Summative Assessment #1:

Ad Hoc Analysis

Approach:

- Accessing Data
  - Import packages and read "transaction-data-adhoc-analysis.json" file
- Making New Dataframes (Price DataFrame: Getting Price Per Item, Transactions DataFrame: Cleaning Data and Adding Months, Customer DataFrame, and Customer Activity DataFrame)
  - Get necessary columns from the original DataFrame. The making of numerous dataframes is to ensure that if entries are split (like in the Transactions DataFrame), other values/columns will not duplicate or affect each other. (Note: The Jupyter notebook has more detailed steps for this part.)
- Extra Information (Total Number of Items Sold, Total Sale Value, Number of Items Sold Per Item, and Total Sale Per Item)
  - This is simply extra information that the coder made for peace of mind.
- Required Outputs:
  1. Count of Each Item Sold Per Month **and 2.** Total Sale Value Per Item Per Month
     - Dataframe was grouped by month, then add the respective columns needed (i.e., quantity and sales value, respectively)
     - Use pivot tables for cleaner visualization.
  3. Repeaters/Inactive/Engaged Customer Activity
     - Steps taken were split by category (i.e., Repeaters, Inactive, or Engaged)
     - The *accumulating customer transactions* are for the accumulation of 1's and 0's. (Note: It's like .cumsum(). But if a customer does not purchase in that month, the count accumulation restarts.)
     - Repeaters: If the accumulated count is greater than 1, that means the customer purchased something the month before.
     - Inactive: The 0's are replaced with a negative number, so that when the previous and next column are compared, the desired True or False results are given (Note: True is for when the count is negative, which is initially 0, meaning that the customer did not purchase on that month).
     - Engaged: Based on the accumulated count for the month, the values are counted and categorized. The diagonal values are taken because it shows the number of customers that have purchases every month.
     - All three categories are put together for cleaner visualization. The values in the first month will always be 0 since there is no existing/known data for the previous month; hence, changing the existing January/first column values is valid/okay.
- Additional Charts/Graphs (Amount of Products Sold Per Month, Amount of Products Sold Per Item, Sales Per Month, Sales Per Item, and etc.)
  - This is for more ways to analyze the collected transactional data. This uses the previously made **new** dataframes, and mainly analyzes the transactions and customer demographics.