Analytics Practicum Supervisor Meeting 01

MEETING CALLED BY	Prof Kam
TYPE OF MEETING	Project Briefing
FACILITATOR	-
NOTE TAKER	Chong Xin
TIMEKEEPER	Chong Xin
ATTENDEES	Chong Xin, Bowei, Hui Min

Agenda topics

1620 - 1630	CHECKING DELIVERABLE REQUIREMENTS 1	ALL MEMBERS

DISCUSSION	Do we need to preload the factors first and allow users to select?
CONCLUSIONS	- The system should have an uploading function so that users can upload factors as and when they need (because there are unlimited factors). - However, we need to upload some factors to form the base of the project. The UI is for them to use to add something extra if required.

ACTION ITEMS	PERSON RESPONSIBLE	DEADLINE
-	-	-

1630 - 1640	CHECKING DELIVERABLE REQUIREMENTS 2	ALL MEMBERS

DISCUSSION	Should we upload the files into a database? - We may not want a database in the formal sense, as we need a server to contain the database. If we need a server then we need extra IT support, hence Prof does not recommend having a database. - In other case, Prof recommends we come up with a new architecture/design, so that we can present to the sponsor to see what they prefer. In most organizations, we will need a temporary license & testing machine to work with their databases, hence we want to highlight when we present our architecture. - It doesn't mean we cannot use database, we can use file-based database or loose data files, so that we can catch everything in a JSON format and write everything in a JSON file. It will be stuck inside our computer but it is not consuming extra space. Alternatively, there is a file-based database, a unique database system, meaning that we do not need data server to do any installations. It is an interesting option that many mobile and web apps are using. Government agencies definitely cannot use open-cloud databases. - Recommends powerful file-based database called SQLite (open-sourced), geospatial support called spatiaLite (open-sourced). - Output will be a desktop based web-app. SQLite does not have model building library hence we need to use R library. A communication exists between R and SQLite. - Prof recommends we explore big data approaches, e.g. Apache Spark (open-sourced) – processing advantage to work with and allow us to experiment big data analytics environment. Considering learning outcome if we select certain approaches to work with.
CONCLUSIONS	- Decide on the technologies & architecture to be used

ACTION ITEMS	PERSON RESPONSIBLE	DEADLINE
Decide on the technologies & architecture to be used	All members	Sponsor Meeting 01

1640 - 1700	CHECKING DELIVERABLE REQUIREMENTS 3	ALL MEMBERS

DISCUSSION	- To calculate the attractive index, we need to go through an iterative process so that it will converge to the empirical data at a certain significant level. - Look at big data engines such as Apache Spark. If we put in a database, it will take a long time before we see any results.

	- Look at the data, this has nothing to do with the geospatial thing, but we can understand the borrowing pattern from different patrons and for different libraries. The other aspect of looking at the data is to use the data to construct the model, use it as a base, but congregate them out into unique users (for each user, which library do they patronize?) - Processing is to be done by the system; so we need to upload the raw data into the system and it fits into the model. - We can look at the transaction data and perform market segmentation using the RFM approach – translate to library context: recency, frequency and monetary. - What is the objective of NLB? No fixed objectives, so that they can test any operational objectives. E.g. if they move to a new library, what is the new catchment area, because of the floor space and the attractiveness index. E.g. If I remove a library, what will happen; If I add a new library, what will happen. - Northpoint Library as an example of simulation, put in a cost factor? Is it cost-effective to put in a new library? - Attractiveness index is derived from the geospatial computation and compilations of the information from the premises. - Bowei enquired about accessing the installation files for JMP; Prof replies that he will give the necessary installation information via email.
CONCLUSIONS	- Calculate the RFM values for each patron, in the context of library & borrowing - Divide the patrons into different groups and fit into the model one group at a time - The web-app needs to cater to NLB operational needs, to include basic functions such as adding and removing a library

ACTION ITEMS	PERSON RESPONSIBLE	DEADLINE
To implement the RFM model and conduct patron segmentation	All members	Team Meeting 02

OBSERVERS	-
SPECIAL NOTES	Next Supervisor Meeting (02) will be tentatively scheduled on 24 August 2016 (Wed), subjected to Prof's availability. All members will present on their findings in the Team Meeting prior.