



# **ANLY482 – Analytics Practicum Project Proposal Group 15 - Edufy**



**Heng Kok Chin**

**Peh Zhan Hao**

**Tan Yong Kiong, Alson**

**Professor:**

**Kam Tin Seong**

# Contents

Project Summary.....	3
Team Introduction .....	3
Project Sponsor Information.....	3
Sponsor & Liaison Information .....	4
Project Background & Motivation .....	4
Project Objectives .....	4
Literature Review.....	5
Data Collection & Description.....	6
Scope of Project.....	6
Proposed Methodology .....	7
Recommendations & Insights .....	8
Stakeholders .....	8
Project Management .....	9
Deliverables .....	10
References .....	10

## Project Summary

For many Secondary 2 students in Singapore, choosing and selecting a subject combination for the next two years leading to GCE 'O' Level examinations can be a tough decision. It is also difficult for teachers to decide whether or not to encourage or let students take on the combination of triple, double or Combined Science. Should schools stream students based on overall subject grades, or should they base their decisions on students' individual Science grades? Often, many parents want and feel that their children are able to qualify for double or triple Science subject combination. Without proper analytical evidence, it makes it difficult for teachers to convince parents that the recommended subject combination would be a better choice for their child.

## Team Introduction

We are Team Edufy. Our team consist of:

- a. Heng Kok Chin. Year 4 student from School of Information Systems. 2<sup>nd</sup> major in Analytics. Proficient in Javascript (AngularJS, D3.js), Tableau, SAS Enterprise Miner and Guide. Basic in R.
- b. Peh Zhan Hao. Year 4 student from Lee Kong Chian School of Business. 2<sup>nd</sup> major in Analytics. Proficient in SAS Enterprise Guide and Miner, HTML, CSS, Javascript, MySQL, Adobe Photoshop and Illustrator. Basic in PHP and Laravel Framework.
- c. Tan Yong Kiong. Year 4 student from Lee Kong Chian School of Business. 2<sup>nd</sup> major in Analytics. Proficient in Tableau, JMP Pro, SAS Enterprise Miner, D3.js, Adobe Photoshop and Illustrator. Basic in Python.

## Project Sponsor Information

Our project sponsor, Edgefield Secondary School (ESS), is a neighbourhood secondary school located in the North-East region of Singapore to provide quality education to students living around the estate. Equipped with the newest facilities and the latest technologies coupled with curriculum innovation, the school is committed to provide the ideal learning environment and experiences for its students. It is currently in the process of setting up a Data Analytics Team to tackle educational problems faced by teachers and students.

## Sponsor & Liaison Information

Name	Position	Contact Information
Mr Lee Peck Ping	Principal, Edgefield Secondary School	6883 9511 LEE_Peck_Ping@schools.gov.sg
Mrs Wong Puay Kheng	Vice-Principal, Edgefield Secondary School	6883 9511 LIM_Puay_Kheng@schools.gov.sg
Mdm Candice Ngau Shu Mei	Head of Department (Mathematics), Edgefield Secondary School	9118 8404 ngau_shu_mei_candice@moe.edu.sg

## Project Background & Motivation

GCE 'O' Levels is a major examination in many students' education life in Singapore. The exams will heavily determine the paths that are available to students in the future. Choosing a subject combination for their GCE 'O' Levels will subsequently affect the future career paths of a student, be it whether the students ends up entering into a Polytechnic or a Junior College. For example, if a student wants to study Pure Biology in a Junior College in the future, studying either Pure Biology or Combined Biology for GCE 'O' Levels is necessary. Also, to take into account that there is a lower entry requirement for Pure Science as compared to Combined Science (e.g. to study Biology in Junior College, you might need to just pass Pure Biology in GCE 'O' Levels, but maybe at least a B4 for Combined Biology in GCE 'O' Levels).

We want to come up with an analysis that will help the parents and students in choosing a subject that is both manageable (considering the student's capability) and desirable (for the future of the student). We want to avoid the regret later where a student's future path is limited by his or her poor subject combination back in their secondary school.

