Time: 11.40am – 12noon	-
Location: Level 4 Meeting Room 4-1	
Attended By: Prof Kam, Janice, Hui Shia and Siong Min	
Absent with Apologies: -	

No	Agenda:	Action By:
1.	Prof Kam mentioned that we cannot based on title and description to determine if it is a sustainability job. For example, a person job's title can be Business Analyst but in the job description, he could be doing things that help to protect the environment.	
2.	If we want to map the region, we would need to have a new column for us to map.	
3.	Look at the dataset and find out: 1. The distribution of all the jobs – do not focus on region wise as it could be too big. How about Asian pacific? In the US, not all states are environmental sensitive. Therefore, regional is not appropriate. Analyse the distribution on smaller geographical instead.	
4.	If we start looking at job scope & description, we can go into more details.	
5.	Regarding the dataset, if we were to focus on trend, date would not matter. Trend is more likely to be possible with about 5 years of data.	
6.	Recommendation from Prof: Devise a content categorization or intelligent database that allows for automation of the classification process. For example, to classify them this particular person with his/her job title and description, and then determine if he/she is environmental friendly or not.	
7.	Apply text mining tactics to examine the job description and summary. With content classification mentioned above, use this content database and train the text mining algorithm. It should be "smart" enough to know how to group accordingly.	
8.	Look into each description (read 5000 records manually) and we will be able to find out more stories. Do not focus on correlation as the team is not ready for it. We need to use the description to find out the existing patterns. Using a java program may be risky as keywords may not help.	
9.	Do not run the risk of overstating or understating. For example, Sg may have 100,000 army man and Indonesia has 100,000 as well. However, they are not the same due to the population size.	
10.	We need to figure out a base (labour force?)	
11.	Additional info: In many Scandinavian countries and selected European countries, they have implemented series of act such that if they conform to certain sustainability requirements, they are entitled to tax exemption or tax rebates.	

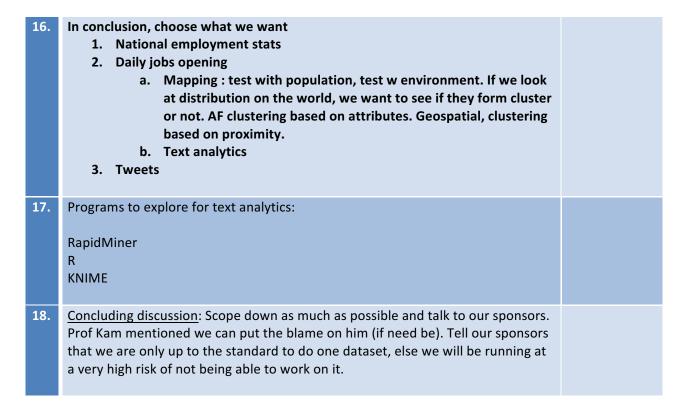| 12.	ISO14001: Prof Kam mentioned that the team can build an investigation to show it. However, it will be a totally different topic because we need to base on the country level and then look at the past last years (e.g. how many of them have adopted this particular act/policy, or any endorsement? After which, use regression model and check if it is statistically significant.) - International Labour Office Website – get international employment patterns (one of UN organisation) and hopefully we can find these kind of data. However, do note that it is totally different from the jobs opening data. Hui Shia asked if we were to use the jobs data to rank the countries that has the most job openings and then go into the country and find out why is this so, would it be appropriate? The main purpose is to try to gain insights into this dataset. Prof Kam's Reply: it will open up a lot of questions. This is job openings but not real jobs. Hence, it is not encouraged to work on that.	

13.	Prof Kam mentioned that we would have to find new data if we want to answer the questions that GTL wants.
	   - We need a nationwide employment data
	   - Need to have an entry called sustainable.
	   - May look at GDP growth, environmental, and then we find out if environmental is significant or not. Does it increase the employment sector?
	   - IF we want such explanation, then the jobs data are too weak. The jobs data are just good for exploratory, such as patterns.
	   - We could base on the environmental sensitivity of the company.
	
	Hui Shia asked if indicators are beneficial to this dataset and Prof Kam acknowledged that it is.	

14.	Concerns about grades: We would not be getting below B+ if we filter everything out properly. The main concern is for us to learn the process (such as the process of text mining). Prof Kam wants us to experience the constraints of text mining.	
15.	Discussion about Tweets data: Prof Kam states that we only need to use one dataset (either daily jobs opening or twitter data). This is because both processes is the same. "We have to let someone to grade us and the person who grade us are more concerned whether we gain good experience on working on some analytics fields. They don't want us to cover 1000 stuff but its very shallow."	

16.	**In conclusion, choose what we want**	
	1. **National employment stats**	
	2. **Daily jobs opening**	
	a. **Mapping : test with population, test w environment. If we look at distribution on the world, we want to see if they form cluster or not. AF clustering based on attributes. Geospatial, clustering based on proximity.**	
	b. **Text analytics**	
	3. **Tweets**	
17.	Programs to explore for text analytics: RapidMiner R KNIME	
18.	<u>Concluding discussion</u>: Scope down as much as possible and talk to our sponsors. Prof Kam mentioned we can put the blame on him (if need be). Tell our sponsors that we are only up to the standard to do one dataset, else we will be running at a very high risk of not being able to work on it.	

Prepared by: Janice Koh