*Stephanie A. Managbanag*
*Ad Hoc Analysis*
*LT#1*
*ITMGT25-B*

**1**

The first few codes open and read the json files. The transdf with the str.split which is seen in the next line separates the values within the transaction items column so that it splits per item. The transdf.explode puts each item in a new row. The def quantity gets the quantity of the transaction_items column and then puts it in a new column. After this, I extracted the non numeric characters in the quantity column and then transformed it into a new variable. For the def product function I subtracted the quantity. Following this, I transformed the transaction date column into a datetime format. Next is grouping by transaction month and transaction items. Finally, sum the quantity and then unstack it so it is viewable in table form.

**2**

I extracted the price manually and then put it in a new column. After this I multiplied the price column with each column in the previous table. Since the for loop also multiplied the quantity price to itself, I replaced the column with the original values of the price.

**3**

I initially grouped the transaction date, month and name. After this I determined the size so that I could see how many times a customer bought items in each month. For the repeaters, I put it in a for loop that checks whether the customer fulfilled the criteria, get the length of the dataframe and put it in a list. The primary reason for this is we have to check whether the customer month in the previous month and the current month, hence, the for loop. Similarly, for inactive, I had to check whether the customer bought in the previous months and buy any items in the current month. Lastly is engage, which is similar to the previous mentioned functions. I had to check whether the customer bought in the previous month and the current month. I put everything in a data frame and then added the months as the indexes. Lastly, I transposed the data frame.