In **tab\_data v2**, we assigned five members to extract data over a two-month period each:

* Usha: May, June 2024
* Jungkyu: July, August 2024
* Mishika: September, October 2024
* Daniel: November, December 2024
* Mhonika: January, February 2024

“Data\_catergorization.R” - Uses dictionary of beauty brand names to extract subset of data containing shopping sessions with at least 1 beauty product. This would give the subset data by the different months stated above.

* cbind ( sep-oct-2024.csv to jan-feb-2025.csv) to get “full\_subset.csv”

After gathering all subset data from our team members, we combined all the CSV files into a single file named **“full\_subset.csv.”**

* Using “full\_subset.csv” we came up with our analysis:
* Loyalty/Unloyalty
  + “Loyalty\_Unloyalty.csv” is the R code to find the customers behavior when they are buying products from unique markets. There are 3 graphs in code with notification.
  + Slides 7,8
* Shopping behavior
  + “Shopping\_behaviour.R” is the R code used to average price of beauty products, average price by gender, matched beauty products by store, proportion of search terms
  + Slides 9,10,11
* Most bought category
  + “Most\_Bought\_Catergory\_byStore.R” is the R code that tells us the main beauty categories by different store
  + Slides 16
* Conversion Rate
  + “conversion rate.R” explain how we got the bar graphs for 3 markets
  + slides 12,13,14
* Macro/Micro categories
  + “Macro\_Micro\_catergories.R” is the R code used to further categorise out beauty products into macro categories (Skin care, Hair care, etc) and micro categories (Skin care - Moisturiser etc)
  + Revenue by macro and micro categories, purchase range by macro categories, association rules, sample for accuracy
  + Slides 17,18
* Sample.xlsx
  + Excel output for a random sample of 50 items per main category
  + Manually validated categorisation of main categories
  + Slides 5