

1. **Loaded the json file**
2. **Processed and cleaned the Data:** I processed and cleaned the data using the split function wherein semicolon was set as the delimiter. Then, I defined a function to extract the integer values in each transaction that represents the quantity sold. Since at most three items were bought by each customer, three new columns were then added to the original data to indicate the quantity sold for each item when that item is bought as the first, second and third item (i.e. Item 1, Item 2, Item 3).
3. **Breakdown of Count of Each Item per Month:** I created three tables using groupby and sum function to indicate the total quantity sold for Item 1, Item 2 and Item 3. I merged the three tables based on the Month and Item Name. I also sorted the values based on the month number (i.e. January =1, ... June = 6) and ascending order of the Item Name.
4. **Breakdown of Total Sale Value per Item per Month:** I solved for the cost per item based on the first 7 transactions using Sympy. Using mapping, I assigned a variable name for each item and assigned item cost to the variable name. I multiplied the columns containing the Item Cost and Total Quantity sold and created a new column.
5. **Repeaters, Inactive and Engaged Metric:** I retained the latest transaction per customer and dropped the other transaction. I created lists to summarize the truth values indicating the customer activity per month. Then, I defined functions using for loop and if, else to count the number of customers per metric. I tallied the repeaters, inactive and engaged per month then created additional columns for it. I created a new DataFrame for each metric. I used the sum function to compute for the values per metric in each month then created another set of DataFrames. Then, I transposed the rows and columns of the DataFrame.
6. **Additional Metrics**
 - a. **Total Quantity Sold and Cost per Item Graph:** I used Matplotlib library, groupby, mean and sum functions on the Breakdown of Total Sale Value per Item per Month.
 - b. **Cohort Counts Table, Customer Retention Table and Retention Rates Table:**

Based on the transaction dates and usernames, I created a cohort counts table by defining functions using datetime and groupby then I created a pivot table. Divide, Iloc and round functions were used to generate the customer retention table. I used Matplotlib and Seaborn Libraries to create the Retention Rates Table.
 - c. **Correlation Heatmap:** I defined the function to generate the ages of the customers. I also assigned 0 and 1 to represent F and M sex, respectively. I used Matplotlib and Seaborn Libraries to generate the heatmap.