## Project Objectives

Utilizing past data of students' grades from the school's database, we would like to discover useful and practical insights which will allow teachers to better decide and advise students on choosing their Secondary 2 subject combination, particularly on whether they should take

Combined, Double or Triple Sciences. We will attempt to analyze the trends of students' academic performance by examining their past subject grades and subject combinations.

To achieve the above mentioned, we will perform an in-depth analysis on the historical data with the following aims:

1. To help secondary schools and teachers better formulate the right streaming practices and criteria that would benefit all students
2. To develop an application using R for the school so that they can input future data to improve the accuracy of the model in predicting students' GCE 'O' Level examinations results

## Literature Review

Since the introduction of the "New Education System" by Deputy Prime Minister Goh Keng Swee in 1979, the practice of streaming students based on their academic performance has been a vital part of Singapore's education. This is reflected by the various major examinations that a student has to partake in - from PSLE, Secondary 2 streaming examination to the GCE 'N', 'O' and 'A' Level examinations. The streaming of students based on their academic performance has been controversial and often contested, as each examination is likely to impact a student's future academic progression and may favour or disfavour a certain group of students.

Various research papers have discussed about the effects of streaming, which is multi-dimensional. Johnston & Wildy (2016) mentioned that a streaming system affects not only the academic learning outcome of students, but also in social and psychological aspects. A study conducted by Tanggaard, Nielsen and Jørgensen (2015) showed that different streaming practices adopted by different schools results in very different experiences for students, which affects the engagement and social inclusion that a student would receive. Hence, schools, as well as teachers, have a huge responsibility in the formulation of clear goals and requirements in streaming. The dilemma that many schools and teachers face is how to aptly adopt and establish the right streaming practices to improve the learning outcomes of students.

This dilemma is not just limited to other overseas countries, but also pertinent in Singapore. For many Secondary 2 students in Singapore, choosing and selecting a subject combination for the next two years leading to GCE 'O' Level examinations can be a tough decision. It is also difficult for teachers to decide whether or not to encourage or let students take on the

combination of double Science or Combined Science. Should schools stream students based on overall subject grades, or should they base their decisions on students' individual Science grades? Often, many parents want and feel that their children are able to qualify for double or triple Science subject combination. Without proper analytical evidence, it makes it difficult for teachers to convince parents that the recommended subject combination would be a better choice for their child.

While various research papers have focused on the effects of streaming on students, little have discussed about how schools and teachers can accurately formulate the right streaming practices and criteria that would benefit all students. As such, we will propose an analytical model to shed light on a more scientific and data-driven approach for schools to formulate better streaming practices.

## Data Collection & Description

Column Name	Description
student_id	The unique id of the student
subject_id	The unique subject id of the course
year	The year in which the data was recorded
level	The academic level which the student is in
class	The class which the student is in
type_of_examination	The various types of examinations (e.g. PSLE, CA1, SA1, CA2, SA2, 'O' Level)
score	The score obtained for a particular examination

## Scope of Project

As we have yet to obtain the complete data from our project sponsor, we will only share our initial plans on how intend to tackle the project. We intend to adopt the following steps in our analysis:

1. Data Collection
2. Data Preparation
3. Exploratory Data Analysis
4. Data Cleaning

5. Data Normalization and Transformation
6. Data Modeling
7. Re-train Prescriptive Model
8. Recommendations and Insights

## Proposed Methodology

### **1. Data Collection**

We will use the data provided to us by our project sponsor in the form of Microsoft Excel.

### **2. Data Preparation**

As the data provided by the sponsor has many columns, we will attempt to organize the data into consistent formats that are easier for our analysis. Additionally, we plan to further split the data tables into tables that are more readable. We will also need to ensure that the data does not contain any confidential information such as the names of the students.

### **3. Exploratory Data Analysis**

We will look into the scores of each type of examination based on the secondary levels of students. From here, we will be able to understand the ability of students and analyze the trends in the scores together with their subject combinations. Additionally, we will go into observing the impact of teachers and CCAs on the student's' examination scores.

