



3103: Cluster Analysis

Date/Time 31 March 2018 3:00PM – 5:00PM

Attendees : Shubhangi, Arushi

Sr. No.	Notes	Actors	Follow up Action
1	Shared with prof updates from our sponsor meeting		
2	<ul style="list-style-type: none">- We shared with prof our revised RFM and routine variables. He said no need to include the routine variable in the analysis . Also we must include all our formulas and their explanations in the report.- We mapped out distributions of all variables and shared with prof how we had transformed and standardized them:to make them compatible	Arushi	Transform the variables. Be ready by next meeting

	<p>with proportions we had used the global transformation that converts ranges to from 0-1.</p>		
3	<ul style="list-style-type: none">- Shared with prof, that our sponsor wants us to include all the categorical variables in the analysis and how we are facing a difficulty dealing with both categorical and continuous variables (RFM) together in the same analysis ,even after converting into proportions since clusters are not coming distinct.- Prof suggested that maybe we should look into latent class analysis since most of our variables are categorical. We started with latent class analysis also and looked into interactive binning- We realized that latent class would require us to use the sheet where each row is a booking, so we would be clustering bookings		

	instead of users and therefore cannot go ahead with it since our goal is to cluster users.		
4	<p>Shared with prof the difficulties we were facing in running clustering:</p> <ul style="list-style-type: none">- Transforming vs Standardizing variables: Sometimes when you transform there is no need to standardize if the scale of the variables get adjusted. There is no problem in transforming variables differently for input in the same analysis. Transforming depends on the shape of the variable distribution .- Ran individual transformations using the continuous fit function in JMP. All of the variables were suitable for the Johnson Su and Johnson Si transformations and therefore decide to go ahead with Johnson transformation. We decided to stick to Johnson transformation.		

	<ul style="list-style-type: none"> - Standardization of variables was not required. 		
5	<ul style="list-style-type: none"> - We first ran a K-means clustering as it is the best approach and then a normal mixtures. - Discussed with prof how to decide on optimal clusters and what is exactly meant by AIC, BIC and CCC. a larger CCC score is better. And if all are negative then change the method to normal mixtures. K-means was giving us a better output - Decided to stick to K-means 	All	Consolidate the clustering sheet by revising.
6	<ul style="list-style-type: none"> - Discussed with how to interpret the clustering output. - For the interpretation always look at the original raw data and not the transformed data by saving clusters to the data table. Then calculate cluster means by grouping by clusters. 	Shubhangi	Execute K-means accordingly keep in mind all factors. Should be ready by next meeting
7	<ul style="list-style-type: none"> - Shared with prof that we have decide don 	Tanushree	Derive clusters profiles

	<p>z-score profiling as our method. It was approved</p> <ul style="list-style-type: none">- Either use parallel plots or z-score profiling to profile. Parallel plots is visual and suited for when there is a lot of variation in the data therefore use z-score profiling.		<p>accordingly. Should be ready by next meeting</p>
--	--	--	---