Supervisor Meeting #6

Drafted by: Tan Rui Feng (09/03/2018)
Edited and Vetted by: Liam Pang (10/03/2018)

Date	Time	Venue
09/03/2018	1100Hrs – 1200Hrs	SOL SR B2-01

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

Agenda:
1. **Updates on Dataset**
2. **Updates on Interim**
3. **Updates on Dashboard**

Meeting Item 1: Updates on Dataset				
S/N	**Issue**	**Action**	**By**	**Due**
1	The supervisor advised on using the graph builder to see the distributions of variables. Subsequently, outlier agents can be removed from the data set before doing boxplot so that observations can be made on what happened to this agent.	Explore graph builder to remove outliers	Ong	13/03
2	The supervisor mentions that it is not useful if looking at data on a 6 months level and recommends the team to look at monthly basis. To determine the timing which people tend to log in, the team has to identify when are people logging in through binning the hours with business sense. The supervisor suggests to see the mean/median for each hour and get the distribution through analysis.	Covert the data into monthly view to have a clearer understanding of how the user behaves	Ong	13/03

	As seen from the distribution done during the consultation, the team observes that very few people are online at 4th hour so this question of when people come online should be answered from business' point of view. The team would like to find out when are users more active across different hours. More number of agents who have sessions at this 1. Business: when are users active when studying distribution 2. Identify usage rates and leave it there Will percentages be better from count? The supervisor mentioned Columns created should have more meaning (e.g. midnight have no meaning since there are many zeros in this column) The supervisor advised on using Johnson transformation or log transformation before embarking clustering analysis			
3	As part of the exploratory analysis, the supervisor wants the team to look at two variables: 1. Frequency (number of sessions in the last one month): someone who logs in the first two months and those who never log in the next three months 2. Recency (categorical – yes/no): want to know	Relook at the clustering variables used for clustering	Pang	13/03

	whether this person has been active in the past one month A clustering variable to be considered would be the count number of session in the last one month. The team can consider counting of number of weeks that the users log in consistently e.g. 1 per week			

Meeting Item 2: Updates from Interim				
S/N	**Issue**	**Action**	**By**	**Due**
1	The supervisor suggests looking at distribution of number of sessions against the user-id. First, click on analyse and then distribution, the team can see that the median is 52 and the mean is 222. The team also needs to look at distribution of N where the median is 14 and the mean is 54 We need to first know 6 months median and per month median The supervisor suggests that there is no need to look at week level because it takes at least a month the transaction using business sense. The team suggests that we should look out for inactive users. When analysed the data, the number of sessions across data set: median is usually around 4- 8 and the 97.5% percentile have around 200 – 300. As such, 0.05% of population are super active on the platform while the the rest are the same			

S/N	Issue	Action	By	Due
	Therefore, it is suggested to remove this 0.05% and cluster the rest and also to remove those with all 0, cluster the rest			
2	There is a difference between infrequent users and users with zero sessions so the supervisor advises to eliminate those with zero sessions because they are not giving useful information. K-means, being an unsupervised technique means that we do not know the characteristics			

Meeting Item 3: Updates on dashboard				
S/N	**Issue**	**Action**	**By**	**Due**
1	Data is very skewed so may not want to look at average instead use median as a good representation The supervisor also mentioned that monthly average is more meaningful than average across 6 months. The supervisor mentioned that since all 4 variables are about how the users are engaged, the sponsor should be able to select what month they would like to see these variables. This is such that a monthly view of these data will provide more meaning. The supervisor recommends that average per month should be equal to total number of session divided by the number of agents. The team brought up that a median and average should be different but the supervisor says that the team should use median number as 'average'.	Changing mean to median for dashboard	Ong	13/03

	The supervisor also mentioned that in an event that the sponsor wants to look at aggregate, the team can provide a feature on the dashboard where 6 months can be provided.			
	As users who have been on the platform for longer period of time, they would have higher monthly activity compared to the three groups. Since the graph is not useful in showing this, the supervisor mentions that the dashboard can be improved. The supervisor also recommends on looking at number of enquiries through frequency rates on a monthly basis.	Exploring how the dashboard can be changed to factor in monthly view on dashboard	Ong	13/03
	The team explains that the number of sessions is placed on a daily view because REO can identify daily spikes within number of sessions which will be useful for them.	No changes need to be made	-	-