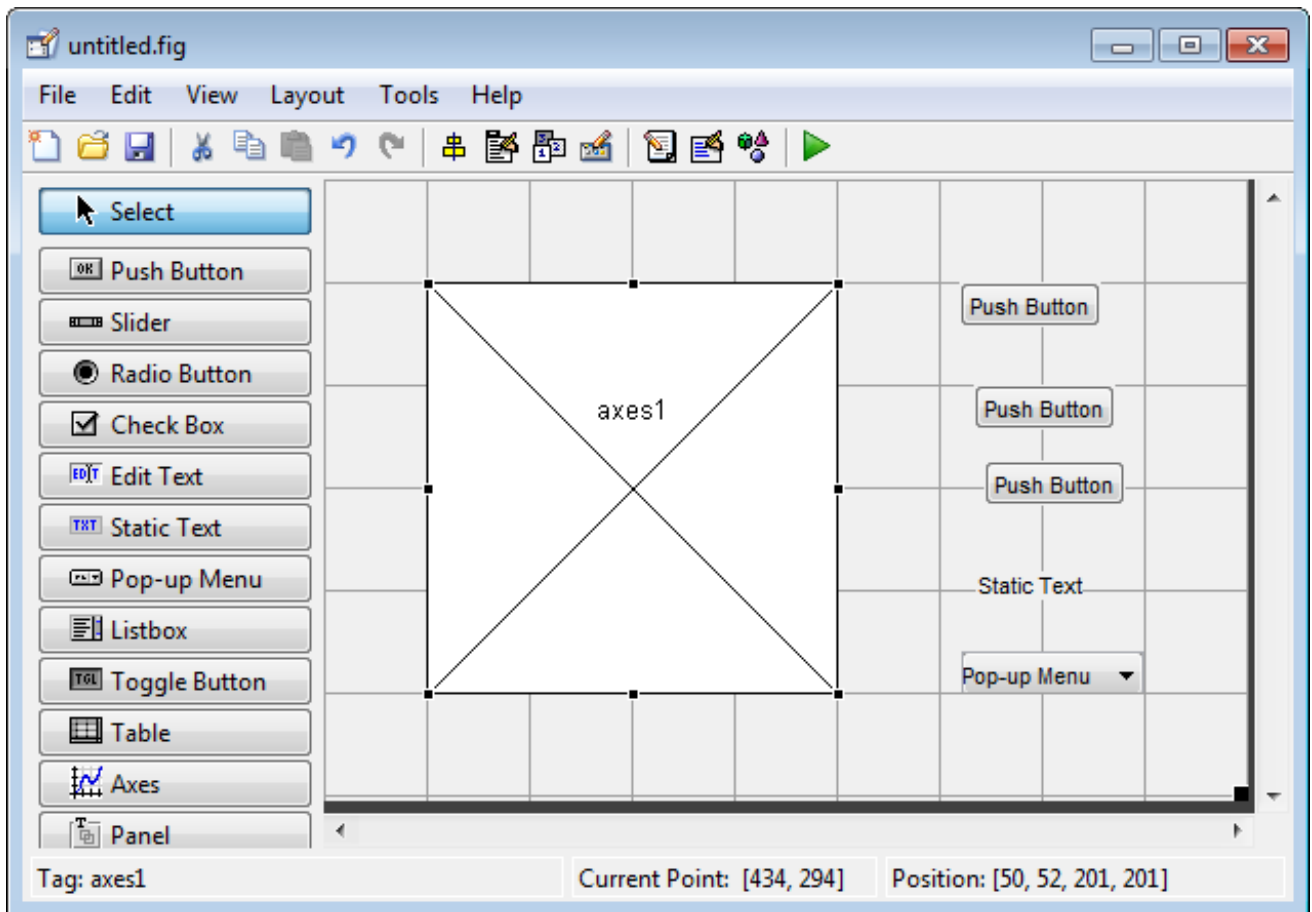


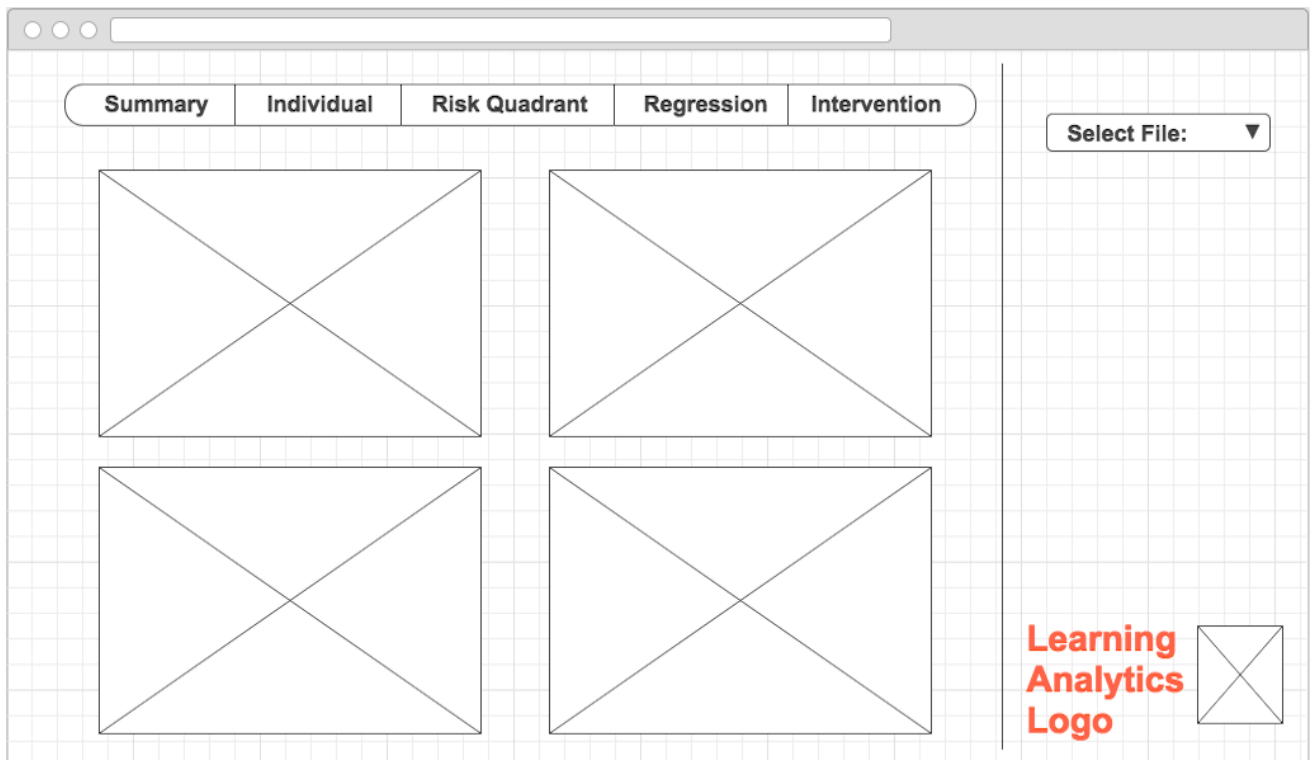
data and predictions are generated into the application based on the information boss Ken given.



NOW, boss Ken is very happy as he is able to monitor employees who is not putting enough effort.







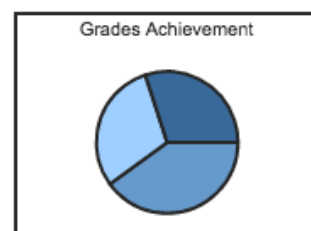
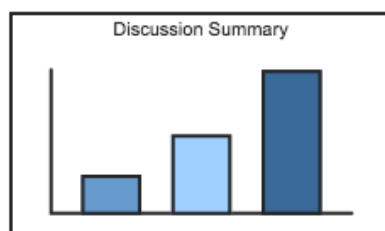
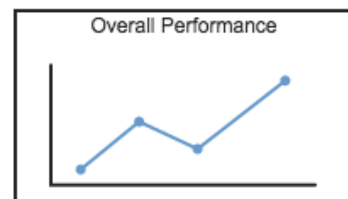
Select Student ID:

- Student 1
- Student 2
- Student 3
- Student 4
- Student 5
- Student 6
- Student 7
- Student 8
- Student 9
- Student 10
- Student 11
- Student 12
- Student 111
- Student 2222
- Student 3333
- Student 44
- Student 55
- Student 66
- Student 77
- Student 88
- Student 99
- Student 100
- Student 111
- Student 122

	Name: School: Gender:	Conduct: Overall Performance: Activities:
---	--	--



Result	Mean	Min	Max	Student's Result
CA	50	34	99	70
Written	45	23	78	60
Practical	45	33	78	76



Select Student ID:

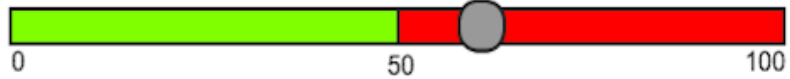
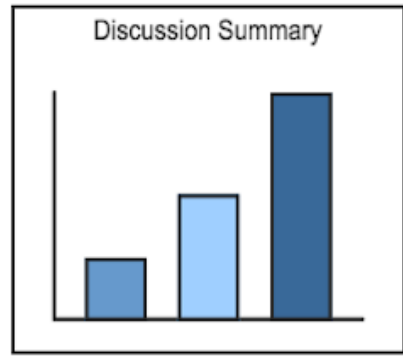
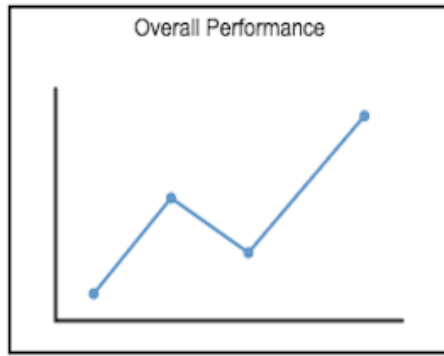
Student111



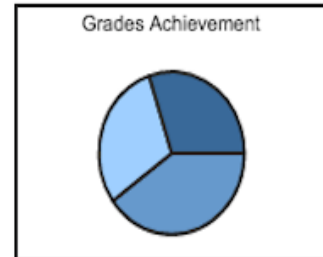
Admin No

Name:
School:
Gender:

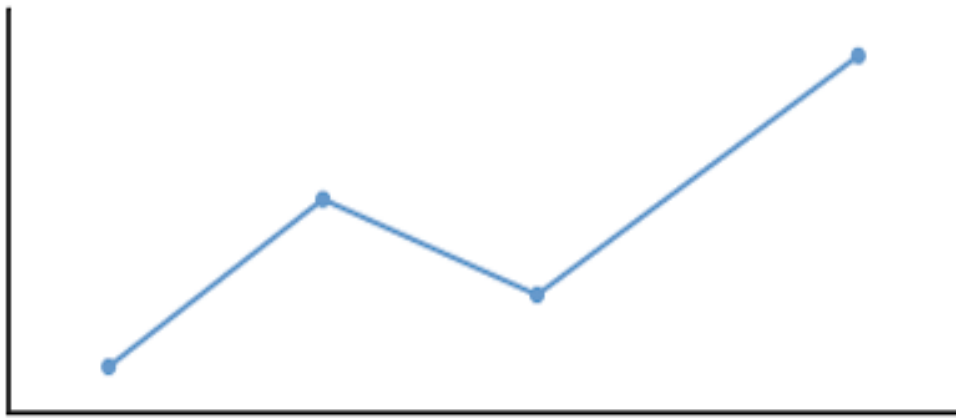
Conduct:
Overall Performance:
Activities:



