

Supervisor Meeting 3

Date & Time: 25 Jan 10:30am - 11:30am

Venue: SIS MR 4.1

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

Absentees: Null

Agenda:

1. Update supervisor on last sponsor meeting information
 - a. Requested data
 - i. Received data for first six months in 2016 and Wi-Fi
 - ii. Additional students' personal information is not available
 - b. Update on data preparation progress
 - i. Rename columns
 - ii. Delete outliers
 - iii. Apply rule on graduation year
 - iv. Add new column for hour bin and day of the week
 - v. Change the data types columns

Details:

1. Update supervisor: Additional student information is not available.
2. Present to supervisor different datasets we have (currently two: main gantry entry data and Wi-Fi connection data).
3. Supervisor's suggestion on our log book: We should give detailed information about which column we changed and how did we change it (recode from xxx to xxx, by using recode in column function, screenshot and put in report).
4. Finding incomplete data (currently we found only one record): by using "missing data pattern" in table function. There might be more records with missing data.
5. Finding outlier (dirty data): Currently we define outlier as the records beyond normal operation hours. However, there might be some operation after hours. When write our procedures, we should give clear justification.
6. Change graduation year based on our business rule:
 - a. Remove "AY_".
 - b. Academic year is very tricky; as different program starts in different times of a year. Undergraduate: August. Master: August or January. PhD: can be any time.
 - c. Use JMP to build formula cell, screenshot the formula and put into work log. □ □
7. Use concatenate to join all six tables before any work. (stack is for data with different columns), and get a smaller dataset to work using resample from the whole dataset.
8. Derived field: Graduation Year, Hour, Day of Week.
9. Pattern that we currently found:

- a. There are more business students in the library, but they also have a higher population in the school. We should set a base (checking SMU annual report for reference) and compare the breakdown of students from different schools.
 - b. The peak hour is in the afternoon.
10. For people that entry and exit library several times in a day, define different analysis (no. of entry/ unique no of person) .
 11. Save script after run it, and give a proper name. □
 12. Always analyze the whole dataset following top-down procedure: by school (average number by day), by time, by time of the day, etc.
 13. Instructions about graph builder:
 - a. Fix the order of hour column: value ordering.
 - b. Change color: right click - gradient □
 - c. Heat map analysis: Hour as X, School as Y
 - d. Save graph: Graph builder- export - interactive HTML □
 - e. Copy graph to document: Edit - Copy graph
 14. Despite studying on per visit, move on with unique student behavior ⇒ derive new field in a new aggregated file (e.g. How many times the person visits each day/ week, earliest and latest time the person visits, the frequency of visit in the different day for students from different school).
 15. Potential analysis: Prove that during exam week, there is not many people come in the early morning (use heat map to show). Focus on April as exam month. Check if there is a significant increase of population coming during the opening hours in the morning (5%).

Action Plan:

Item	Person in charge	Deadline
Concatenate different data file and finish data preparation.	Mengxi	Jan.30
Prepare meeting minutes and update wiki	Sijia	Jan.28
Finish overall analysis for the whole dataset and find patterns	Tianjing	Feb.1
Prepare work log and document for next meeting	Mengxi	Jan.30