

Time: 10.45am

Location: SIS MR4-5

Attended By: Siong Min, Hui Shia, Janice

Absent with Apologies: Nil

No	Agenda:	Action By:
1	<p>Siong Min started with the explanation about filtering out 1 shipment first, and then move on to filtering 20 of it. 98%. Prof Kam's comments: It might not matter much. It depends on what we want to find.</p> <p>He would interpret it differently when it comes to 1 shipment only. Few stories to tell: For the low volume, esp those with small number of shipments, they can have different utilization rate. Their range can be from <10% to >90%. On the other hand, when we notice that when the shipment goes further, the utilization rate become more consistent, say 35% to 70%. We may find these things happening.</p> <p>His observation: - who are the contributors? By one shipping company? One shipping route? We cannot simply take out the 0 because they have a wide separation. For this variation, could it be because they are representing different route and shipping company? Or time?</p>	
2	<p>Generation of charts – up to us if we want to use JMP, EG etc.</p>	
3	<p>“Pack better” & “Utilization Rate” → Both are two different things.</p> <p>Pack better and utilization might not be the same thing because pack better is how they pack it. Utilisation – there are containers due to their nature, we need to reserve more places. For eg, automobile. Utilization rate might not be that full because they are spaces. → It doesn't mean high UR means pack better. Unless we have some measures to prove it, else we should not infer that way or draw any conclusion.</p> <p>What we want to do as a start is to focus on the data and draw different observations and insights.</p> <p>There are no indicators or any factor to prove that both of them are same. → Use UR to understand their efficiency.</p>	
4	<p>Hui Shia explained about our concerns about using volume as our calculation. More volume → More utilization rate???</p>	
5	<p>Kam mentioned that with the data we have, we should not use mean. We are given variables to give us insights.</p>	

6

Missing link:

The context is not very clear. We might have something in mind but its not very clear.

Context → maritime shipping; when we talk about UR, look at the relationship based on the shipping route. We know that different shipping route will have different volume. Or if we use port of origin as our context, or port of destination.

Lets assume we start with shipping route.

We start looking at what are the routes, and look at their volume and so on, then give him some understanding of the different routes and the volume that they have over the last 2 years period. Look at their utilization rate. If we just look at one particular route, they might vary different season, but time, shipping company not fixed. Then in actual fact, out of the 7 companies that uses the shipping route, we found out 3 companies that are consistent.

If we look at port, we look at what are the routes. For e.g., these 3 routes have higher UR compared to others.

WE NEED A CONTEXT TO GUIDE US TO DO THE ANALYSIS!

Prof Kam stated that he don't know which one is a better one but we can approach any of it. Taking the example of one route, if we start looking at it, we take one route that has high total volume (not container size) and then we see how they varies over time. Also, what kind of ships are they using? (usually there is a relationship between the total volume and the type of ship that they are using)

Challenge: shall we only focus on the top 10 or that? We should work on 5, the rest are very small. Most proper one 5-10. After which, bring in the different parameters. Temporal patterns, port they are going through, carrier lines they are using.

With such analysis on the individual shipment, we can get more in depth analysis. And then, we will really be able to construct an explanatory model. We should be able by the sheer insights of the data construct a good, reliable and predictive model. However, we need an explanatory model at the start.

Key point: If we lump everything together like we have done, we wont be able to see it.

Siong Min's question: should we focus on 5 only? Do we have to justify them?
Prof Kam's reply: just do top 5. Divide them by percentile (what we did was correct). In general, we divide the whole frequencies into 10 equal intervals and then we take the top 10 percent. → We should use the total volume as a gauge and then we sort it. We do not use UR because it is a response variable. After which, do a **summary** based on the ship route.

The team has decided on top 3 for the time being.

7

JMP: Graph Builder (Prof Kam's walkthrough)

1. Bring carrier_name 1 to X axis. We are able to tell which one has a lot. CMA CGM, Hanjin Container Lines, mediterranean company
2. Bring utilization rate (change to mean using summary statistics) to left
3. Click boxplot.

The 3 large one have a lot of variation.

We look at the 4 (CMA, hanjin, Lloyd, Mediterranean)

Whether they pack well or not is based on the company. But based on the same company, they have a variation. Therefore, we can show this to them.

→ Find out what are affecting the UR. Due to carrier? Time? Ship? (Technically its all)

→ Do an overview based on the month.

Drag carrier to top right corner then we can compare both.

Conclusion: there are a lot of different variations. We can thus build an explanatory model to explain the rate of change. We want to make sure that the variables we have is in line with what the sponsors want.

Load the different variables to let the algorithm find out which are the parameters that are more significant. Run multivariate regression. Find which are the independent factors that have greater influence on the UR.

We isolate them one by one. After which we need to build a model to see the relationship between them.

If the kind of influence is not only one but also a combination of factors, take those possible factors, and then try to find out how they interact with each other.

Next part: Which model are we gonna use? Step wise? Normal regression?