



## SUPERVISOR MEETING 1

<b>Date</b>	25 January 2017
<b>Time</b>	2pm to 3pm
<b>Venue</b>	School of Information Systems Level 4 Sofa
<b>Attendees</b>	Prof. Kam Tin Seong (Supervisor)  Wei Xiaoxin  Wu Di  Zheng Xiye
<b>Absentees</b>	Nil
<b>Agenda</b>	<ol style="list-style-type: none"> <li>1. Discussing project scope</li> <li>2. Exploring potential R libraries for data handling</li> <li>3. Project management miscellaneous</li> </ol>

## AGENDA

No.	Task	Follow-up
1	<p><b>Discussing project scope</b></p> <ul style="list-style-type: none"> <li>✓ Went through documents provided by the sponsor with Prof. Kam. These documents include, data sets, data cleaning procedures as well as sample report format.</li> <li>✓ Meanwhile, of all reports requested by the sponsor, SOC Report is the most tedious and thus, should be prioritized as of now.</li> <li>✓ For data cleaning, input file is KTPH's monthly VTSOC data. Output should be a separate excel file highlighting all suspicious data rows. However, it was pointed out by Prof. Kam that this should be an iterative process, ensuring the final output is as clean as possible.</li> <li>✓ Prof. Kam suggested us on using R as backend language, which is free &amp; open source, enables easier customization &amp; modification and powered by many readily available packages. On top of which, tools like, RStudio, Shiny and R Markdown streamlines the development process and provides basis for data visualization construction.</li> </ul>	<ol style="list-style-type: none"> <li>1. Explore R tools like, RStudio, Shiny and R-Markdown. Think of how these tools may be of use in this project.</li> <li>2. Check with sponsor whether final submission should be in excel format.</li> </ol>

	<p>However, we should design the interface to be as intuitive as possible, which hides backend logic from the end-users.</p> <ul style="list-style-type: none"> <li>✓ Prof Kam. also highlights that data cleaning should be as robust as possible and report generation should be as simple as possible.</li> </ul>	
2	<p><b>Exploring potential R libraries for data handling</b></p> <ul style="list-style-type: none"> <li>✓ Since we are new to R development environment, we consulted Prof. Kam regarding some of the R libraries that may be useful in our project. R libraries highlighted by Prof. Kam include: <ul style="list-style-type: none"> <li>• dplyr</li> <li>• tidyr</li> <li>• readxlsx</li> </ul> </li> <li>✓ On top of which, Prof. Kam highlighted that for each CRAN has a pdf file for each R library, which functions like an API, highlighting all functions available in this library.</li> </ul>	<ol style="list-style-type: none"> <li>1. Explore 'readxlsx' library to load excel data as R frame.</li> <li>2. Explore 'R-Markdown' to create report and visualization in html or pdf format.</li> </ol>
3	<p><b>Project management miscellaneous</b></p> <ul style="list-style-type: none"> <li>✓ Prof. Kam suggested us to decide on internal task allocation as soon as possible due to time restriction. He suggested that two of us should work on data cleaning while the remaining one start working on report generation.</li> <li>✓ Some midterm presentation admins were clarified with Prof. Kam: each team will be assigned a half-an-hour presentation slot mid-term presentation slot to showcase their problem statement and check whether their approach is on track.</li> </ul>	<ol style="list-style-type: none"> <li>1. Zheng Xiye and Wu Di will be focusing on data cleaning process.</li> <li>2. Wei Xiaoxin will be focusing on report generation.</li> </ol>

The meeting was adjourned at 3pm. These minutes will be circulated and adopted if there are no amendments reported in the next three days.

**Prepared by,**

Zheng Xiye

**Vetted by,**

Wei Xiaoxin, Wu Di