

## Supervisor Meeting #7

*Drafted by: Tan Rui Feng (13/03/2018)*

*Edited and Vetted by: Ong Geok Ting (15/03/2018)*

<u>Date</u>	<u>Time</u>	<u>Venue</u>
13/03/2018	1755Hrs – 1855Hrs	SIS MR 04-06

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

**Agenda:**

1. **Updates on Interim Report**
2. **Updates on Exploratory Data Analysis**
3. **Abstract**

Bring VGA adapter

Meeting Item 1: Updates on Interim Report				
S/N	Issue	Action	By	Due
1	<p>There is not enough thought on data preparation</p> <ol style="list-style-type: none"> <li>1. Segment for dashboard</li> <li>2. Segment for agent</li> </ol> <p>The supervisor emphasises that segmentation should be done to provide users with characteristics.</p> <p>For instance, for sessions wise, users can be more active on weekday/weekend</p> <p>The exploratory data analysis (EDA) should be done using business sense to analyse the data:</p> <ol style="list-style-type: none"> <li>1. General distribution of raw data</li> <li>2. look at different time bins</li> </ol>	To relook at the data preparation	Tan	20/03
2	The supervisor suggests a few tasks to be completed	Explore the segment profile node in SAS	Pang	20/03

	<ol style="list-style-type: none"> <li>1. The team has to first think about what does the team want to convey to REO to determine the clustering variables.</li> <li>2. Subsequently, the team can perform clustering based on all the variables you have and also to break down into weekend and weekday and to further break down into more timeslots</li> </ol> <p>The supervisor mentions that from business point of view, they would want to identify 3-4 segments based on 6 variables first. Subsequently, once the clusters are identified, the team can look at the distribution so that the team may find more characteristics of this cluster For instance, those who logged into the sessions who do not get any enquiry or those who do not log in and get many enquiries. may not want to have both do clustering for paid and free</p> <p>the supervisor also advises that even though the team should use the optimal number of clusters e.g. 5, the team can characterise by combining clusters together when interpreting e.g. cluster 2 and 3 are both freeloaders</p>	<p>Miner so as to see how different is the cluster from the population</p> <p>Through saving the clusters and run the graph for visualisation, the team can interpret clusters on graphs by seeing how different the clusters are as well as seeing how different the cluster is from the population</p>		
3	<p>The supervisor asks whether there is a need to break down listings into timing. The team suggests that the listings allow 99.co to know what they are using. However, this may be</p>			

	applicable to dashboard and not clustering. Therefore, the team needs to see the graph of sessions and listings to have an idea through performing statistical test like correlation analysis between total number of sessions and total number of listings			
	<p>The supervisor then shares the next step for clustering where the team needs to perform data transformation to transform sessions to a lower value. Some of the options the team can consider include:</p> <ol style="list-style-type: none"> <li>1. scale individual columns</li> <li>2. Johnson transform (untick scale individual columns)</li> <li>3. Logarithmic transformation by adding one to the zeroes</li> </ol>	Performing data transformation	Pang	20/03

**Meeting Item 2: Updates on Dashboard from Interim**

<b>S/N</b>	<b>Issue</b>	<b>Action</b>	<b>By</b>	<b>Due</b>
1	The reports in the interim report should include screenshots of metadata when explaining The suggested new variable “active” can be used for dashboard	Edit interim report	Ong	20/03
2	<p>The supervisor suggests two key things to think about for dashboard:</p> <ol style="list-style-type: none"> <li>1. What are the metrics to be used <ol style="list-style-type: none"> <li>a. number of free and paid users</li> <li>b. number of active people: based on 6 months (at least log in once a month)</li> </ol> </li> </ol>			

	2. What are the trends? We can see a monthly trend for each month via total enquiries			
3	<p>The team shared that the dashboard is used by client servicing team so these metrics identified have to be useful for them. For instance, 35% of all users who are active but how much can REO maintain or increase?</p> <p>The supervisor advises on defining on what is activeness based on data. For instance, the team needs to observe users' activity level in terms of sessions and listings. This is done through looking at the monthly distribution Some of the options that can be considered include:</p> <ul style="list-style-type: none"> <li>- Users who have at least two enquiries and listings</li> </ul> <p>The team also expresses that the sponsor wants the team to work on retention through showing proportion of people who are logging in across months</p>			
<b>Meeting Item 3: Abstract</b>				
<b>S/N</b>	<b>Issue</b>	<b>Action</b>	<b>By</b>	<b>Due</b>
1	The supervisor is unsure on how the abstract is being graded	Writing the abstract	Pang	