Singapore Management University
ANLY482 Analytics Practicum

Supervisor Minutes 3 as on 6th February 2017

Time Start:	2.35pm
Time End:	3.40pm
Location:	SIS Meeting Room 4-6
Recorded by:	Heng Kok Chin
Vetted By:	Peh Zhan Hao

Attendees:	
Dr Kam Tin Seong	Associate Professor of Information Systems (Practice)
Heng Kok Chin	Undergraduate, Singapore Management University
Peh Zhan Hao	Undergraduate, Singapore Management University
Tan Yong Kiong, Alson	Undergraduate, Singapore Management University

Agenda
1. **Update Supervisor Regarding Our Dataset**
2. **Supervisor's Advice**
3. **Administrative Matters**
4. **Set Agenda for Next Meeting**

No.	Discussion:	Action by:	Deadline:
1	**Update Supervisor Regarding Our Dataset**		
	• Prof. Kam mentioned that some schools stream their Secondary 1 students based on PSLE results and their Secondary 2 students based on Secondary 1 results • He suggested that we check with our sponsor on how they stream their Secondary 1 students • The team excluded the results for other subjects and focus mainly on the Math and Science results as the school gave us the criteria that is taken account into when they stream students • Prof. Kam mentioned that it is not a wise idea to exclude the data without first exploring them. Some students are being dragged down by other subjects (e.g. foreign students who are forced to take Mother Tongue like Mandarin) • We should confirm that the school's current streaming system and criteria is either working well or there is some loop hole; we don't want to assume that their system is working well • English must be included because it is compulsory and an important subject • Zhan Hao asked Prof. Kam on how to quantify the letter grades results for PSLE and 'O' Levels? • Prof. Kam explained that the PSLE and 'O' Levels letter grade results is what we call independent variables (these are all categorical results) and we can make it as a dummy variable	Edufy	13th February 2017

	• We can also use <u>Latent Class Analysis</u> (with JMP Pro) to analyze it • Make use of the L1R4 and L1R5 to know the overall performance (those who get 4 or 6 is definitely better than those who get 8) • By right, we should also exclude CCA score (because it minus from L1R4 or L1R5) and use only the raw academic score for 'O' Levels • <mark>For our missing data in Secondary 4 CA1/CA2, we need to confirm whether it is due to Prelims or what</mark> • Our data structure is not considered correct in Tableau, the second row should be the first row to have the correct header • Need to tidy up, reorganize the data by rename the columns (PSLE_SC, PSLE_MA etc.) and decide what kind of data structure we want to adopt • End result of the data should all be in one sheet • If we append it down and make it long with many rows, it is easier to compare column by column • If we need to compare horizontally, retain it the current way (can compare across) • Zhan Hao asked that if a student performed well in CA1 and CA2 is very hard and the student didn't do well and affect his score, should we take into account the percentile or the overall score change? • Prof. Kam replied that in general, some school do moderation, some school report back the actual score, we need to take a look at it • Certain year, if they are setting tougher questions, then the score will shift over but usually the HOD will ensure moderation • We should think from the analysis perspective, we need to clean the data and finalize the data structure • Let's say Secondary 1 and Secondary 2 each have five classes, is there any significant difference between their performance? • If there is significant performance issue, then there is problem and we need to find out first • Some classes will perform better and some will not perform so well • How well can we compare how well their model is? • Based on their current system, we can use the school examination results as a respond variable to help us do the separation • We are looking at how they stream their students • We take the sec 2 results as the response variable then use the different SA/CA results as separation • First we do a predictive model (cluster analysis) on the combination of results that they have • Separation, from what score to what score go which combination • Based on that separation, we compare the 'O' Levels performance • For PSLE, need to monitor the English (it can mean something to the students' progress), there are two extremes, do badly in English but other subjects pull up or do well in English but other subjects didn't	Edufy	13th February 2017

		do so well • For the categorical variable (like the letter grades A, B, C), we CANNOT assign a median score • Do the <u>Latent Class Analysis</u> to see what kind of combination they have, segment the students based on their PSLE • The PSLE score is weird (by right should be transferred, should be complete and correct) what is the possibility that this kind of error happen • The team should quickly clean and run some statistical analysis and see if there are any discrepancies		
2	**Supervisor's Advice**			
		• When we perform the EDA, we want to examine the distribution; do a <u>histogram</u> and <u>boxplot</u> to see the distribution • Tableau is good for quick view but not good for in-depth statistics, the team should look into using JMP Pro • For example, first of all, look at PSLE, assume we look at the overall score, we want to know whether the school already streamed them into their classes when they first get into Secondary 1 • If they stream already, then the histogram should be curved to the left, already sorted by score. What we want to compare is whether this is a good way to prepare them for the streaming for Secondary 3 • Do the same for the Secondary 1 to Secondary 2 results, how it leads into their Secondary 3 streaming • Each subject and each assessment need to have a boxplot • Determine that they are consistent and get a good view on how the grades are being used		
3	**Administrative Matters**			
		• Interim presentation is in Week 8 • By then, we should have the dataset nicely cleaned, finished EDA • We should focus on one batch of students only if there is time limitation • It is best that we meet with sponsor and get their feedback before interim presentation • NDA is now signed by Prof. Kam and the team members • <mark>Get it sign by the sponsor</mark> • If we require Tableau, we can apply for the educational license using our SMU credentials	Edufy	13th February 2017
4	**Set Agenda for Next Meeting**			
		1. Finalize data structure 2. Plan timeline to prepare for interim - Hope to meet sponsor at least twice before interim (one time to get all the missing data from them and another time to present to them our findings from EDA)		

| | | 3. Seek Prof. Kam's advice before presenting findings to sponsor | | |