

Supervisor Meeting 11

Date & Time: 05 Apr 10:30am - 12:00pm

Venue: SIS Meeting Rm 4.7

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

Absentees: Null

Agenda:

1. Demo Shiny application
2. Midterm Review

Details:

Further works on application:

1. Add on titles for each graph and add business questions to make it clearer and more understandable
2. Recode month and the rest of the binary values, etc
3. Change to percentage; Add label; Modify y axis accordingly
4. Redo dean's lister/singaporean vs. hour line graph
5. Rename user group to another more meaningful name: degree
6. Don't use scientific notation
7. Stack graph should be change to mosaic plot
8. Three Line graph together : use trellis graph GGally

Recommendation from supervisor:

1. If we want to compare between foreign and local students, we should have a complete flow, i.e. start from the heatmap. followed by bar chart
2. Compare the capacity rather than the exact number. What is the average number of people inside the library at 11pm-12pm? Get an idea about how many people come to the library. Use average per day, which makes more sense, then go further to distribution, ANOVA test. If it is significant enough, acknowledge the variations. Using total number does not give distribution.
3. Degree distribution: should show percent of number of entries per unit student. Otherwise, it does not provide us with significant results. Didn't tell us much. Guide the users to read the data. Should present in a more informational way.
4. Why use R? Explore and dig out the information cannot find from tableau

Conference Paper:

1. Sample paper. There are lots of data need to be cleaned, but main focus was on how can funnel plot help. What can be used to market strategy. Carefully read the two student sample post in e-learn
2. Your topic is what do you want to discuss about. Choose one aspect where analytics skills are put into use, either visualization analysis (box plot, heatmap) or confirmative statistics (ANOVA), or combine.

3. This sample paper is a library case. The main thing is the analytical method they are using. ANOVA. See how people can only one method, but still can write interesting journal paper. They didn't go into detail. But you can have session to explain the details: mean and median. Non-parametrics test. Refer to them to see how they derive the findings.
4. This is just one example. However, if you insist to focus on the visualization, you should take time to look at box plot, mosaic plot.
 - a. Box plot is not commonly used, thus you can say it can help.
 - b. Heatmap, do a comparison on the conventional methods people do and how can heatmap helps. What is the most common case? In paper, you can say we want to into the heatmap, what is it all about and so on.
5. Indicators are very important. Feature selection. Remember to ask yourself should I use variables I take from the database directly? If we use the absolutely number. DO NOT USE THE DATA DIRECTLY. This is exactly the point to do feature engineering. It is a part of the data preparation. It is time for you to rethink.
6. Likewise you can have confirmatory analysis and visualization both. Contingency table, distribution, ANOVA. Is data visualization enough or not. Not enough. Therefore, we need to go through both.
7. Topic of the paper: How does statistic analytics rescue the data analysis?
8. Add interpretations: Some things actually can be joined together.

Midterm feedback:

1. median: 74
2. You can still add in a lot after 7th
3. It's time to cut off development work

Action Plan:

Item	Person in charge	Deadline
Conference paper draft - abstract	Ren Mengxi	07 Apr
Conference paper draft - Data Preparation	Ren Mengxi	07 Apr
Conference paper draft - Data Visualization	Wang Sijia	07 Apr
Conference paper draft - ANOVA test	Wang Tianjing	07 Apr