Result	Mean	Min	Max	Student's Result
CA	50	34	99	70
Written	45	23	78	60
Practical	45	33	78	76



Overall Regression of Marks VS GPA



Engineering Students



Non-Engineering Students



Overall Regression of Written VS GPA



Overall Regression of Practical VS GPA



Overall Regression of Written VS GPA





Learning
Analytics

Summary Individual Risk Quadrant Regression Intervention


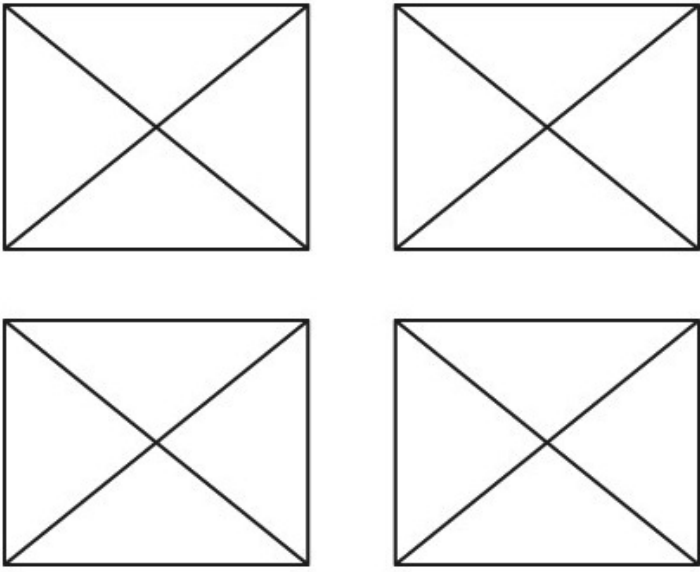
Q Search Select File

Type of Summary

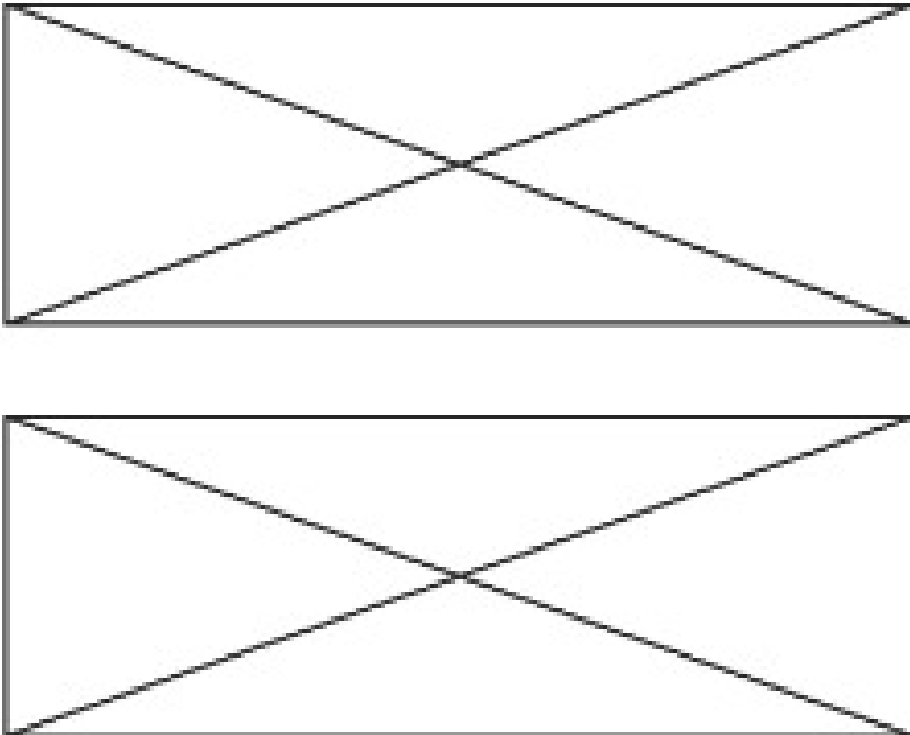
- Activites
- Grade
- Statistics

Type of Student

- Student Type 1
- Student Type 2



Summary Individual Risk Quadrant Regression Intervention




Key in GPA

56

Summary Individual Risk Quadrant Regression Intervention

Select Student ID:

- Student 1
- Student 2
- Student 3
- Student 4
- Student 5
- Student 6
- Student 7
- Student 8
- Student 9
- Student 10
- Student 11
- Student 12
- Student 111
- Student 2222
- Student 3333
- Student 44
- Student 55
- Student 66
- Student 77
- Student 88
- Student 99
- Student 100
- Student 111
- Student 122



Name:


School:

Gender:

Contact:

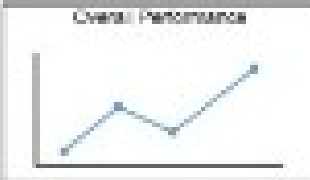
Overall Performance:

Ratios:

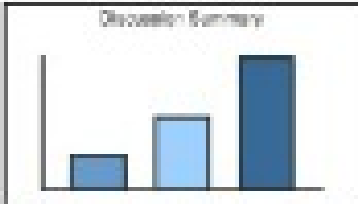


Result	Mean	Min	Max	Student's Result
CA	60	34	99	70
Written	45	23	78	60
Practical	45	33	78	79

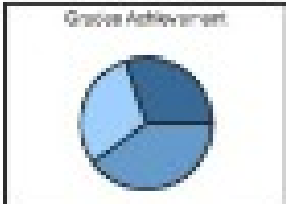
Overall Performance

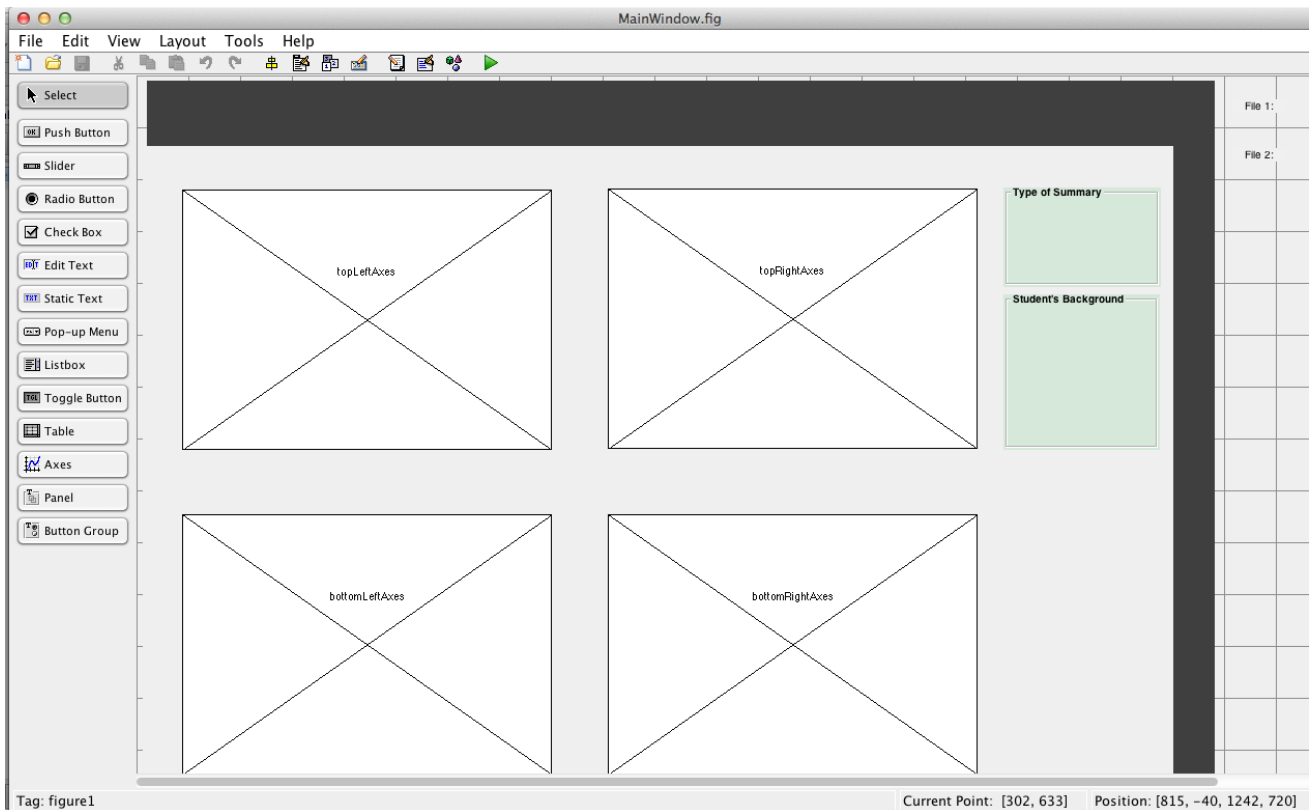
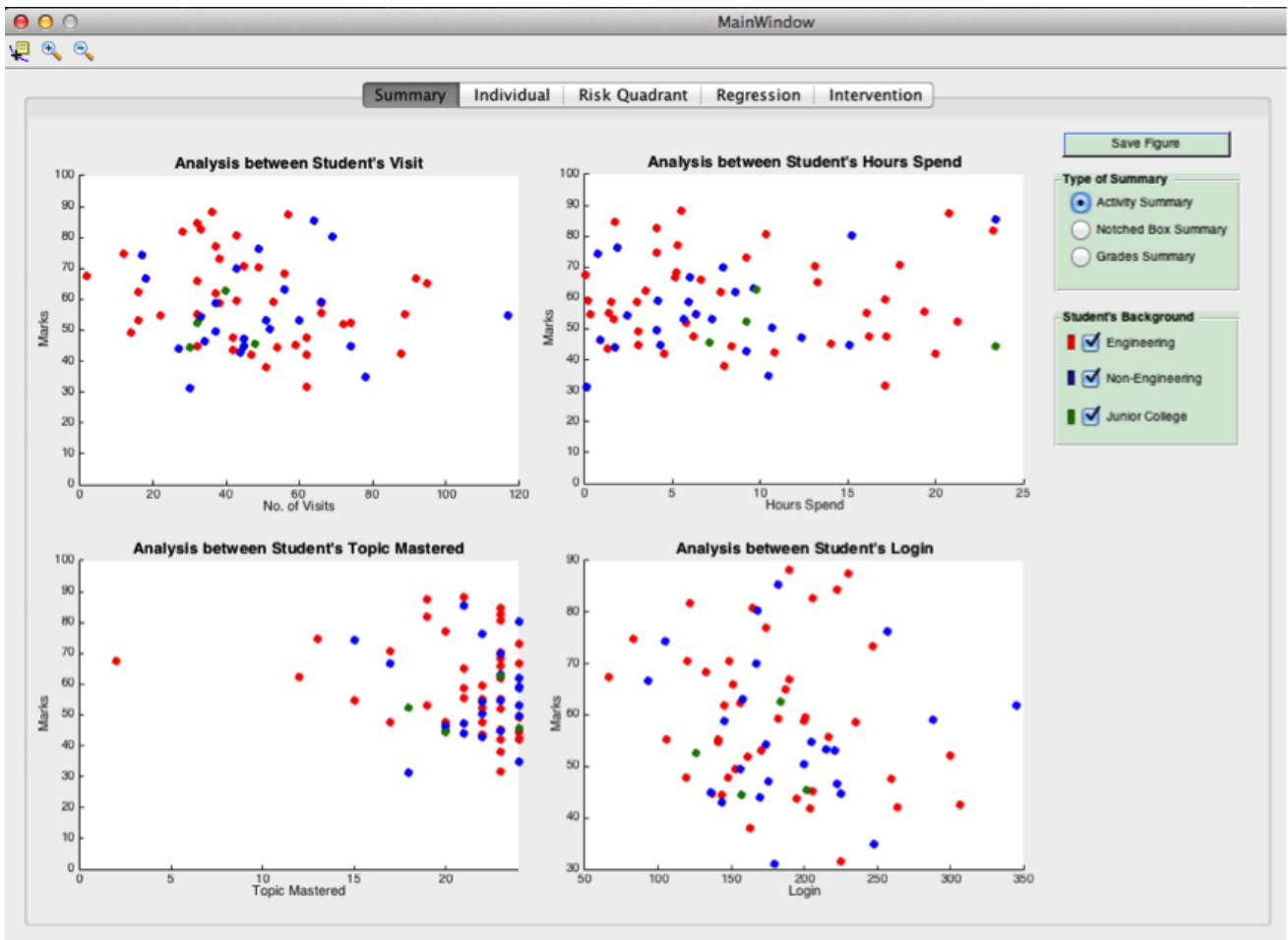


