

## Team Meeting 4

Date:	17 Jan 2017
Time:	4.30pm – 5.30pm
Venue:	SIS GSR
Attendees:	ALL
Absentees:	
Agenda:	Data Pre-Processing & Cleaning

No.	Details	Action By	Due Date
1	<p><u>Pre-Processing</u></p> <ul style="list-style-type: none"> <li>- Discussed on how to extract the search query out from the %20 separators</li> <li>- Came up with ideas for the following steps on how to do so: <ul style="list-style-type: none"> <li>o Transform query into string</li> <li>o Transform into basic words, search and replace</li> </ul> </li> <li>- Used RStudio for EDA, figured out that Tableau may be used too for data filtering and is fast, unlike JMP Pro 13 which crashes our laptop because of insufficient memory</li> </ul> <p><u>Data-Cleaning</u></p> <ul style="list-style-type: none"> <li>- Used Tableau and filtered by %20 for space between search queries (thinking behind it is that users usually search for more than 1 word as a search query, thus when they input a space, it will be converted into %20 in the string) <ul style="list-style-type: none"> <li>o However, this raises an issue of users who searched for 1 word only. As these are filtered away due to lack of %20</li> </ul> </li> <li>- Split by database because different database has their own way of expressing searches</li> <li>- Cleaned up and reduced the dataset into a size small enough to be opened on JMP Pro 13 without taking up too much memory space on our laptops</li> </ul>	Jing Xuan, Bendexter, Jun Rong	-

The meeting was adjourned at 5.30pm. These minutes will be circulated and adopted if there are no amendments reported in the next three days.

### Agenda for next meeting

1. Request for help with Data pre-processing and cleaning from Supervisor

Prepared by,  
Tan Jun Rong

Vetted and edited by,  
Bendexter Lim