Date: 26/1/2017
Time: 1530-1730
Venue: SIS Project 3.2

Attendees: Thavanesan s/o Sivananthan
Wang Tian Tong
Gauri Bhatnagar
Prof. Kam Tin Seong

Agenda:

1. To present the slides we have prepared for our client (NLB) to prof Kam, regarding the progress of the previous group in te project as well as to run a demo to show how the application functions.

The following were the clarifications made by prof:

1	<u>Library collection data</u> There is no point in pursuing such data as all libraries are meant to have the same type of collections except for the national library which carries special collections for other reasons (research library and has 2 private schools)
2	<u>Huff's model</u> The group's huff model is very limited. We should also be looking at additional factors for the attractiveness index such as: 1) The users that go to a particular library 2) The propensity they go to the particular library 3) Transport cost and distance from location to library The concept of propensity in this case is that in one year, the frequencies of visits to different libraries varies for every patron. Hence, we need to use empirical data regarding every particular patron's library visits. And fit it into the model instead. Eventually we need to use the MCI model to derive the attractiveness index and decide the significance of each variable. Also in Huff's model the probability never changes. If we do what-if analysis, whether you want to change the profile of a library or you want to build a new library (take the average of all the other libraries), you use huff's model to recalculate the value
3	<u>Library types</u> There are 3 types of libraries - home base, workplace base and school base (school usually have their own libraries except for private schools) 1. School based - Primary school students tend to go to the public libraries for hang out, as they can borrow books from their school libraries itself 2. Home-based - There has been no such data in the previous model, that's why we need the datasets regarding HDB blocks and condominiums (should be able to cover 80% of housing data in Singapore)

	For MBS and orchard areas, they may become outliers because there are no neighborhoods around
4	Map generation We can use leaflet R to generate the map, and use ggplot in R to plot choropleth map. We should also look into (leaflet) ggmap & tmap in R for the map generation
5	Patrons Need to identify who is our entertainer (target), and based on their profiling, we decide what can be the key variables that determine the patronage. It is important to note that there are people going to the library without borrowing books, they just kill time. There is a need to note the difference between people who just visit the library and those who borrow books. But we don't have the data for those who just visit the libraries without borrowing books, in this project we only look at people who borrow books, which builds up to 60% of the users of library.
6	RFM analysis RFM analysis has no much problems at all, but better to provide recommendations in terms of how we can utilize the results i.e. proper data visualizations for clusters
7	Requirements from application Crawling, geocoding, data abstraction and preparation are expected to be done using the application.
8	Use of Centrality indexes There is a need to decide which of the centrality indexes we want to use. Suggetions for use of indexes: MRT (centrality index, and betweeness index) and bus stations (centrality index for all the bus routes) near the particular library. e.g. for a library, for a bus station near the library, how many buses are reaching there
9	Application based clarifications 1. Overlay of patrons is is not necessary because it will affect the results. If it is present in the app then one should be able to toggle and remove it. 2. We should be dealing with site profile analysis not retail area analysis. 3. Bar chart on the number of mrt, bus stop, childcare centre etc → that's what they want

	4. Buffer should be from 500m onwards. Once we select then it should take all the layers in and summarize the count of the data 5. The data file is currently separated and not in 1 data file. Ideally it should combine all the files into a master layer, with xy coordinate, identifier & 1 column POI type, description etc 6. Change the workflow whereby there's a Masterfile 7. Attractiveness activities and facilities present at a library should be shown 8. The template that the previous group used is wasting a lot of space on the left, so we have to do some formatting of the application 9. Only patron flow should be displayed and not other factors like malls for example. Also, 1. % values of how many patrons there are coming from different destinations for any given library should be shown 2. Patron flow should be the FIRST TAB, then PATRON SEGMENTATION → SITE PROFILE ANALYSIS 10. Patrons segmentation 1. Under optimal cluster → if there is no meaningful cluster, then it should not show the ones you don't need. Always start from a k value of 3 for clusters 2. Display box plot showing only the distribution of that cluster, to allow one to interpret the characteristics of the cluster

The meeting was adjourned at 1730 hrs. These minutes will be circulated and adopted if there are no amendments reported in the next 24 hours.

Prepared by, Thavanesan

Vetted and edited by,Thavanesan