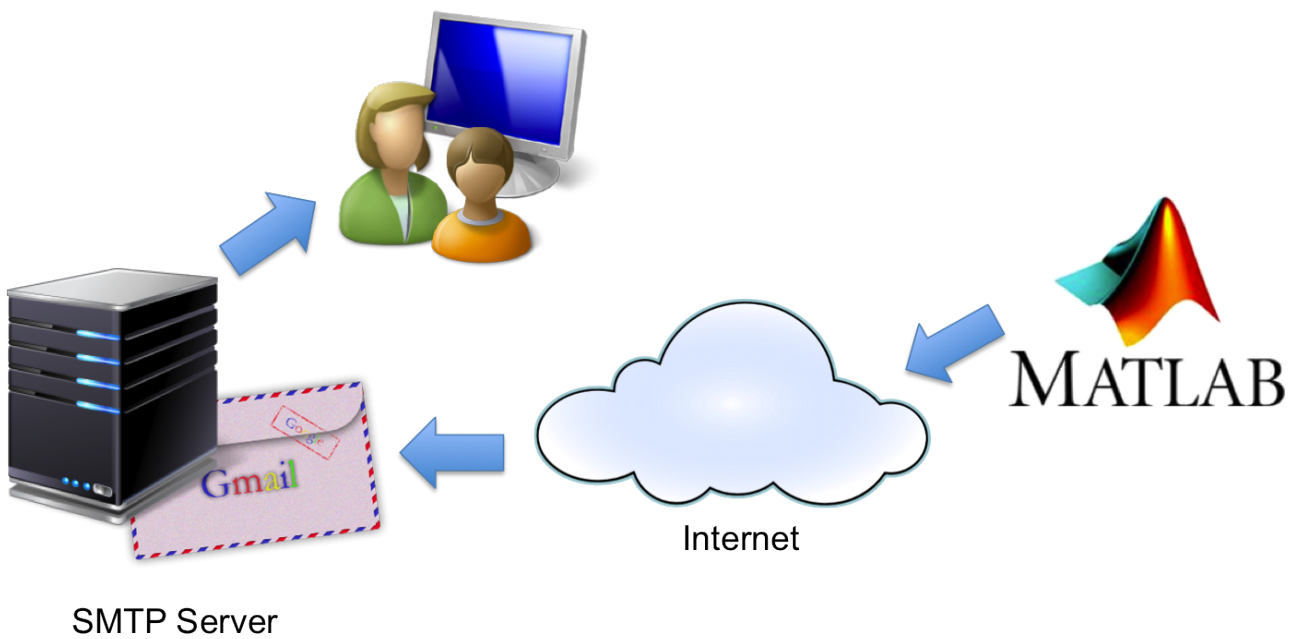
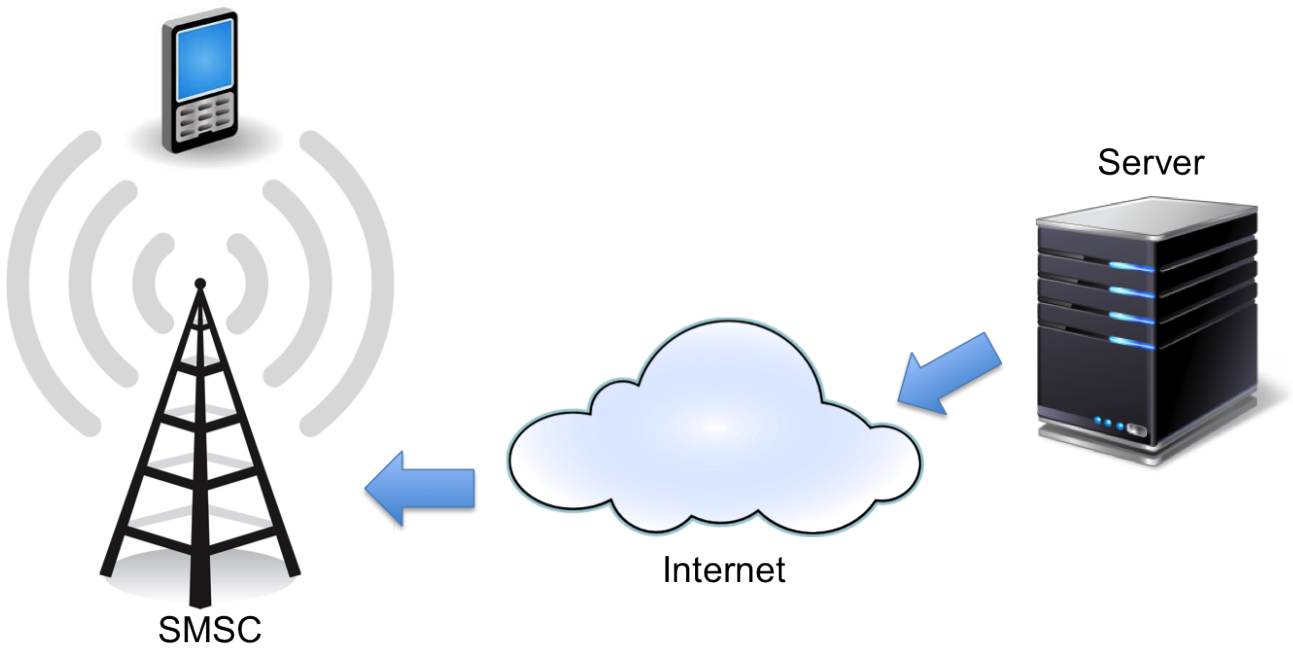
Discussion Summary



Grades Achievement







Summary

Individual

Risk Quadrant

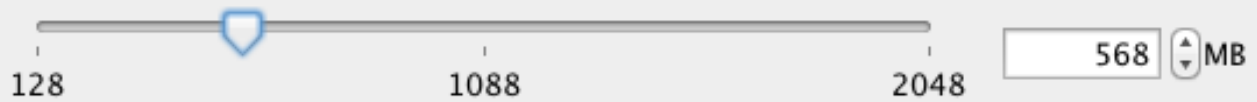
Regression

Intervention

MATLAB General Java Heap Memory Preferences

Java Heap Size

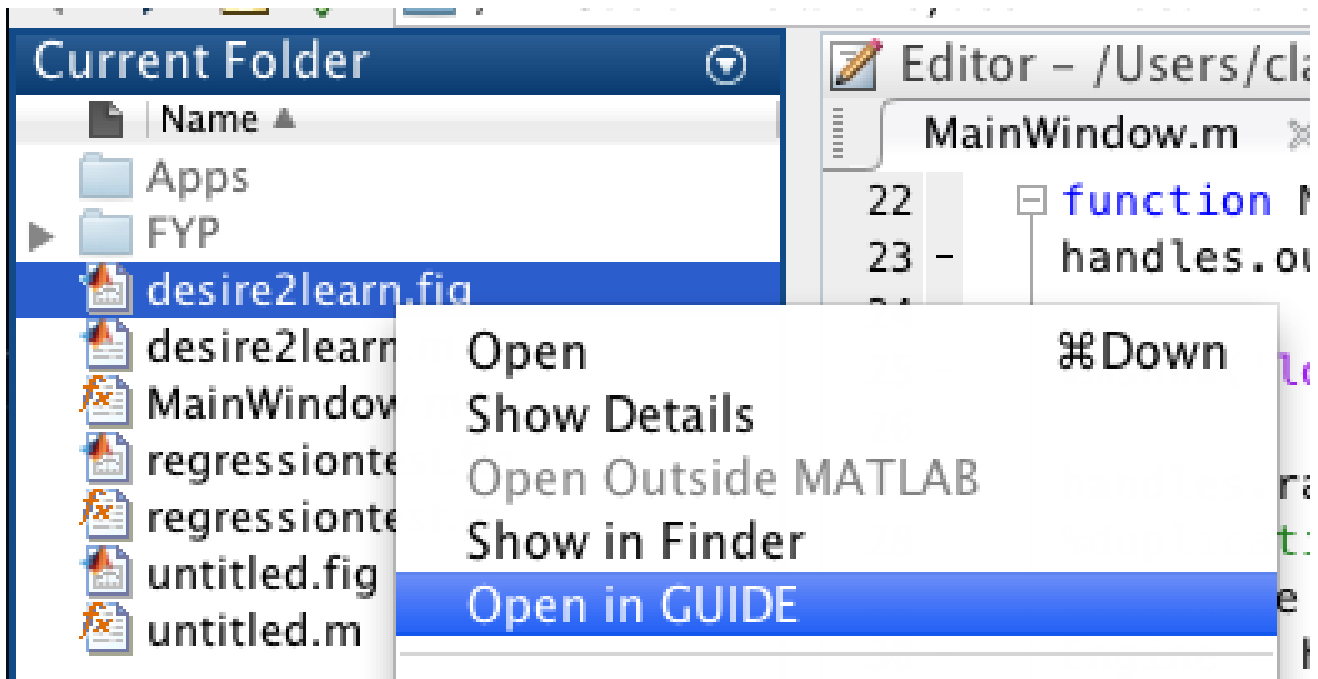
The heap size determines how much memory is available for Java objects. Increasing this value may help you avoid "OutOfMemoryError: Java heap space" errors.



