Table 1 (Count of Each Item Sold Per Month):

In determining the number of products sold per month, I mainly made use of the "split" and "apply" functions, as well as pivot tables. Firstly, I used several "datetime" functions to extract the months from the raw data. After doing that, I used the "str.split" function in order to split the "transaction_items" columns into three columns; and then I used the apply function to reformat the values again. In doing this, I was able to extract the quantities of the products. Subsequently, I formatted the quantities into integers by using the apply function. I then split the existing dataframe into three specific data frames to store each transaction column and their corresponding quantities. In order to group the products per month, I used a pivot table that specifically used the aggregation function "sum". After doing that for each dataframe, I merged the three pivot tables into one.

Table 2 (Total Sale Value per Item per Month):

For this table, I mostly used my code from table 1. The only main difference is that I had to get the price for each product. I treated the transactions as a system of equations with 7 variables since there were seven products, as seen from the previous table. I assigned each product into one variable. I made an array that would store the quantities purchased for each product for the first 7 transactions, and then another array that would store the transaction values. Using the built-in linear algebra function in "numpy", I was able to get the price for each product. After getting the prices, I multiplied them to the values I got from table 1.

Table 3 (Repeaters, Inactive, Engaged):

For this table, I made use of boolean values in order to get the sum for each category. Firstly, I used the "apply" function to generate new columns that would store boolean values for each month. I then separated the existing dataframe into six different data frames. For each dataframe, I made it into a pivot table in order to sort the data together for each month. After that, I extracted the data values I needed for the applicable month into new dataframes. I did this method for the other 5 dataframes based on the month I needed to filter them with. After getting the purchasing history of each customer, I used the "np.where" function to check if a customer would be classified as a repeater, an inactive, or an engaged. Since this would result in boolean values, I used the sum functions (0=False,1=True) to get the number of repeaters, inactive, and engaged customers per month.

Extra Tables (Sex, Generation, Count of Transactions (Orders) per Month):

In these tables, I mostly used the apply and "str.contain" function in order to transform the values into boolean expressions. I would then add these data points in order to get the desired values.