My approach to determine the count per item per month is shown in the following sentences. When the dataframe for the whole json file was loaded, it was clear that I needed to split the column which held the “transaction_items" into three distinct columns as each row possessed around 1-3 items with their corresponding quantities. After which, I was left with a new dataframe with three columns. Each column’s rows either housed one item with its corresponding quantity or the “None” value. Simply, I needed to split each column in this new dataframe so that I could yield a better dataframe which isolated the quantity of each item in a new column. Simply, this new dataframe would have this set of 6 columns: a. Item 1; b. Quantity for Item 1; c. Item 2; d. Quantity for Item 2; e. Item 3; f. Quantity for Item 3. After yielding this new dataframe, I further splitted it into three dataframes. Each new dataframe housed two columns: the item name and the quantity. Since again there are three items in the original dataframe, there will be three new dataframes in this step. After which I added the month column to each dataframe. Thereafter, I simply created a pivot table for each one, and I added all of them together to get the total count per item per month. Overall, my approach to **determine the total sales value per item per month** was similar to that of my approach to the previous problem. The only two steps that I needed to add was to use numpy to determine the price of each item and I also had to input an additional three columns which held the transaction price of each item based on its actual price and its corresponding quantity. To do this, I did the same process wherein I created three new dataframes, each with the following columns: 1.) item name; 2.) sales value; and 3.) month. After which I created pivoted tables for each one and added them together to get the final table. My approach to **determine the different customer types item per month is shown in the following sentences.** After extracting the month using datetime, I created a new dataframe from the original one with only the names and the months. I realized that I can create 6 new columns with each column holding an array of boolean values (True or False). Each column from this set of 6 will be titled by a specific month and will serve as the category for that column. Hence, in the Month 1 column, a specific row will only show True when the value of that same row in the original months column is 1. I did this for all 6 months. After which I needed to determine the amount of times a person purchased in a specific month. To do this, I created 6 separate pivot tables which essentially added up the boolean values from the previous step for each month. Ultimately, I concatenated everything together and thereafter, I merely used “np.where” as my functions to determine the different customer types.

(Note: I used these same processes to generate other tables which categorized data by sex, generation etc.)