Supervisor Meeting Minutes #7

Date / Time	7th MAr2018, 3.00pm to 3.30pm
Venue	SMU Library, Project Room 4-3
Attendees	Taffy Joan, Jerlyn & Jaehyun
Agenda	1. Comments on Slides

S/N	Notes / Task	Action by	Follow up
1	**Recap and Elaboration on Interim's comment** **Data Preparation**: Prof commented that the team should have shown how the staff list and training records were combined before the major concatenation of the 2 data tables. **Data Exploration:** Cross tabulation, Contingency tables should be used to represent the data. Mosaic Plot was recommended by Prof Kam for ease of comparison between the employee retention status and the period of analysis. Each year should have 1 Mosaic Plot. This ca be done by dragging the data into JMP, then selecting distribution analysis, fix y by x		
2	**Normality Test and Confirmatory Test** This test should first be done, under the distribution analysis. A suggested test is called "Fit Curve Test" to test the hypothesis of whether the data is normally distribution. There are 2 possible outcomes. First, that the data follows a normal distribution. In this event, parametric confirmatory tests should be done on the observations found during data exploration. Second, the data may not be normally distribution, in this case, non-parametric tests should be used. **Chi Square statistics tests** This test how significant the differences observed are. The team would be required to interpret the results of this test. It shows if the data is independent of or dependent on each other. This would help to establish if the employee statuses were random or biased.		
3	**Box Plots** The box plot is meant to show the distribution of hours and the average value should only be 1 single point. Currently the team has plotted the distribution of averages instead of		

	the sum of hours, hence the median shows the median of averages rather than the sum of training hours each employee undergone. The mean value can be added as a marker because it is not present in the default setting. **Average Hours** Currently, bar charts were used to represent the average hours, however, Prof commented that this visualisation is insufficient as it does not show the distribution		

4	**Additional Comparisons**
	Currently the team has only explored the distributions and difference of training hours and placements over the years for each individual business unit. Prof Kam suggested that additional comparisons should be made to uncover more findings.
	
	Mean comparisons can be done for:
	    1. EMOS vs JLT for each year (side by side bar chart)
	    2. EMOS vs Company
	    3. JLT vs Company
	
	The median should be used as the basis of comparison and the upper and lower quartiles to be used as compliments to the discussion. Doing do, the team may observe that there may be no significant difference in 2016 among EMOS and JLT while there is significant difference in 2017. Statistical tests would be needed again to confirm this.		

5	**Reorganise Graph** It would be ideal for the team to be able to reorganise the charts to show a bar chart with the Company, EMOS and JLT side by side for each year. With this revised comparison, the team may realise that while it may be true the JLT's staff undergone longer training hours in 2017, but, the hours may be the same in other years. The sponsor may run the risk of over generalisation as it easy to focus on the most recent data of 2017 and result in an inaccurate assessment of the data in the earlier years (2013-2016) One motivation for the revised graph is when there are year to year difference for each difference, hence a further comparison to the other business unit and the company would help to provide deeper insights		The team is unsure of about how the company can be plotted together with the business units because it is the sum of the both. The team should clarify in the next Supervisor meeting.
6	**New Hire** Similar Mosaic plots should be plotted for this section and the trend line in the bar graph should be removed as the trend is evident enough.		

		New hires compare Avg cost and placement Currently, bar graphs are used to display the average training cost and placement for the new hires, however, Prof Kam recommends that the team should use box plots to better study the distributions instead. JMP graph builder can do this plot.		
7		**Dashboard** **Monitoring** Prof Kam commented that the dashboard should not have the sole function of plain reporting. Instead he suggests that it should be versatile in being able to monitor expenditure quarterly or monthly to ensure the training expenditure is well spread out. **Indicators** He points out the tendency to be conservative with spending at the start of the year, only to realise that the company has underspent towards the year end and engages in a surge of training to meet the training budget plan for the year. This could help avoid sudden pumping of resources in the last minute by allowing the sponsor to regularly adjust the spending quarterly with the in-build indicators of the dashboard. The indicators would be able to reflect is the current expenditure status is above or below the target and it would provide as a signal to the sponsor to decrease or increase the spending correspondingly. **Turnover** This is an aspect that should be included in the dashboard. Since the research found out that the turnover is high, it should be highlighted to the sponsor to inform them of the number of new hires that have left the company. An application of this information could be in the forward panning of training as it shows the training needs derived from the number of staff that have left and thus require a replacement, which lead to the demand of training. Work specific courses can thus be planned accordingly with this information		
8		**Dashboard Requirements** Choice of platform to build the dashboard on is highly dependent on the sponsor. • If an interact and server-based platform is expected, JMP is a suitable candidate. • If an interactive dashboard is Not required, R shiny is good as it acts as a webservice that is able to generate a web report.	Jerlyn	Email the Sponsor to find out the current dashboard used by the company – DONE

	• However, if the Sponsor is unable to spend money on tableau/JMP/d3.js server, the team may consider using open source platforms to build this.		
9	**Administrative Matters** **Interim Submission** All tableau and excel files should be uploaded. Parameters The team ran into issues display the employee status in a time series format due to the nature of the parameters used which obtains data from 2 years. Thus, plotting a time series results requires the same year of data to be processed twice and displayed as output for 2 difference years. Prof Kam would look into the data and our files to get back to us with a proposed solution.		

Vetted by: Taffy, JaehYun