

Supervisor Meeting 2

Date & Time: 18 Jan 18:00pm - 19:20pm

Venue: SIS MR 4.3

Attendees: Prof. Kam Tin Seong (Supervisor), Wang Sijia, Ren Mengxi, Wang Tianjing

Absentees: Null

Agenda:

1. Confirmed with professor on the weekly meeting slot.
2. Discuss with professor about project details and confirm on the scope.
3. Seek suggestions on what techniques are suitable for our project.

Details:

1. Our group have confirm with the Professor on the weekly meeting schedule from 10:30am to 11:30 am on Wednesday.
2. We presented to prof about our ideas about the library entry analysis project. Prof Kam stated that it may be hard to tell whether the user works in library as individual or group because there are frequent cases where a group of students meeting inside the library but reaching at different time so that they would not tap together. We agreed on that and will consult our sponsor in next sponsor meeting to discuss about potential bias results based on this scenario.
3. We have informed prof Kam that we are still in the process of gathering the data. It would be great if library can provide us more personal information about library users like gender and address, so we can derive more interesting insights from it. Regarding the 'out' data which may help us to determine the duration of each visit, we decide to leave it for next phase since our sponsor is still in progress of exporting the data.
4. We showed the Professor about the raw data we have and briefly went through each variable and shared with him about some issues we had concerns on.
5. Prof Kam gave us suggestions on the missing value issue. (solution only suitable when proportion is not so big)
 - a. Missing program name - category them as another group named 'unknown'.
 - b. Missing graduation year - build own rules to determine alumni
If has graduation year \Rightarrow alumni
Else If (phd and ≥ 7 years) or (master and ≥ 4 year) or (undergraduate ≥ 5 years)
 \Rightarrow alumni
Else \Rightarrow ongoing
6. Most work will be about data exploratory
 - a. Use contingency table analysis to find different pattern of accessing library by school (answer question like are second year students tend to go library in the morning?)
7. Other work should be about segmentation
 - a. Detection of frequent users / early birds

- b. Explore LCA for clustering on categorical data. But not all of our data are categorical. Continuous columns like frequency and time need to be converted (frequency to ordinary, time bin to basket)
- 8. There is not enough data for us to study the occupancy and seat hogging issue because the wifi connection data cannot map with specific users and it is not feasible for us to access the login account information as well.
- 9. He also addressed the importance to delivery our final result with reproducibility which should allow library team to upload annual data into our system and run report automatically.
- 10. R, as one of the most commonly used language in data analysis, can be considered as the main language to code our application in our project. It is not about writing our own algorithm but more on understanding the currently available libraries, reading documents and follow the syntax in order to make fully use of them and apply them into our project scenarios.

Action Plan:

Item	Person in charge	Deadline
Clean up and vet minutes	Mengxi	Jan.19
Update the sponsor with the issue raised during our supervisor meeting and ask for the newly exported data	Sijia	Jan.20
Upload minutes to wiki and elearn	Tianjing	Jan.20
Start data exploratory	All	Jan.25