2103: Supervisor Meeting: Distance and Clustering

Date/Time   21 March 2018, 2:00 PM – 3:00 PM

Attendees   Shubhangi, Tanushree, Arushi

Tasks	Actor	Follow up Action
Updated prof on our discussion with the sponsor i.e. to revise the distance analysis to check for home range of the customers instead of travel distance - Home range: Mapping radius of sorts for customers in which they are traveling - Travel Distance: How much they are willing to travel from where they make the booking	Arushi	
Updated prof on the travel distance analysis: - Resolved the error - How to interpret the output file - Converted degrees to metres so that the data is usable and the projections readable. - With the lat long format, the distance is calculated in degree form , so we need to convert it: Create a new field> Call it Dist> input decimal value> specify 16 in the D field and 4 in the M field > Choose the type as $length>okay. We should get a new column with the distances calculated - Download map file from data.gov.sg by	Tanushree	Work on distance analysis

inpuuting ura master plan 2014 and look for the map		
- Home range analysis: ADE Habitat HR- sooftware to calculate home range. They have functions to help us calculate the home range	Tanushree	
- Chande on abstract to unsupervised methods - Clustering: Clustering cannot be used to predict (it is unsupervised) - Behaviour- RFM, distance travelled, Times (shopper behaviours) - Choosing the clustering bahviour: Ideally we want their social demographics but since we don't have that we map their behaviour: RFM, , days, times- use all of them as your input parameter. - What about variety? You can deal with it in 2 ways i.e. either use as input variables or after use them for EDA type analysis to see patterns within the group Suggested clustering variables (5-8) 1. Frequency 2. Recency 3. Spend - Restaurant Tier 4. Weekday Weekend 5. Time – breakfast, lunch, dinner 6. Distance Analysis For clusters : - Cuisine Variety - Att, canc, no-show (redeemed and not redeemed) - Discount Type : K - means clustering but if too many outliers normal mixtures -> JPM can do	Arushi, Shubhangi	Work on Clustering model