

Analytics Practicum Supervisor Meeting 05

MINUTES

SEPTEMBER 22, 2016

1000 - 1100

SMU SIS BUILDING MEETING ROOM 4-3

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

1000 - 1020

RFM DISTRIBUTION ANALYSIS

ALL MEMBERS

DISCUSSION	<ul style="list-style-type: none"> - User can choose a cluster and observe the distribution of the RFM. It shows the boundaries of each cluster distinctly. - Prof mentioned that the no. of members in each of the cluster – we can still do it within the table but it will be good to display as a visualization and show how many % belong to cluster 1, 2, 3, etc. - Showing the user simply the transformed data will be hard for them to interpret. - Prof says to use the median of the original RFM data to label the results of the clustering to be “Low” or “High”, depending on whether the cluster results fall below or above the median of the original distribution - Add a reference line (median line) to better visualize the proportion of data falling below or above the median, which will make the concept of “Low” or “High” clearer to the user. - Keep the width of the bars constant across all 3 graphs so it will be easier to compare and contrast. 				
ACTION ITEMS	<table border="1"> <thead> <tr> <th>PERSON RESPONSIBLE</th> <th>DEADLINE</th> </tr> </thead> <tbody> <tr> <td>- Add a reference line to the graphs</td> <td>All members Sponsor Meeting 01</td> </tr> </tbody> </table>	PERSON RESPONSIBLE	DEADLINE	- Add a reference line to the graphs	All members Sponsor Meeting 01
PERSON RESPONSIBLE	DEADLINE				
- Add a reference line to the graphs	All members Sponsor Meeting 01				

1020 - 1040

HUFF'S MODEL

ALL MEMBERS

DISCUSSION	<ul style="list-style-type: none"> - Unclear about “select only distances within 50km and the 10 closest destination from each origin”. - Calculate all the distances so that there will not be any constraints. When we do the calculations, we should divide it - We have all the distances from each subzone to all the 28 libraries. - For the attractiveness of each library, we added the MRT stations, malls and tuition centres. - We are able to set the beta value ourselves, or if we use the other model, we use the data that we have to run a regression model to derive all the beta values. - The alpha and beta values are constant, so either we add it arbitrarily or we use a regression model/linear programming method to ensure convergence between the values and the data. - By the end of the day, we should have a probability value and it says that “if I stay here, there is a ___ % chance that I will go to a specific library”. - After getting the probabilities, we need to multiply the probabilities with the population in each subzone, to get the theoretical demand for each library - This will give us the patronage pattern of each library. - After getting the Huff's Model, we will map back the data so that we can derive the borrowing characteristics of each library. - We want a Huff's Model to calibrate the alpha and beta values. The purpose is to minimize the difference between the theoretical demand and the actual demand. We will only calibrate the alpha and beta values to achieve this minimum stated. - Alpha is the attractiveness coefficient, and to calculate it we will need to calibrate the weights of each variable. The final Huff's Model will have constant weights that are calibrated beforehand. - Can look at the inverse properties of the variables, to get a relative weight for each of them - Now what we can do is to fit a regression equation to get the optimal alpha value. - And also to calculate the optimal beta value. - It will be better to split the alpha value based on location. 				
ACTION ITEMS	<table border="1"> <thead> <tr> <th>PERSON RESPONSIBLE</th> <th>DEADLINE</th> </tr> </thead> <tbody> <tr> <td>- Calculate the optimal alpha and beta values</td> <td>All members Update by next Meeting</td> </tr> </tbody> </table>	PERSON RESPONSIBLE	DEADLINE	- Calculate the optimal alpha and beta values	All members Update by next Meeting
PERSON RESPONSIBLE	DEADLINE				
- Calculate the optimal alpha and beta values	All members Update by next Meeting				

1040 - 1050

DATA CRAWLING & SPONSOR MEETING

ALL MEMBERS

DISCUSSION	<ul style="list-style-type: none"> - User can search for a key term, e.g. "Tution", and we will be unable to show all the advanced filters. Prof asked is there a way after we crawl everything, then we can implement selection filters based on the categories and sub-categories. - No replies from NLB for the last few weeks. - Prof asked whether we have confirmed the tools of that we will be using. A: Shiny R. However, Shiny R is not very interactive compared to other platforms. - We should continue our project and build the Huff's Model even though the sponsor is unconfirmed about the details. - Try to implement all our work into Shiny R. - We will be able to put the D3 inside Shiny R, where all the visuals will be using D3 instead of R graphics – using Shiny R as an interface via R Server 	
ACTION ITEMS	PERSON RESPONSIBLE	DEADLINE
- To implement all works to Shiny R	All members	Update by next Meeting

1050 - 1100

MIDTERM PROGRESS REPORT & PRESENTATION

ALL MEMBERS

DISCUSSION	<ul style="list-style-type: none"> - We need to submit a midterm progress report, with a presentation - The presentation will be held internally - The deadline will be confirmed at a later date, either Week 7 Friday or in Week 9. - Report all our findings so far in the presentation, as well as the future plans. - To update the wiki based on all our findings so far as well. 	
ACTION ITEMS	PERSON RESPONSIBLE	DEADLINE
- Prepare for the midterm progress report and presentation	All members	Sponsor Meeting 01

OBSERVERS	-
SPECIAL NOTES	<p>Sponsor Meeting 01 will be tentatively scheduled soon. All members will present on their findings in the Team Meeting prior, and complete the PPT slides for the Sponsor Meeting.</p>