



1403: Supervisor Meeting: QGIS

Date/Time 14 March 2018, 2:00 PM – 3:00 PM

Attendees Shubhangi, Tanushree

Tasks	Actor	Follow up Action
<p>Distance Analysis : There is dirty data which is not usable or recorded wrongly: Latitude: - Should be around 1, >1.4 is not good Longitude: - Should be around 100 Singapore is between: - Latitude: 1.15 and 1.4735(1.5) - Longitude: 103.6 to 104(105) Therefore, we need to filter out data accordingly.</p>	Tanushree	Using JMP: <ul style="list-style-type: none"> - Sorting needs to be done to clean up the data in ascending or descending order - Filter out data according to the boundaries of Singapore - We should be left with only 307,052 records (about only half of the data)
<p>Import files into QGIS and save the file: 1. User data (where the booking is made from = origin) Map it out on the street map (use a plugin to add that) Change the file projection to svy 21 We will notice a few things: <ul style="list-style-type: none"> - No bookings from the port area - We will notice a few bookings from JB (might want to exclude those or keep those depending on what analysis you can draw from there) 2. Import files into QGIS and Save the file Vendor sheet (destination) <ul style="list-style-type: none"> - We will notice no missing values - All restaurants are in Singapore, so we find no dirty data (points in the Sea) </p>	Tanushree	<ul style="list-style-type: none"> - Convert the file to csv and import with QGIS - Check the vendor sheet for similar dirty data and then import (Layer>add Layer>Add delimited text layer>Choose your file (Integrated sheet)> OK. <ul style="list-style-type: none"> - Once you map it out, save file X: Booking Long Y: Booking Lat
<p>Do a check, See Properties→ Check they're svy21</p>		
<ul style="list-style-type: none"> - Go to Plugin> install plugin> mmqgis. http://michaelminn.com/linux/mmqgis - We need to use the function, called 'Hubline'. - Specify Where is the hub point (destination) and where is the spoke point (origin) - Then they will need the HubID and the 	Tanushree	<ul style="list-style-type: none"> - Do the Analysis as taught (the data will take a while using HubLine - Repeat the analysis using Hub Distance. Hub Distance gives you the

<p>SpokeID to calculate the distance</p> <ul style="list-style-type: none"> - Go to MMQGIS> Create> Hubline <p>Hub point layer: Destination Spoke point layer : Origin Hub Point ID : RestaurantID Spoke Point ID: RestaurantID</p> <ul style="list-style-type: none"> - So QGIS will look for restaurant ID from the origin data and map it to the restaurant ID in the destination data - Then look for the DIST column which will give the distances 		<p>nearest restaurant to the booking ID. We ultimately want to gain insights into whether people go to the nearest restaurant or not. Hub distance will give you that analysis and by matching it to the Hub Line analysis we can derive some conclusions</p>
<p>Clustering: We first need to identify the variables for clustering:</p> <ul style="list-style-type: none"> - If we want to cluster restaurants: <ul style="list-style-type: none"> - What do you want to derive from the clustering? Behavioural variables - Look at the unique id of the customer and look at variables such as frequency of visiting different restaurants/cuisines vs. frequency of using eatigo - Divide based on time→ Morning/Lunch/Dinner. - Weekend/Weekday. - Key Behaviourial Variables - Aggregate 	<p>Arushi, Shubhangi</p>	<ul style="list-style-type: none"> - Cluster using single variables first before getting into multivariate clusters to get an essence of the cluster sizes for each of the different levels per variable - Be careful to use users who have made more than one booking
<p>Steps Ahead for next week</p> <ul style="list-style-type: none"> - By next meeting have the univariate clustering ready and the distance mapped out 	<p>Arushi, Shubhangi</p>	<ul style="list-style-type: none"> - Analysis to be done