Supervisor Meeting #3

Drafted by: Liam Pang (09/02/2018)
Edited and Vetted by: Tan Rui Feng (11/02/2018)

Date	Time	Venue
09/02/2018	1430Hrs – 1500Hrs	SIS Student Corner

Participants: Ms Meenakshi Gopalakrishnan, Ong Geok Ting, Tan Rui Feng and Liam Pang

Agenda:
1. To share current team's data exploratory results
2. To share the team's obstacles with clustering

Meeting Item 1: Data Exploratory Results				
S/N	Issue	Action	By	Due
1				
2	The distribution of each variable was shared with the supervisor. The supervisor mentioned that the data can appear to be coarse.	Upload the merged file onto the team's wiki	Ong	14 Feb
3	Tan asked what is expected for the interim presentation. Supervisor shared that it is important for the team to share exploratory findings, the reasons behind the team's decision and what was shared between the team and client. She mentions the key is to derive more and meaningful insights from the team's limited data	Prepare for interim report	All	21 Feb
4	Supervisor remarked that it would be meaningful to conduct analysis to see how they would be able to convert more free users to become paying ones. However, this is	None	None	None

S/N	Issue	Action	By	Due
	challenging as there are no demographics data available.			

Meeting Item 2: Challenges with Clustering				
S/N	**Issue**	**Action**	**By**	**Due**
1	Shared that there are problems with clustering due to a significant portion of users having their counts at 0 and that free users have been removed from clustering analysis. Supervisor advised the team to compare each individual cluster to population mean to identify the traits and within each cluster. Users with zero count in various variables should in theory, go together into a single group. After that, these users can be split into active/inactive for further analysis. Ideally, the data fed into the K-Means should not contain too many 0. Supervisor proposed treating 0 as a separate group and segment only the active users.	Try two clustering scenarios, with the inactive users and without the inactive users. The analysis is subjected to the team's interpretation. Understand the proportion of inactive users via descriptive distribution and conduct an analysis of them separately.	Tan	21 Feb
2	For paid or unpaid user, Supervisor mentioned that K-Means should have helped the team to develop a label for the data group while doing an analysis. However, the team have the count of paid/free users.			
3	Supervisor mentioned that recency and frequency variables may help with segmentation. She proposed some variables may be created as a proxy for frequency, should as if the user is logging in daily, average sessions per week/day etc. This would spread out the behaviour in a more granular manner.	To look into new variables for analysis	Pang	21 Feb

Others

- The NDA was shared with the Supervisor. The copy was signed and returned to the team.