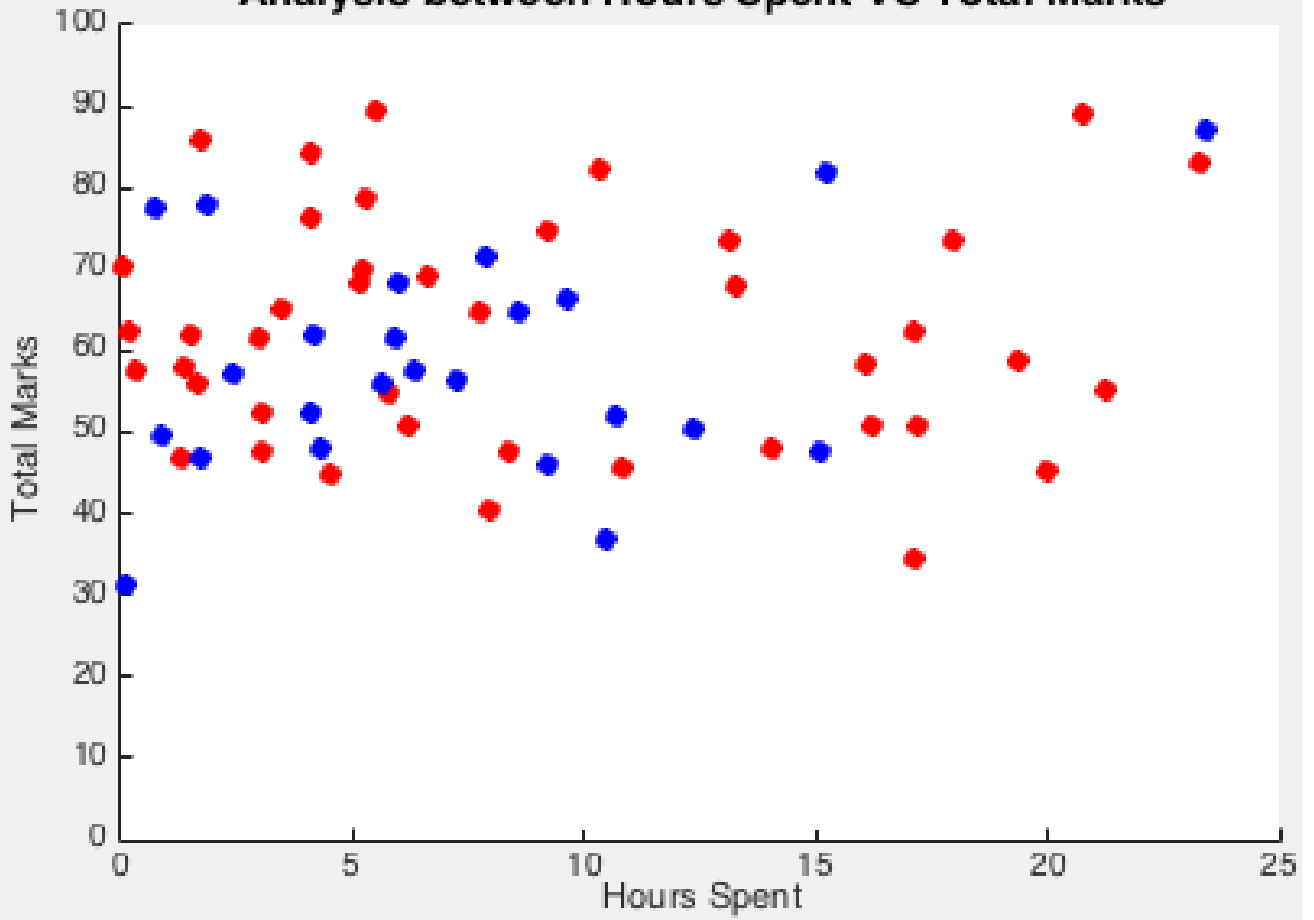
Current value 512 MB. Default value 512 MB.



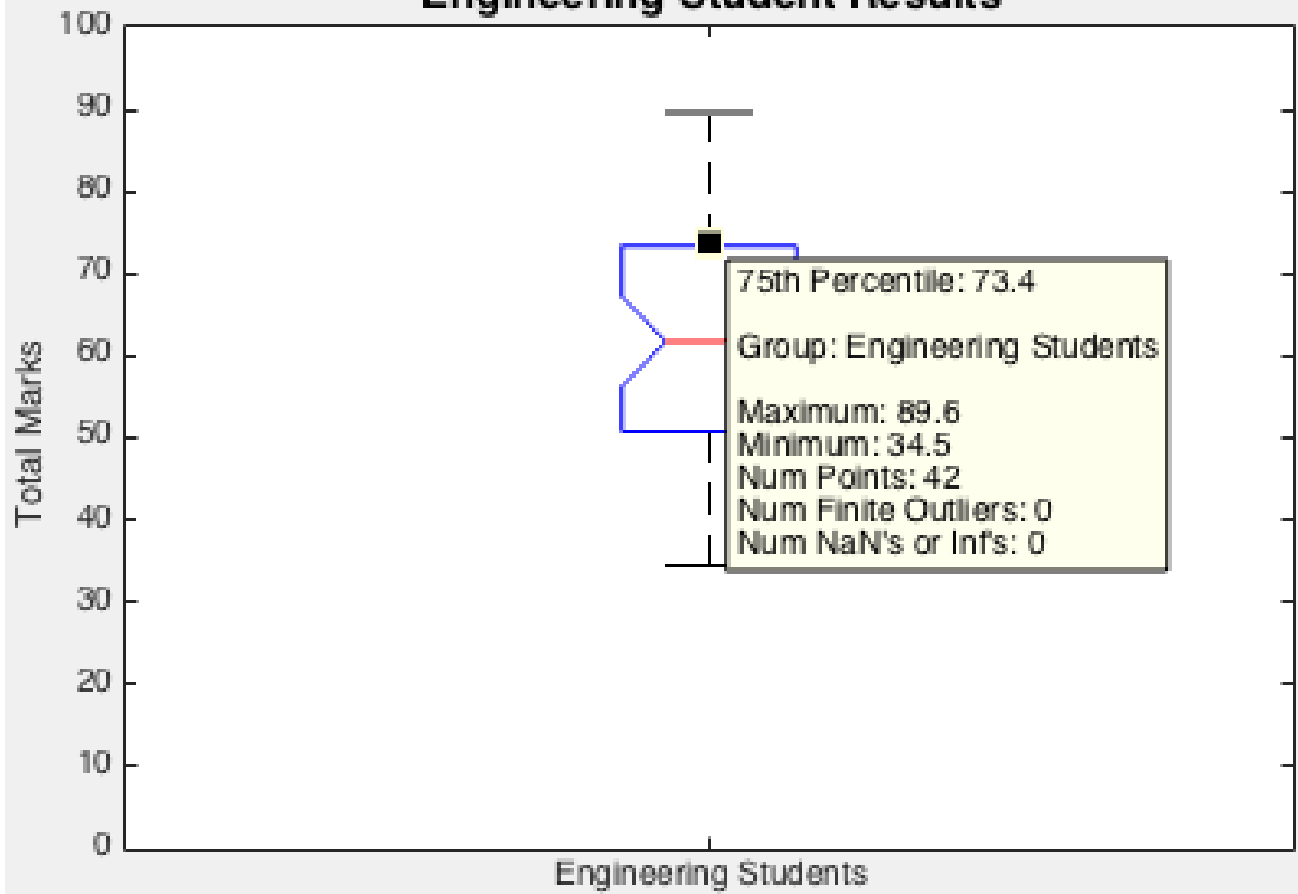
You must restart MATLAB for the above change to take effect.



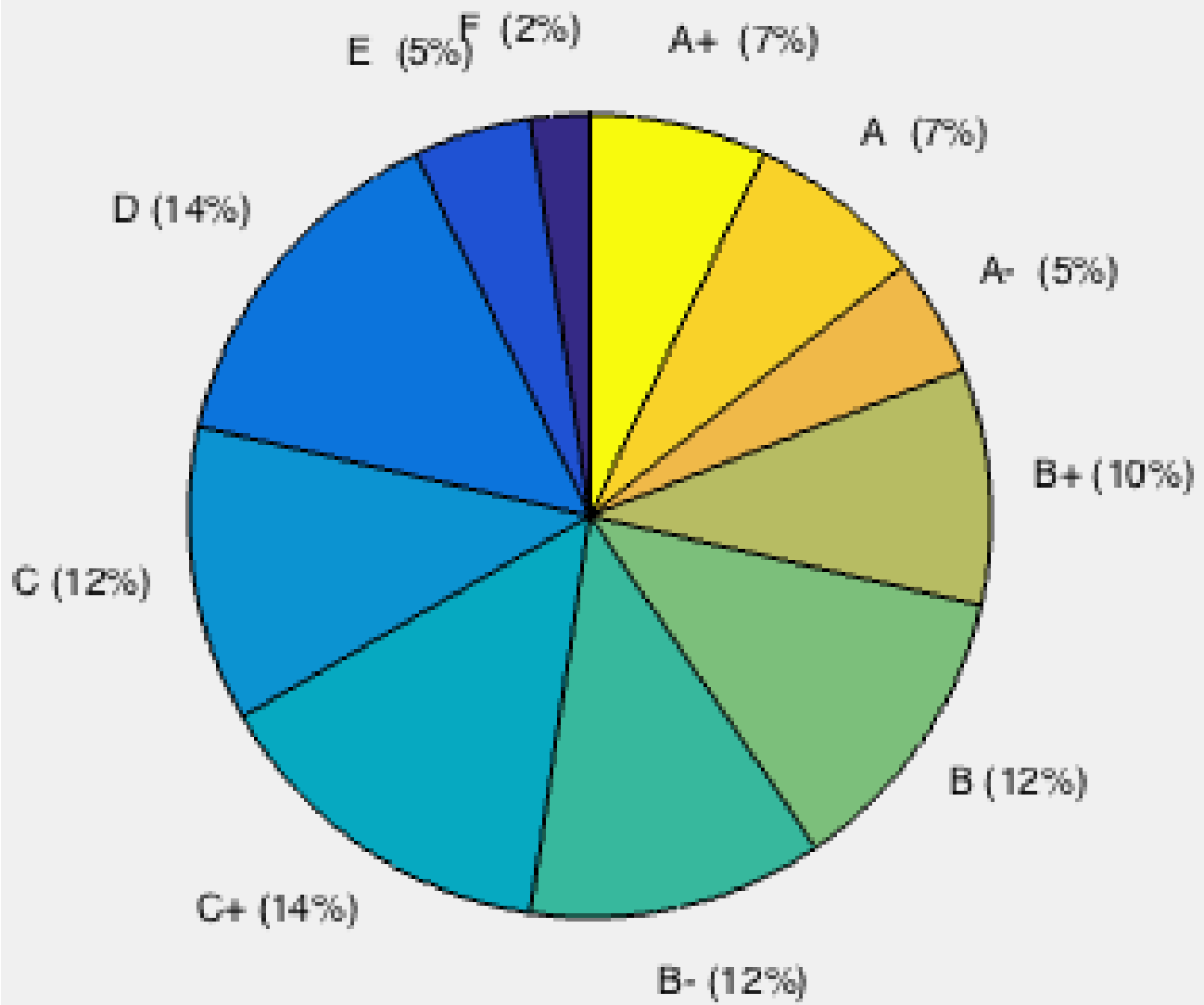
Analysis between Hours Spent VS Total Marks



