**DATA SCIENCE PROJECT 1 (MAST90106)   
Semester 1 2021**

**GROUP 17**

Meeting No. 5

**Date & Time:** 22nd April 2021, 12.00 PM - 01.00 PM

**Agenda:** Update on the project with clients

**Attendees:** Kartika Waluyo, Vrinda Rajanahally, Roberto Bonelli, Brendan Ansell

**Discussion:**

Weekly Meetings**:**

ONLINE MEETING: Wednesday, 3 PM – 4 PM

IN-PERSON MEETING: Friday, 10:30 AM – 11:30 AM

Notes on Task 1:

1. Samples and Genes – only two datasets in TMM\_norm\_filt\_count\_Tissue required for the task

1. Create 2 files: sample\_tissue1 & sample\_tissue2
2. To count the overlap between 2 tissues: sum(sample\_tissue1$SUBJID %in% sample\_tissue2$SUBJID)
3. To find the proportion: sum(sample\_tissue1$SUBJID %in% sample\_tissue2$SUBJID) / length(sample\_tissue1$SUBJID)
4. To count the shared genes: sum(gene\_tissue1$gene\_id %in% gene\_tissue2%gene\_id)
5. To find proportion of shared genes:

**Next steps/Decisions:**

New Tasks – four tables:

1. Tissues x Tissues : count of sample no.

2. Tissues x Tissues : Sample proportion

3. Tissues x Tissues :  count of gene no.

4. Tissues x Tissues : Gene proportion

Packages in R for heatmaps and clustering:

1. CorrPlot

2. Complex Heatmap

3. Tidy Heatmap

Functions in R:

1. Map
2. Apply

For Wednesday, 3rd May 2021:

1. Four new tables

2. Visualizations

Send the subject description to Roberto and Brendan

**Resources:**