TEAM V
ANLY482 SUPERVISOR MEETING
MINUTES 9

Date	15 March 2017
Time	17:20 - 18:00
Venue	SIS Meeting Room 4.5
Attendees	Team V: Andrew, Sarah Supervisor: Prof Kam
Agenda	1. Update supervisor of revised Cluster Analysis progress 2. Discuss with supervisor about the use of Association Analysis

S/N	Item Discussed	Remarks
1	Cluster Analysis	- Andrew updated Prof Kam that booking_monetary_average is changed to booking_monetary_total using total value instead of average. - Andrew updated Prof Kam that $\log_{10}$ was used to reduce skewness in booking_frequency and booking_monetary_total. For booking_monetary_total, the distribution of the transformed values were more normal but for the booking_frequency, the skewness was still present. This issue was raised by Andrew to Prof Kam to seek his advice on what we can do. - Prof Kam guided us in doing a distribution continuous fit in JMP. We did a fit for all distribution types and Johnson SI transformation was the best. - Prof Kam said we can use this transformation. Alternatively, we can transform the data into categorical values to see it as quantiles instead of the current continuous values. - Prof Kam also guided us to download an

		interactive binning plugin/tool for JMP, using version 2, then installing it. - Andrew asked Prof Kam if it is okay to use the untransformed version of of booking_frequency for the cluster analysis. In this case, booking_monetary_total is transformed and booking_recency will be used as it is (untransformed) because the distribution is quite normal. - Prof Kam said it is alright to go ahead. He said because the difference between value is not very large, the range is smaller, the results might be okay. - Andrew asked why booking_monetary_total is transformed nicely but booking_frequency will not. Prof Kam said that for booking_monetary_total, the values are continuous and the range is larger so difference is bigger. Got booking_frequency, the value are more ordinal where values are 1,2,3,4…, not in decimal places. - As the cluster analysis was done according to the above specifications, Andrew showed Prof Kam the cluster analysis results. The cluster analysis was done using K-means and a cluster size of 6 was chosen. Even though the optimal size stated in JMP is 8, we chose 6 because at size of 7, there was a very huge dip in CCC value. - Next, Andrew showed Prof Kam the profiles of the clusters. Cluster 3 is a good cluster that Vanitee can look into because it has high frequency, low recency, high monetary. Cluster 4 is a cluster where vanitee can go after to attract these customers to come back to Vanitee. - With the cluster results, Prof Kam is okay with it. Andrew mentioned that next we will look at the age distribution and EDA for each cluster. Prof Kam also suggested that we should look at the monetary contribution of each of these clusters.
2	Association Analysis	- We moved on the update Prof Kam about the Association Analysis. Andrew told Prof Kam that our team decided to use only transactions that contains 2 or more services. In this case, we will only have about 2000+ transactions to work with. So, this might be a challenge that we thought about hence Andrew asked Prof if a dataset of 2000+ be good enough.

		- Prof Kam said that typically, even the transactions with 1 item will be used in the analysis. But for our data, there is a very high percentage of transactions that contain only 1 service.
		- Sarah asked if should we then do the analysis to include transactions with 1 service. Prof Kam said we can do both. One with transactions with 1 service while another without. But we will need to see the results and the relative strength of it.
		- He also mentioned that we will need to observe the how strong the associations are. For e.g. there is only one rule in the analysis. nails and makeup service has an association of 80%, then this combination has a strong association. However, if we see that there are many rules such as {nails -> makeup}, {nails -> hair}, {makeup - facial} etc. and each of these has an association of 5%, then that means the association are not strong even though there are many rules.
		- Prof Kam briefly mentioned about the final deliverables. Sarah asked what the team can focus on in our paper. Prof Kam then taught us about the 3 stages of a business. He said that our project falls within the first 2 stages.
		  - Stage 1: How to acquire customers, who are our customers – segmenting into clusters
		  - Stage 2: to keep customers – cross selling
		  - Stage 3: retain customers/customers churn – survival analysis (set a time limit as to when customers will drop off)
		- But, Prof Kam said our team does not have enough time to do stage 3.
		- So for the final paper, our team can write clearly how the two analysis are used to help the sponsor.

S/N	Action Item	Action By	Deadline
1	Profile the clusters	Andrew, Sarah	By 21 Mar 2017
2	Perform EDA on clusters - find out the age, total amount spent for each cluster	Andrew, Sarah	By 21 Mar 2017