- Data preparation is very important, and needs to be thought through. We should ask the sponsors if it makes sense to remove outlier or remove a certain field.

- For the sample interim report, one group did more on continuous work and the other on exploratory. However, both groups made it clean on how they cleaned the data.

- Even if you think the data is dirty, it might be an essential part of business logic. Need to discuss discrepancies with the sponsor.

- Prof Kam expected at least 30 pages of our research findings and we need to check with the sponsors if it makes sense to them.

- Before the group moves to predictive, there needs to be a more in-depth discussion of findings.

<u>Presentation</u>

- The presentation should be in a story-telling format. It needs to start from a broad overview

- First slides needs to recap on the project

- The next few slides are the title and the main objective of the project

- In part of understanding, highlight the objective. The sponsors might ask non-technical people to listen to the presentation. Hence, the presentation should be clear enough to a layman

Slide 1:

- Put the graph together to give a comparison. When Prof Kam was practicing in the industry, he made it a group to meet different groups like the business development to have a clearer picture of the situation.

- It is ideal to put the industry combined ratio together with Tokio marine's ratio in the same graph

Slide 2:

- There is a need to show lost ratio in a time-series format. You will need to let the audience know what period of data is used for the presentation of graph.

- Not wise to lump everything together

- Call graphs "figure" rather than "chart"

- There should be a slide that established how the ratios are calculated. The way you calculate might be different from the company.

- Performance of ratio

-> Show average performance as a whole before going into specific trends. This is to keep the audience focused.

- Do not underestimate exploratory. It can be very insightful if done properly.

<u>Predictive model</u>

1) Forecasting

- Is revenue accurate? Should you use GWP or use the accounting form of revenue?

- Separate new and renewal policies

- Good to focus only in new policies

- use all the time series you have

- Have forecast for new sale base on Japanese and local, corporate and personal. Break down and don't assume all market. Forecast by submarkets.

2) Predictive

- There is a high drop off rate

- Predict who the people that will renew are (You might not be able to do a complete picture of it but you can try to give something basic)

- Okay if explanatory is too low (ensemble method) or error is too high (logistic, decision tree, random forest, bootstrap or bootstrap forest).

- Japanese, local, private, corporate, what will happen in these demographics?

- Make sure people don't change their taste. Validate and make sure response data don't change much.

- Predictive looks at the lost and see if they renew or don't renew policy.

- Talk to business owners which are possible churn areas. Take a quick look and anything that might be missed out. Choose appropriate variables.

3) Text mining

- Claim amount & vehicle description can mine report & predict claim amount.

- Pick key word and predict the claim amount.

- Gather Claim Report that show claim amount if possible

- Data that only appear one or twice can be ignored

- Focus on car accident, claim amount and text

4) Training and validation data

- Use one by random and the other by stratified random. By the end of the day, it there any difference? Is stratified random okay?

- 20% training, 20% validation.