Engineering Student Results



Engineering Students

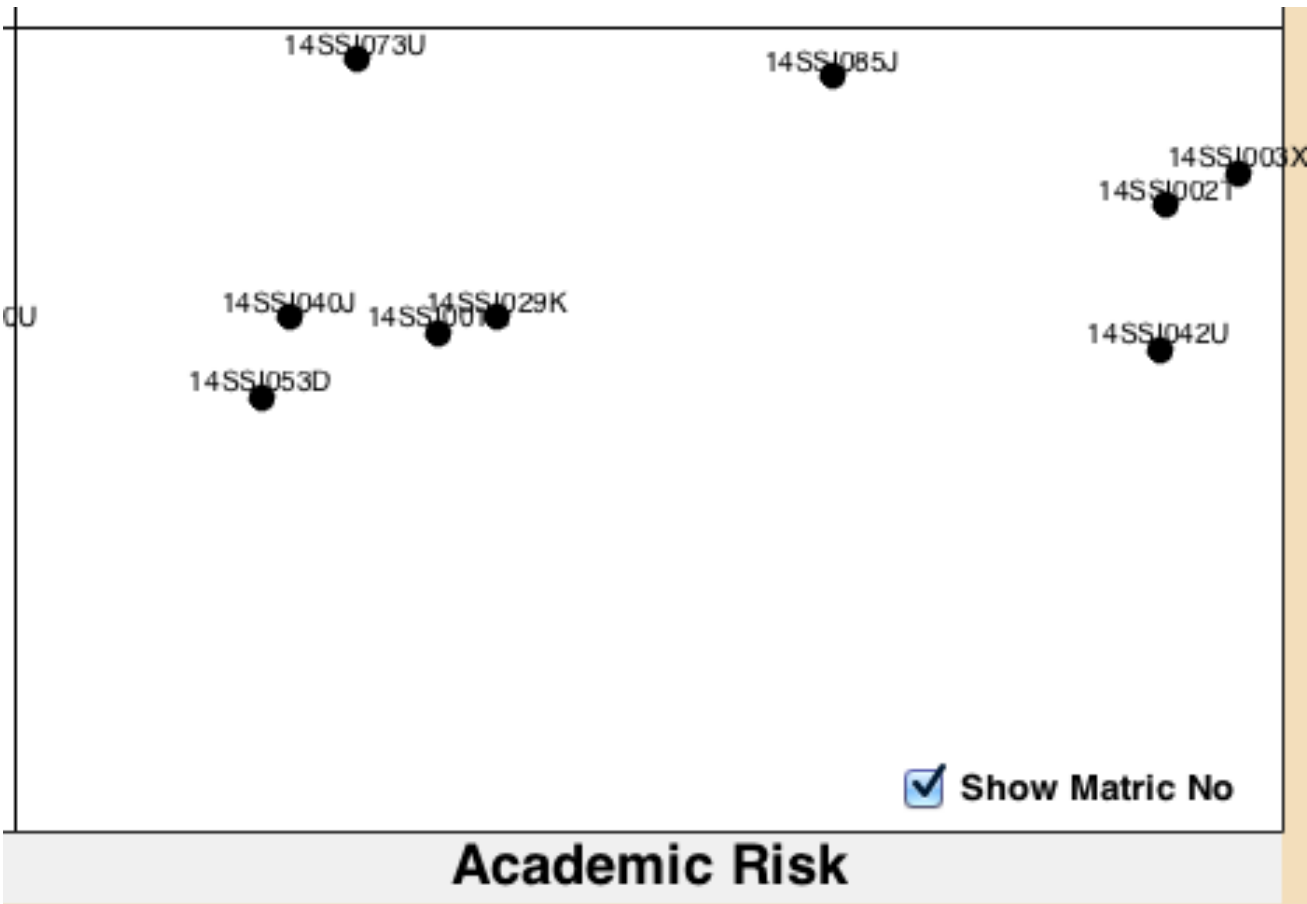
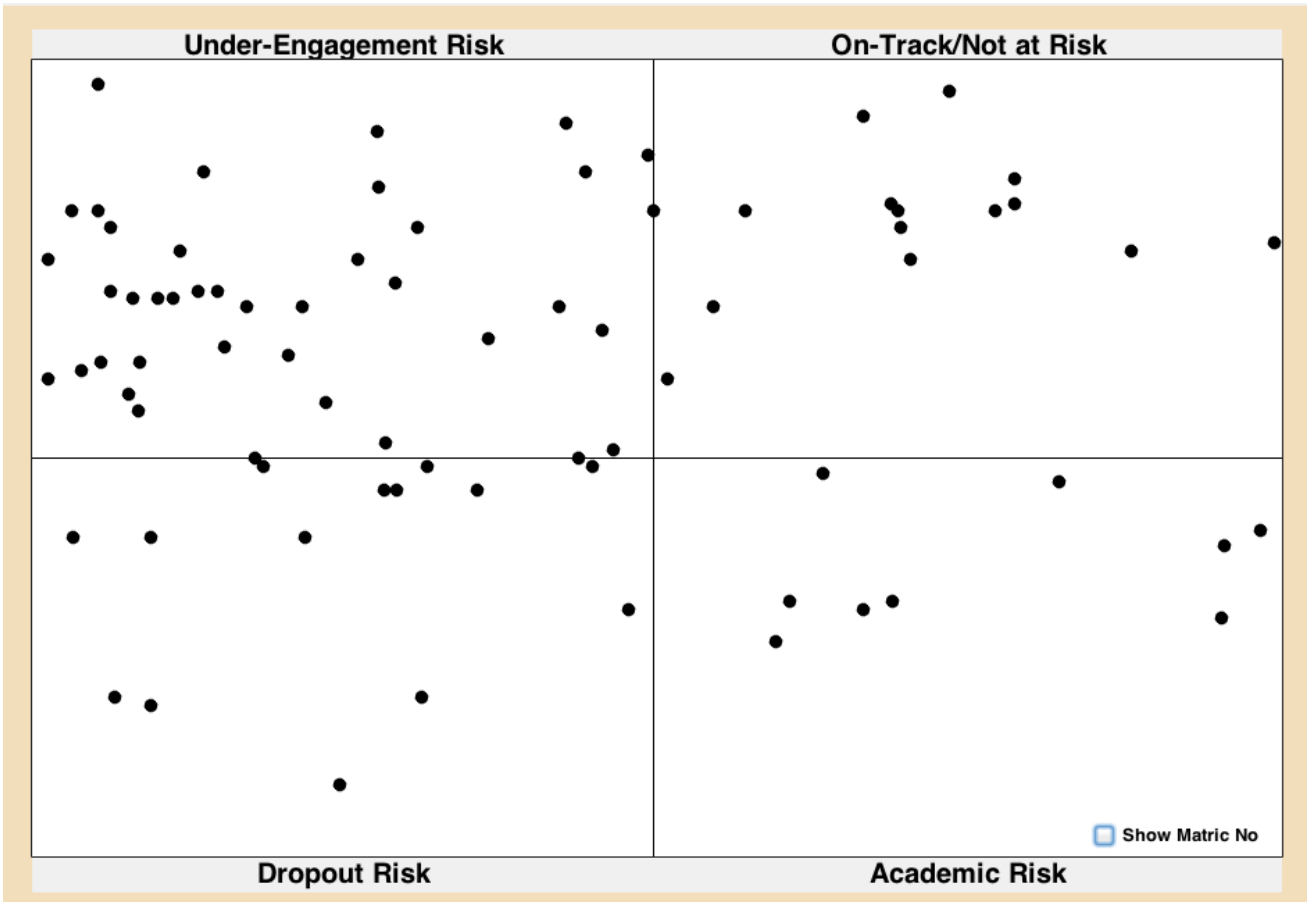


