

Get what is needed for processing

a. name | transaction\_items | transaction\_value | transaction\_date  
based on this .json file item formatting the 'transaction items' should be split into different entries. After splitting the entries per item, the entry should be split into :  
(Brand, Item, Quantity)

I also checked for the year, but they are all the same, so we really only need the months here, so Final working df columns should be: (Name, Brand, Item, Quantity, Value, Month. [Use a pivot table and sum all the quantities within the month](#) to create the final table. Use item as index and columns as the months and for other dataframes, use groupby or pivot\_table, whichever is better suited

Repeaters:

For repeaters I used the list of items, then removed duplicates of name and month, this way I only have unique (month,name) combos. After this, because the dataframe isn't efficient for getting repeats, I sorted it first according to name and month, in this case using month\_num since sorting month will put April first as it starts with 'A'

After this I check for repeat names. this checks if the current cell that I'm in still has the same name. If it still has the same name, then I will check if the month that I'm in now, is consecutive to the month of the previous entry, if yes then that instance is a repeater

Inactive:

Create a table of transactions per person. This consists of months as the columns and the members as the indices. If person had a transaction in month x, then data point = 0. In a given data point, if it's 0 I should check if the values to the contain 1, and if yes, then that person is considered inactive for that month