### **4. Data Cleaning**

To ensure accuracy of our model, we will identify missing values and outliers that are observed during the Exploratory Data Analysis. We will go through these missing values individually and decide on how we should handle it (whether we should estimate by taking the average or simply remove the entire row). As for handling outliers, we will try to analyze and come up with a reason for the outlier and see if it will affect our analysis.

### **5. Data Normalisation and Transformation**

To better cater the data to our needs, we will perform data transformation to transform some of the columns into rows and transforming between categorical and numerical variables so that we can better analyze the data. If the values in certain

attributes varies too much, we will normalize these attributes to ensure that the analysis will be accurate.

#### **6. Data Modeling (Steps 6-8)**

We will develop an analytical software application using R Markdown or R Shiny to visualize the impact of the ability and subject combination of students to the final GCE 'O' Levels results.

## Recommendations & Insights

To be completed after our analysis.

## Stakeholders

The main stakeholders of this project include:

- Project Supervisor
  - Prof Kam Tin Seong, Associate Professor of Information Systems; Senior Advisor, SIS (Programme in Analytics)
- Sponsor
  - Mr Lee Peck Ping, Principal of Edgefield Secondary School (ESS)
  - Mdm Lim, Vice-Principal of ESS
- Students of ESS
  - Teachers and Heads of Department (HODs) of ESS
  - Parents of students studying in ESS



# Project Management

Tasks	Week 0 (Before 2/1)	Week 1 (2/1 - 8/1)	Week 2 (8/1 - 15/1)	Week 3 (16/1 - 22/1)	Week 4 (23/1 - 29/1)	Week 5 (30/1 - 5/2)	Week 6 (6/2 - 12/2)	Week 7 (13/2 - 19/2)	Week 8 (20/2 - 26/2)	Week 9 (27/2 - 5/3)	Week 10 (6/3 - 12/3)	Week 11 (13/3 - 19/3)	Week 12 (20/3 - 26/3)	Week 13 (27/3 - 2/4)	Week 14 (3/4 - 9/4)	Week 15 (10/4 - 16/4)	Week 16 (17/4 - 23/4)
Project Preparation	Understanding Sponsor Requirements	A, K, Z															
	Project Acceptance	A, K, Z															
	Create and Update Wikipedia Proposal in Word Document & Wikipedia	A, K, Z															
<b>Milestone 1: Project Proposal Submission</b>																	
Data Gathering & Scoping		A, K, Z															
Research & Preparation	Finalize Requirements With Sponsor		A, K, Z														
	Explore Software		A, K, Z														
Data Cleaning	Finalize Proposal		A, K, Z														
	Data Collection		A, K, Z														
	Data Cleaning		A, K, Z														
Data Modelling	Data Transformation					A, K, Z											
	Stage 1: Exploratory Analysis																
	Stage 2: Clustering																
Interim Preparation	Stage 3: Time Series Analysis																
	Gather Feedback From Sponsor																
	Prepare Interim Report & Slides																
<b>Milestone 2: Interim Report &amp; Presentation</b>																	
Application Building																	
Iteration	Code Application																
	Test Application																
Final Preparation	Gather Feedback From Sponsor																
	Adjust Analysis																
	Refine Results To Improve Clarity																
<b>Milestone 3: Abstract &amp; Full Paper Submission</b>																	
Final Preparation	Prepare Research Paper																
	Prepare Poster																
<b>Milestone 4: Final Presentation &amp; Poster Submission</b>																	

Legend
A- Alison
K- Koh, Chin
Z- Zhen Hao
Planned

## Deliverables

The final deliverables of this project are as following:

- Project Proposal
- Mid-term Report
- Mid-term Presentation
- Final Report
- Final Presentation
- Project Wiki Page
- Project Poster
- Recommendations to Sponsor

## References

Tanggaard, Lene, Nielsen, Klaus, & Jørgensen, Christian Helms. (2015). Students' Experiences of Ability-Based Streaming in Vocational Education. *Education & Training*, 57(7), 723-737.

Johnston, O., & Wildy, H. (2016). The effects of streaming in the secondary school on learning outcomes for Australian students – A review of the international literature. *Australian Journal of Education*, 60(1), 42-59.