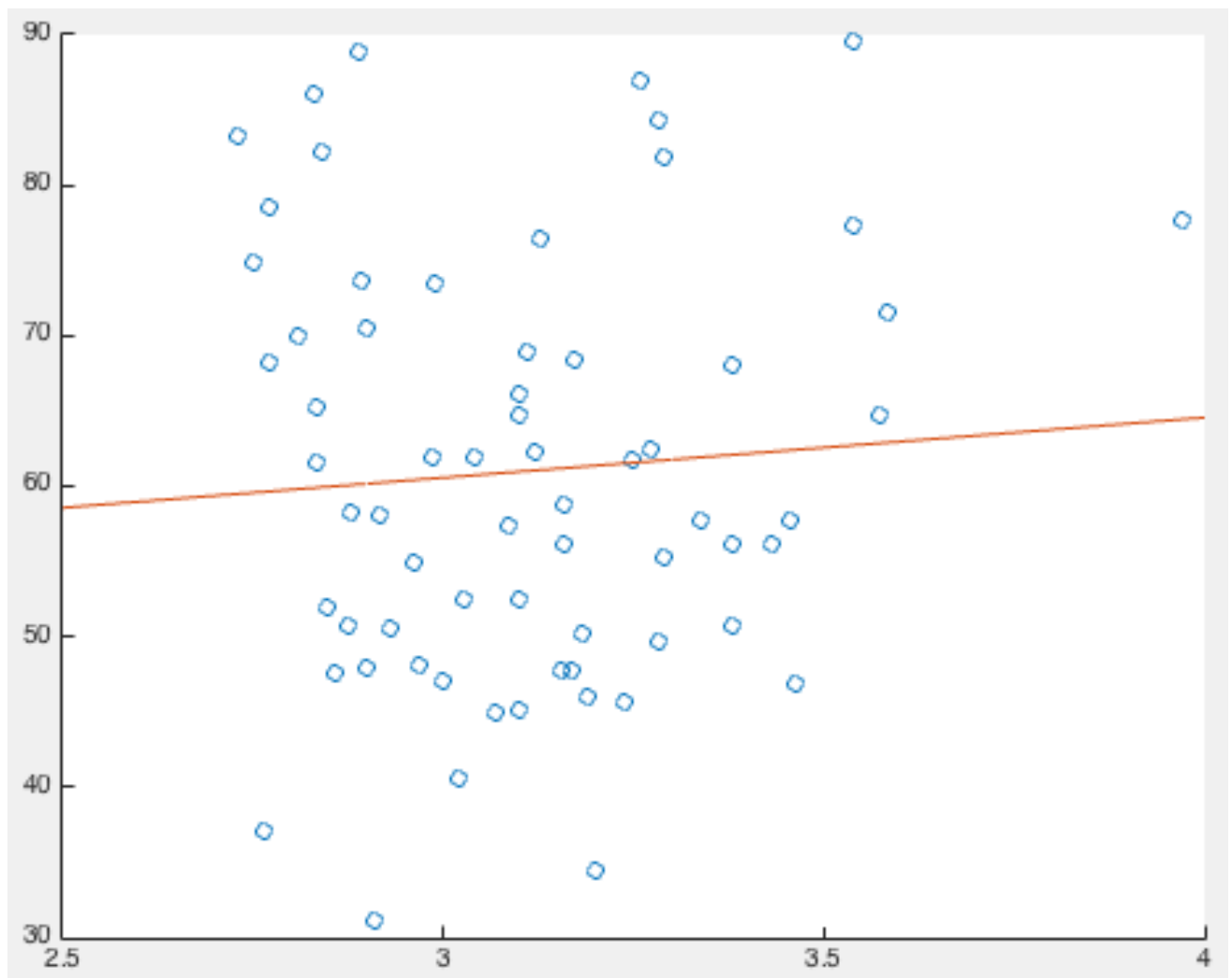
Overall Achievement

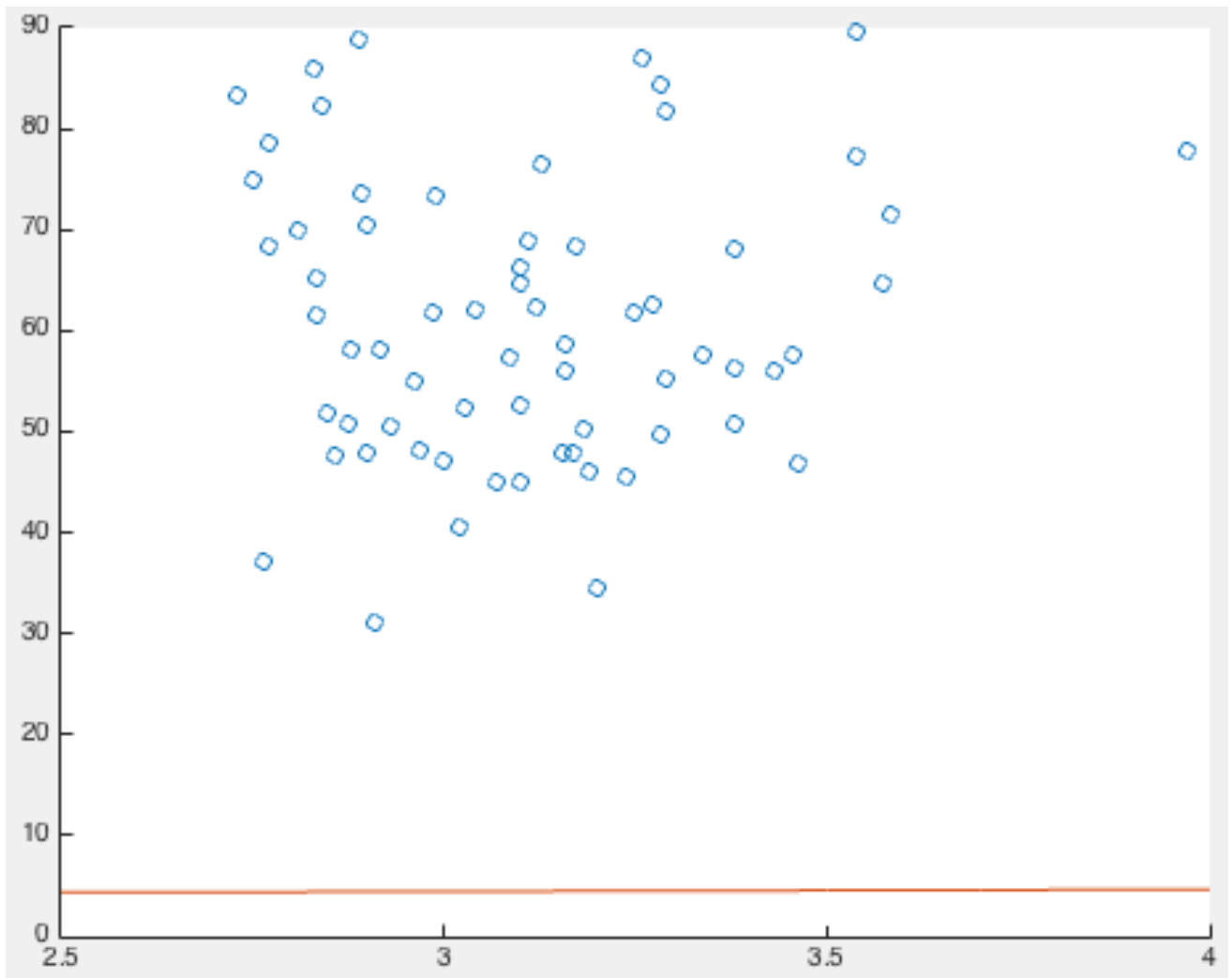
Total Marks (%): 45.6

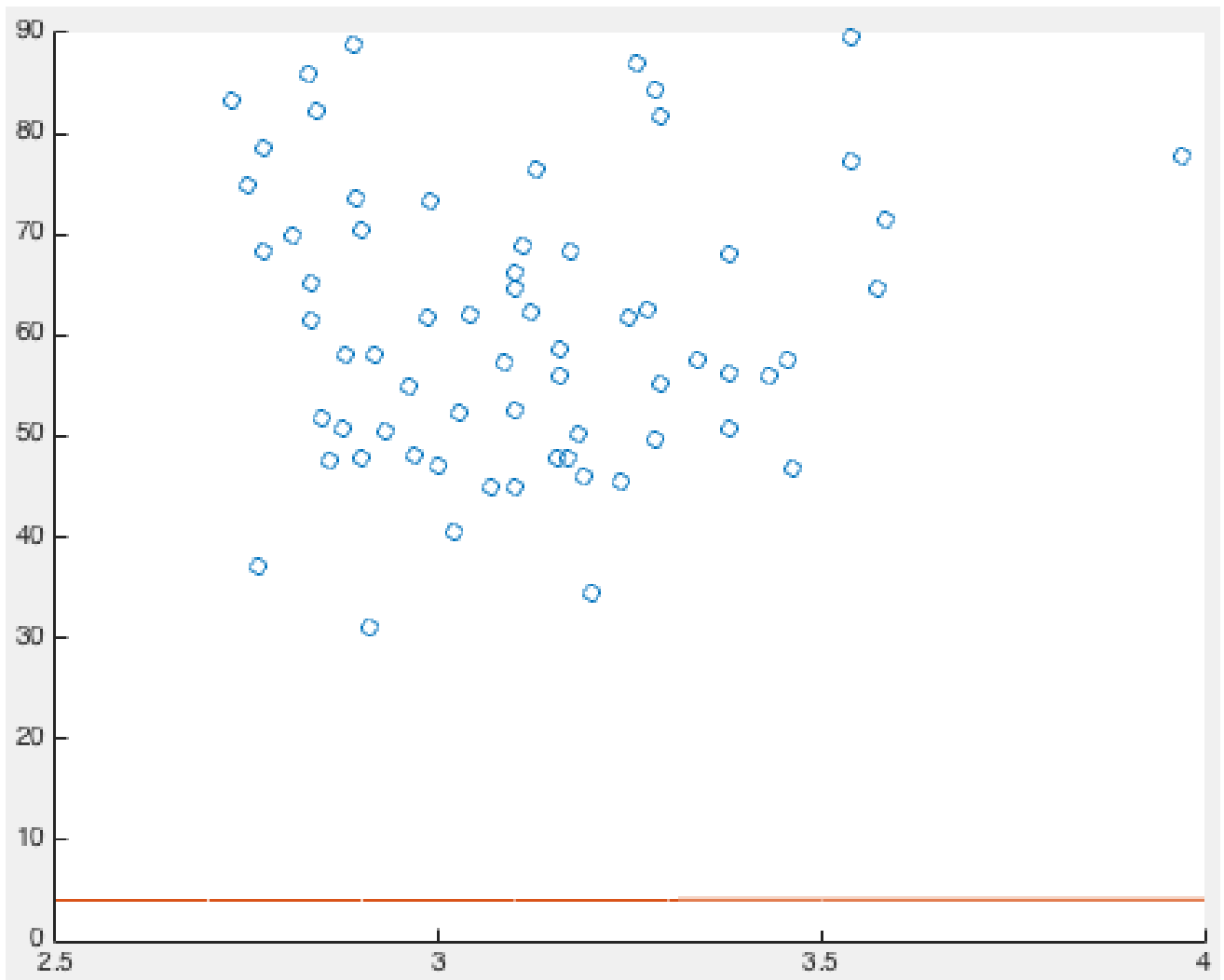
Grade Indicator: Weak

