

Time: 10.16am

Location: SIS MR4-2

Attended By: Prof Kam, Janice, Siong Min, Hui Shia

Absent with Apologies: Nil

No	Agenda:	Action By:
1	<p>Prof Kam:</p> <ul style="list-style-type: none"> • Cannot perfectly determine that the approach is correct or wrong • Grouping by quartile we should group accordingly to the discrepancy of the volumes. (Like said in the last meeting) • When looking at volume aspect, we will look at the actual volume, so there is no actual problem in grouping them accordingly. At the end of the day should talk to the shipping providers • Should group accordingly to our tradelanes as different tradelanes have different volumes 	
2	<p>Siong Min/ Janice / Huishia:</p> <ul style="list-style-type: none"> • Janice: How to present the summary statistics more professionally? TEU grouping, utilization rate & carrier name. • Showing the formula of the TEU grouping. Kam say: Do a bivariate (Utilization VS Volume) > Fit Line - This graph gives you a very weird result, if you take a certain group of data then you'll get a good fit. Whole Idea: When volume increases, the utilization rate increases. • Using lasso tool and highlight the certain group of volume. When doing a bivariate fit on this group of volume, then you will get a really good fit. This might mean that the graph is too nice and the sponsors are not giving us the correct values (R2 is = 1) • Using the Multivariate and correlations – Y: Volume m3, volume gd line m3, total volume and actual TEU. The results means that Volume and volume gd line is the same thing and it is calculated out. From this we suspect that the values are derived values and are not complete volumes – Volume m3 is derived from total volume. Total volume has variations but not for the volume m3 and volume gdline. Hence volume m3 and volume gdline is not standard values. ALL THESE VALUES CANNOT BE USED TO CONSTRUCT THE MODEL. • Now do Fit Y by X: Y- utilization Rate, X- total volume, actual TEU. From this we can show that Total volume and actual TEU are the same but technically they should not be the same. Practically all data points are identical. HENCE ONLY CAN USE EITHER ONE OF THIS VARIABLES. • Should use GROSS WEIGHT – it shows better distribution. Karway approach of asking us to group accordingly to different TEU and different volume, IS WRONG. DON'T USE IT. • Should pair it with 20 FT, 40 FT containers. As it would have a better distribution and build a global model. Use Gross Weight and the different weight of the containers to dissipate them. This would help us to understand how well they pack the containers. 	

- As a data analyst we should be the one that advise them what is going on and the problems in the data. Should just structure the report accordingly.
- Fit model Y: Utilization Rate, Construct model effects: Gross weight, No 20 FT and No 40 FT. Don't need to change any variables. Results: have a very small R2 (very poor – there are many other factors affecting utilization rate, that we have not tried), two very strong explanatory variables between gross weight and No 40 FT. **EACH TRADELANES MIGHT HAVE DIFFERENT VARIABLES/CONTRIBUTORS.**
- **Maybe can use and analyze the 20, 40, 45, 20HC, 40HC, 40RF with actual TEU. And do a fit model. Add in the Carrier and seasonality (Month) to fit model as well. For months should do step wise regression to pick out the more significant months. Least square has more control (also can be used) – Step wise is better than least square.**
- **Not enough information to answer the LCL and FCL question.**
- 2 ways to approach this: build it as an **explanatory model**; **spend time to explain the different factors that would affect the utilization rate.** Else you can use it as a predictive model, and then we would need to test the data with the existing data available.
- There are also two types of containers: lying horizontally and vertically.
- Building an explanatory model: R2 value should be good (0.75 to 0.8 at least), analysis of variance has a v small p value and the F ratio is large, hence model is good. Parameter estimates: month, the different size of containers, carrier name and gross weight. Use step wise to eliminate those months that are not significant (not just using least square regression). **Factor profiling > Profiler** (helps you to understand what are the factors and how they actually affect the utilization rate (elasticity – shown as estimate on the model, high elasticity means it will affect the response variable by a lot.) Can change the value of the variables to simulate accordingly to our model. Must write down the equation to explain as well. **Save column > prediction formula** then just click on the column and view formula.
- **Gross weight should be a dummy, months cannot be an ordinal scale, should be nominal scale.**
- In the actual data we should have the volume, the weight and the number of pieces (Packages is zero or one) Electronic good should be in smaller cartons.