# Team Meeting #6

*Drafted by: Liam Pang (07/02/2018)*
*Edited and Vetted by: Tan Rui Feng (08/02/2018)*

| Date | Time | Venue |
|---|---|---|
| 07/02/2018 | 1900Hrs – 2200Hrs | SOE/SOSS GSR 3-1 |

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

1. **Updates on Data Cleaning**
2. **Data Exploration**
3. **Preliminary Clustering**

| **Meeting Item 1: Updates on Data Cleaning** | | | | |
|---|---|---|---|---|
| **S/N** | **Issue** | **Action** | **By** | **Due** |
| 1 | Ong has completed merging various sheets provided by 99.co to help us and shared the derivation of each variable | None | None | None |
| 2 | Around 12000 users are inactive as they do not have any sessions logged in the last two quarters of 2017 | None | None | None |
| **Meeting Item 2: Data Exploration** | | | | |
| **S/N** | **Issue** | **Action** | **By** | **Due** |
| 1 | The team tried filtering out those paid users from the data set as free users can only post at most 5 listings | None | None | None |
| 2 | The team looked at the distribution together and discovered that the distributions across all variables are positively skewed | Correct clustering variables via data transformation | Tan | 13 Feb |
| 3 | Ong will look at creating dashboard for these information | Explore dashboards | Ong | 21 Feb |
| **Meeting Item 3: Preliminary Clustering** | | | | |

| S/N | Issue | Action | By | Due |
|---|---|---|---|---|
| 1 | As many users have zero for variables like listings or cobroke or organic or synced, it is difficult to form clusters of similar sizes | Filter out those with zero and try clustering | Tan | 13 Feb |
| 2 | The team attempts to filter out outliers via hierarchical clustering (k-nearest neighbour). However, the cluster size after filtering out outliers is still not similar across clusters | Explore JMP for alternatives to filter out outliers effectively | Tan | 14 Feb |
| 3 | K-means clustering may not be the ideal method | Explore alternatives to clustering | Pang | 10 Feb |