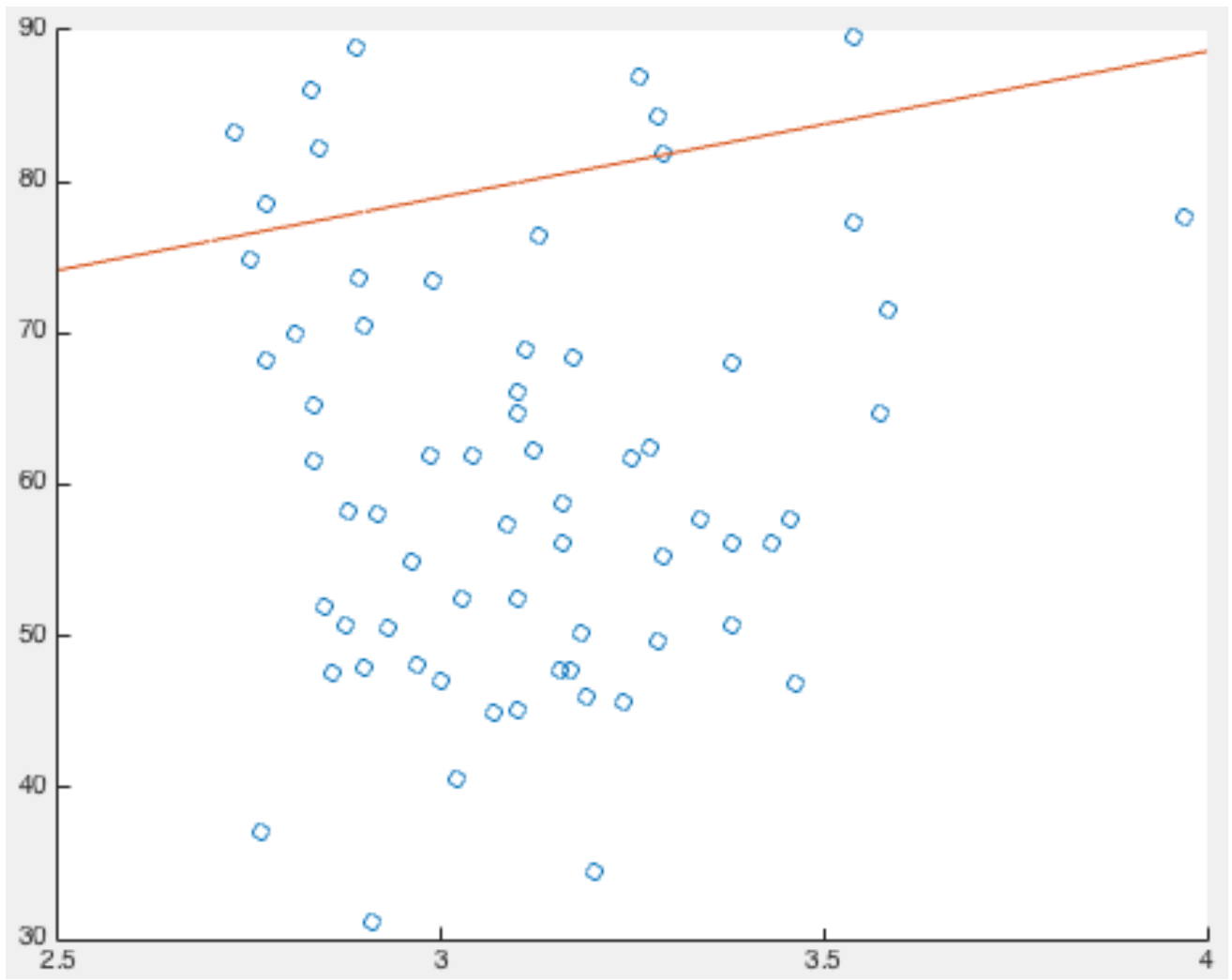
Class Position: 60 / 66

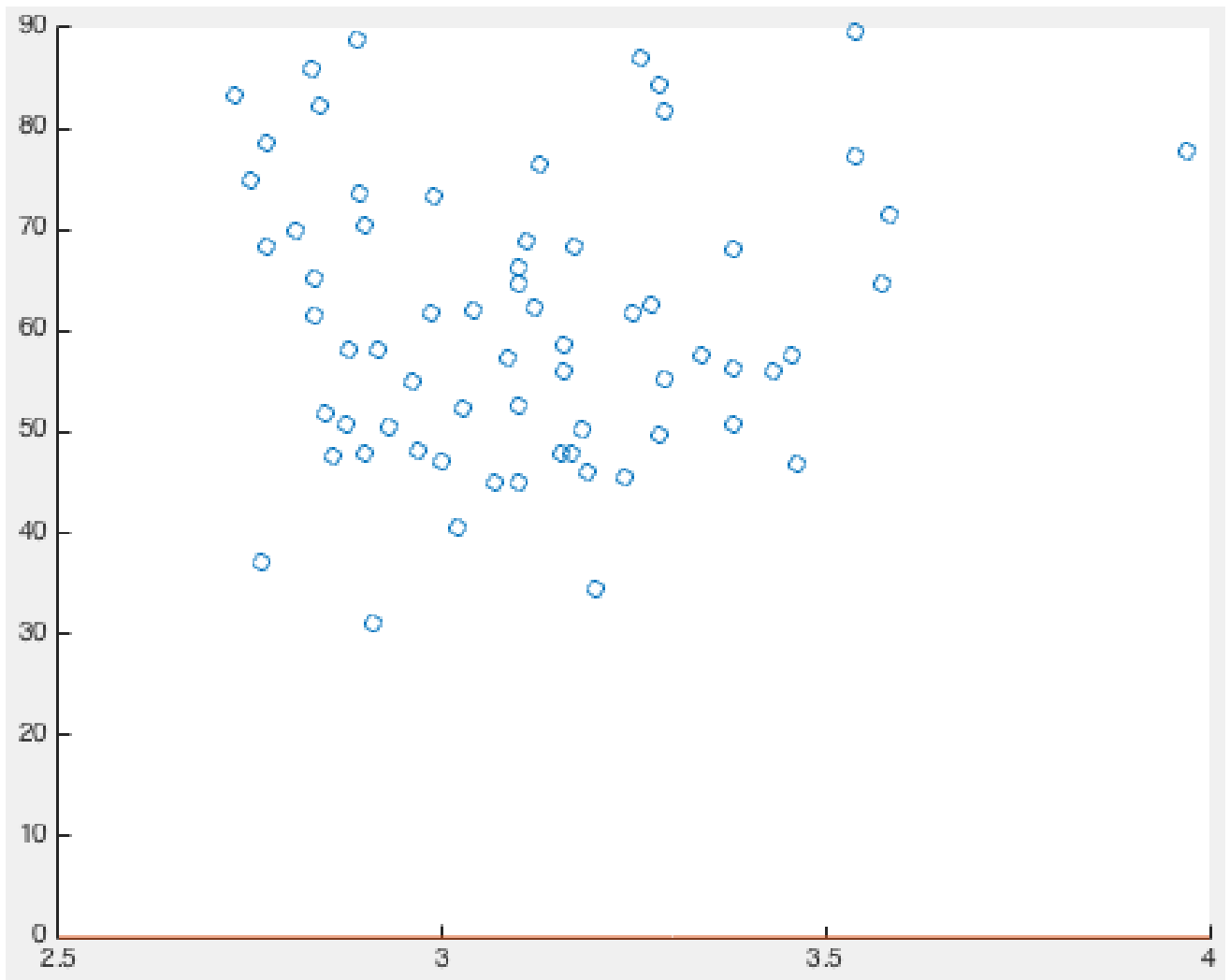


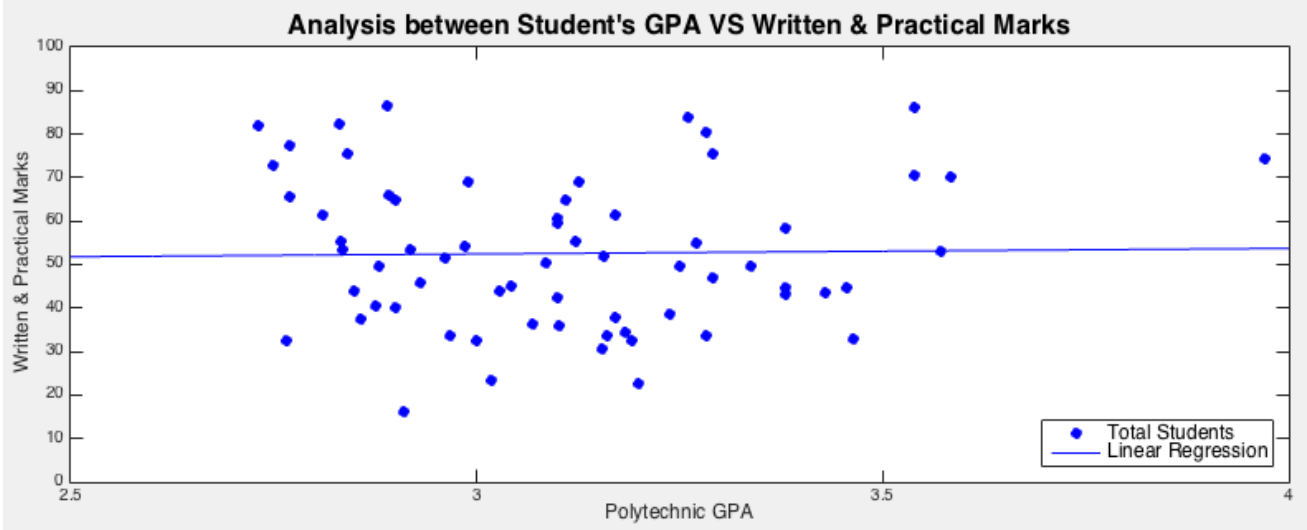
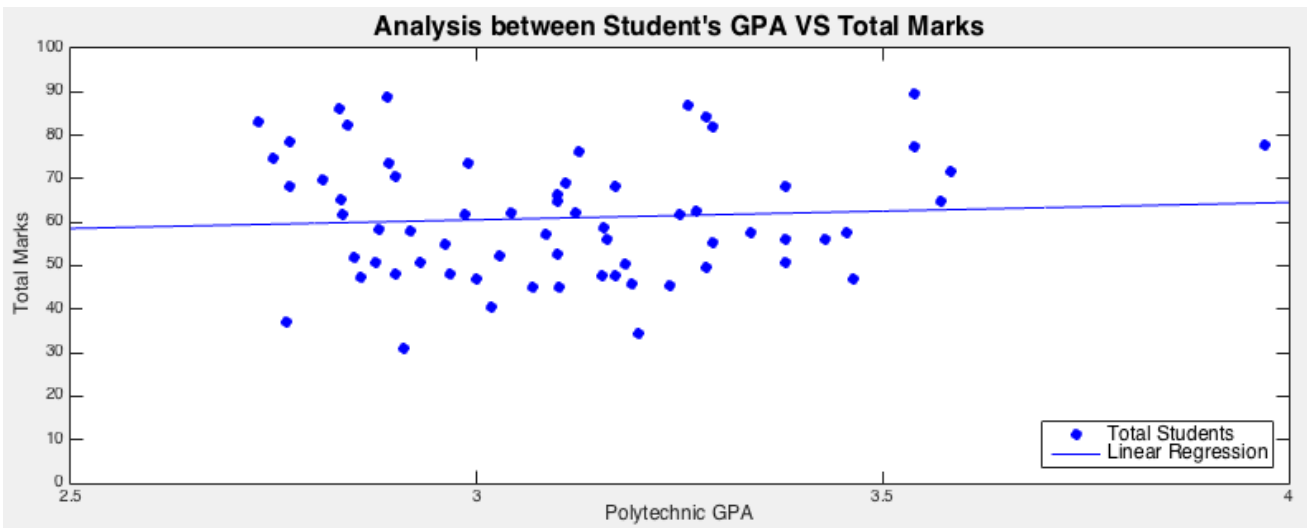












Enter GPA:

3.6

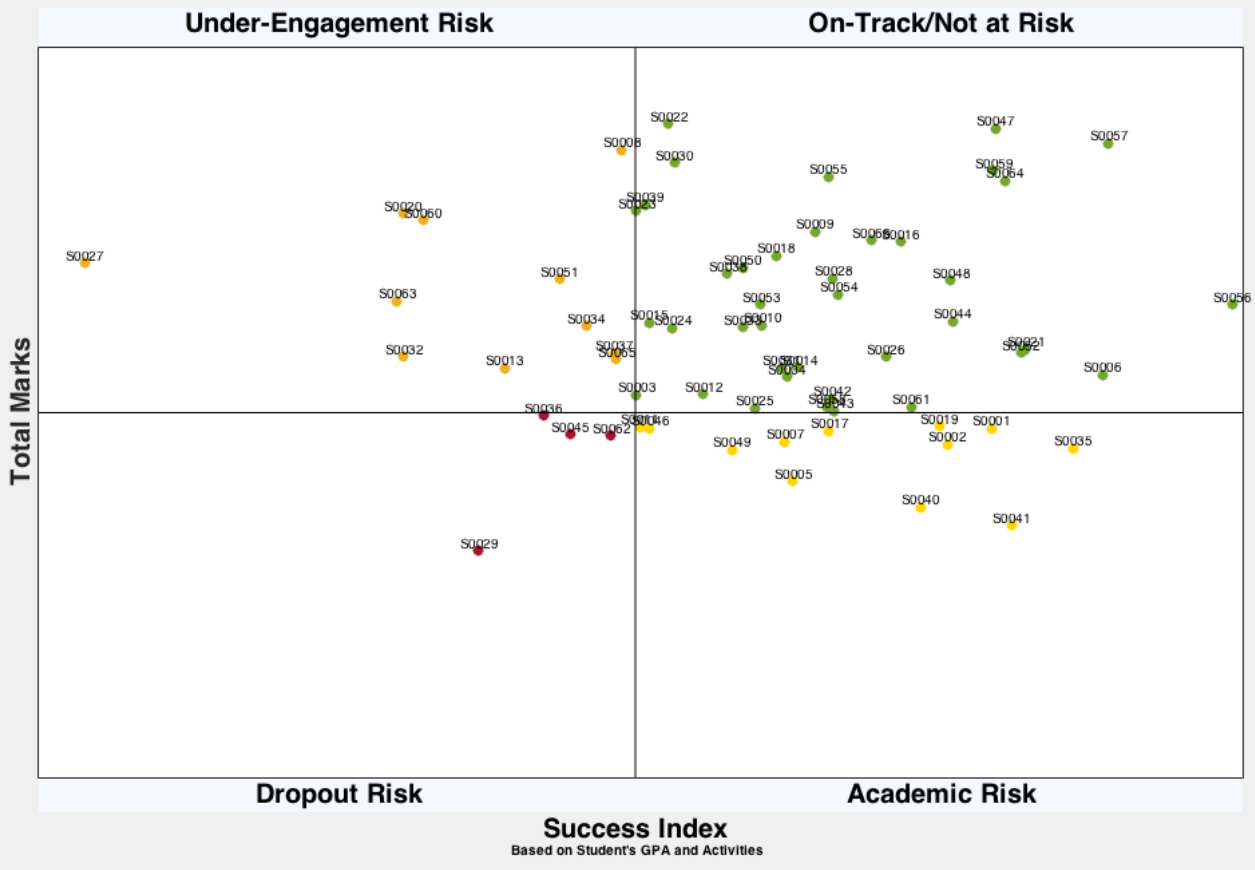
Calculate

Expected Total Marks:

62.96

Expected Written & Practical Marks:

53.21



Under-Engagement Risk

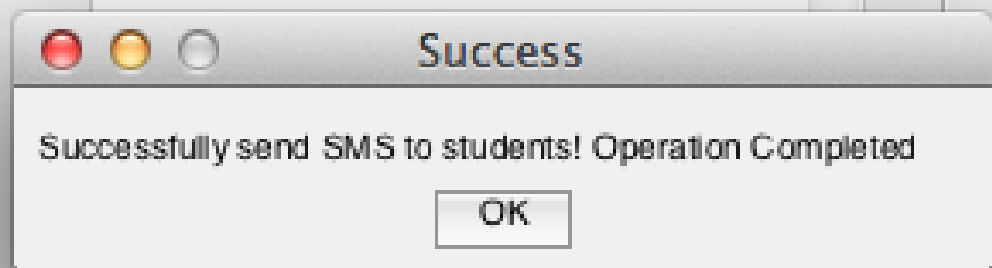


Send SMS

Number: 9177_____

Message:

You have a new notification from
LEARNING ANALYTICS. Please
check email for more information.



Send SMS

 Messages

Details

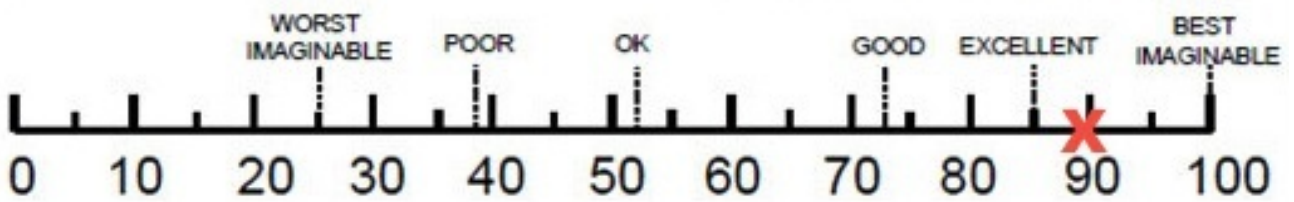
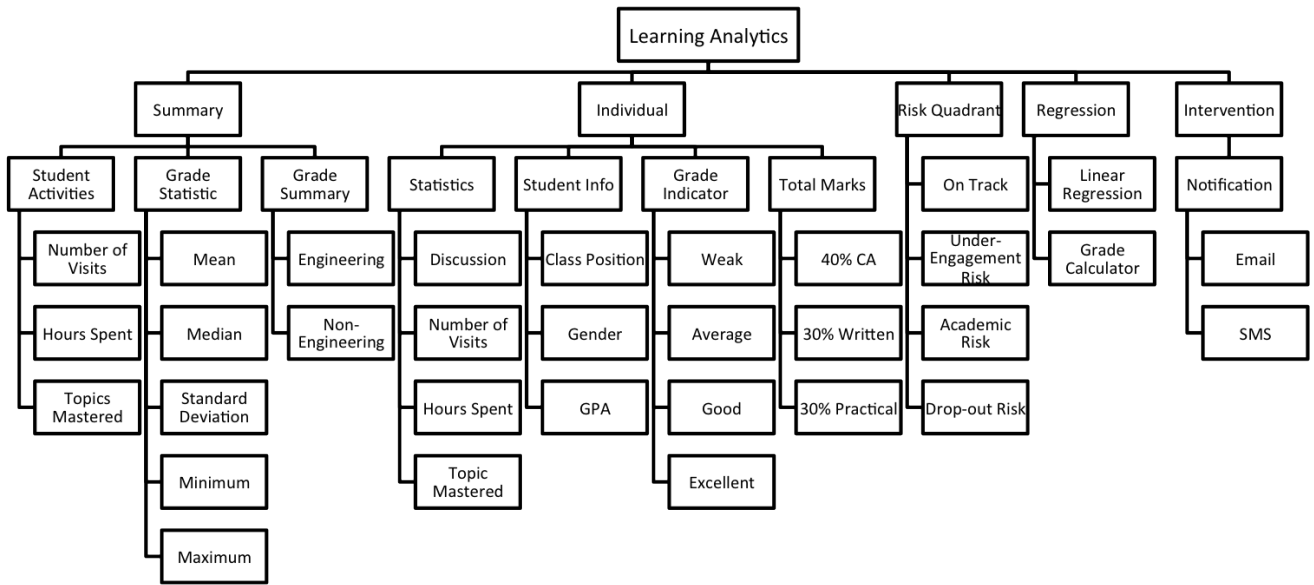
Text Message
Today 7:17 pm

You have a new notification
from LEARNING ANALYTICS.
Please check email for more
information.



Text Message

Send



Appendix F

Ethics Checklist

**School of Computing Science
University of Glasgow**

Ethics checklist for 3rd year, 4th year, MSci, MRes, and taught MSc projects

This form is only applicable for projects that use other people ('participants') for the collection of information, typically in getting comments about a system or a system design, getting information about how a system could be used, or evaluating a working system.

If no other people have been involved in the collection of information, then you do not need to complete this form.

If your evaluation does not comply with any one or more of the points below, please submit an ethics approval form to the Department Ethics Committee.

If your evaluation does comply with all the points below, please sign this form and submit it with your project.

-
1. Participants were not exposed to any risks greater than those encountered in their normal working life.
Investigators have a responsibility to protect participants from physical and mental harm during the investigation. The risk of harm must be no greater than in ordinary life. Areas of potential risk that require ethical approval include, but are not limited to, investigations that occur outside usual laboratory areas, or that require participant mobility (e.g. walking, running, use of public transport), unusual or repetitive activity or movement, that use sensory deprivation (e.g. ear plugs or blindfolds), bright or flashing lights, loud or disorienting noises, smell, taste, vibration, or force feedback
 2. The experimental materials were paper-based, or comprised software running on standard hardware.
Participants should not be exposed to any risks associated with the use of non-standard equipment: anything other than pen-and-paper, standard PCs, mobile phones, and PDAs is considered non-standard.
 3. All participants explicitly stated that they agreed to take part, and that their data could be used in the project.
If the results of the evaluation are likely to be used beyond the term of the project (for example, the software is to be deployed, or the data is to be published), then signed consent is necessary. A separate consent form should be signed by each participant.

Otherwise, verbal consent is sufficient, and should be explicitly requested in the introductory script.
 4. No incentives were offered to the participants.
The payment of participants must not be used to induce them to risk harm beyond that which they risk without payment in their normal lifestyle.

5. No information about the evaluation or materials was intentionally withheld from the participants.
Withholding information or misleading participants is unacceptable if participants are likely to object or show unease when debriefed.
6. No participant was under the age of 16.
Parental consent is required for participants under the age of 16.
7. No participant has an impairment that may limit their understanding or communication.
Additional consent is required for participants with impairments.
8. Neither I nor my supervisor is in a position of authority or influence over any of the participants.
A position of authority or influence over any participant must not be allowed to pressurise participants to take part in, or remain in, any experiment.
9. All participants were informed that they could withdraw at any time.
All participants have the right to withdraw at any time during the investigation. They should be told this in the introductory script.
10. All participants have been informed of my contact details.
All participants must be able to contact the investigator after the investigation. They should be given the details of both student and module co-ordinator or supervisor as part of the debriefing.
11. The evaluation was discussed with all the participants at the end of the session, and all participants had the opportunity to ask questions.
The student must provide the participants with sufficient information in the debriefing to enable them to understand the nature of the investigation.
12. All the data collected from the participants is stored in an anonymous form.
All participant data (hard-copy and soft-copy) should be stored securely, and in anonymous form.

Project title Design & Develop of a Learning Analytic toolkit

Student's Name Clara Kertyn Ke Yue'e

Student's Registration Number 2109935K

Student's Signature 

Supervisor's Signature 

Date 24/03/2015