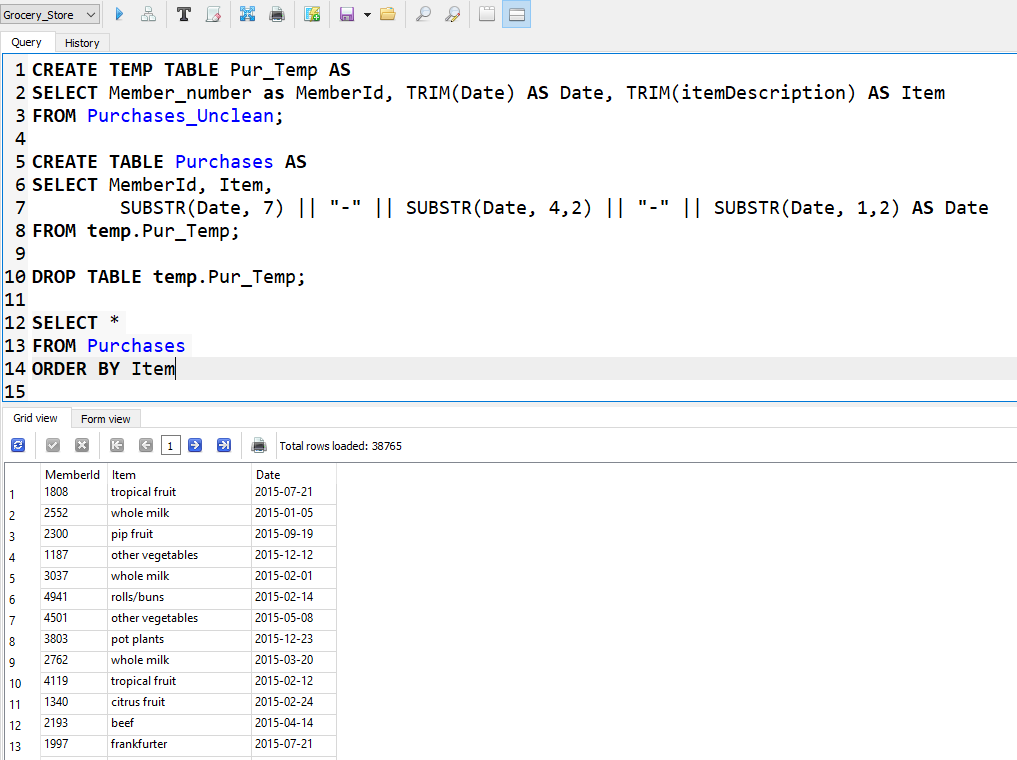
**Question 1**

I. Cleaning and creating the table for analysis

Going through the data, I realized there was no missing data nor any format inconsistency. Member Ids were all 4 digits and correct. No typo in Item Description was spotted either. Columns Date and Item description were trimmed and the date was changed into yyyy-mm-dd format. Additionally, Column names were standardized.



II.

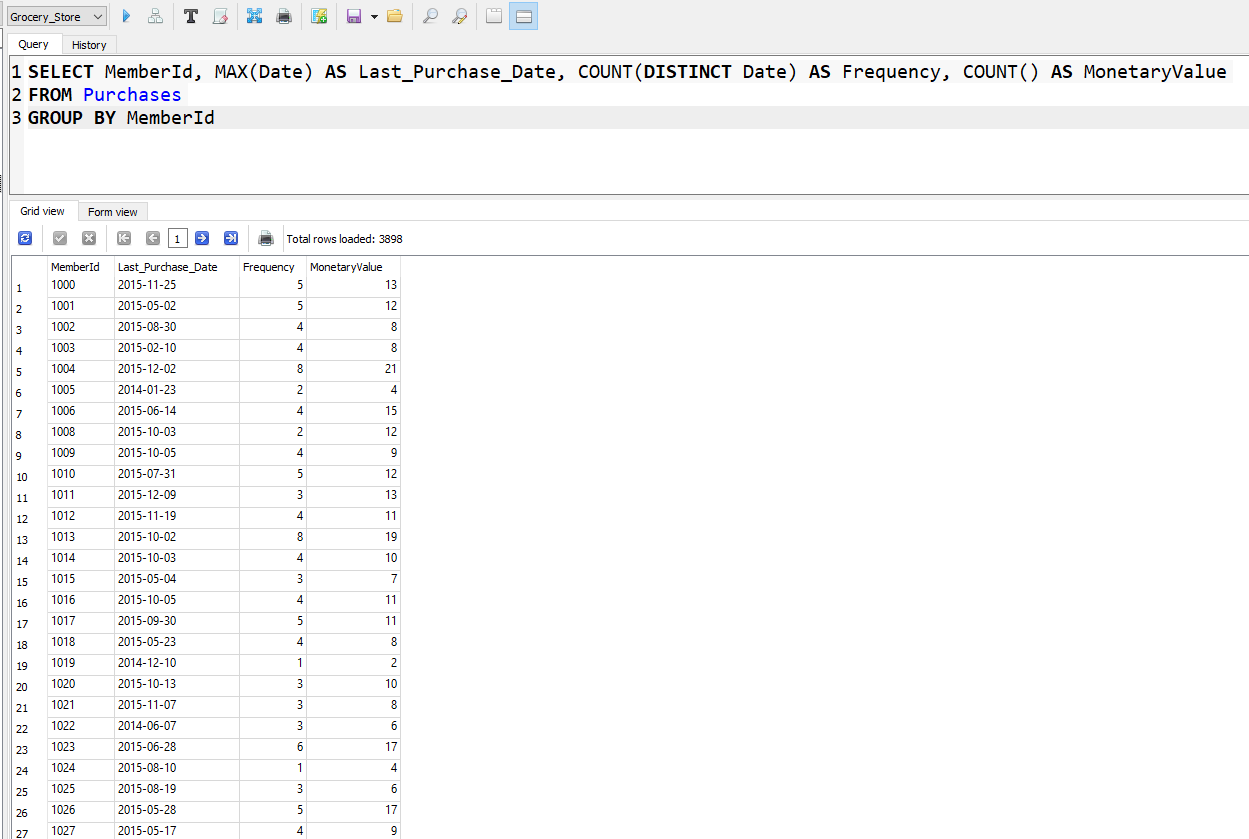
**Question 2**

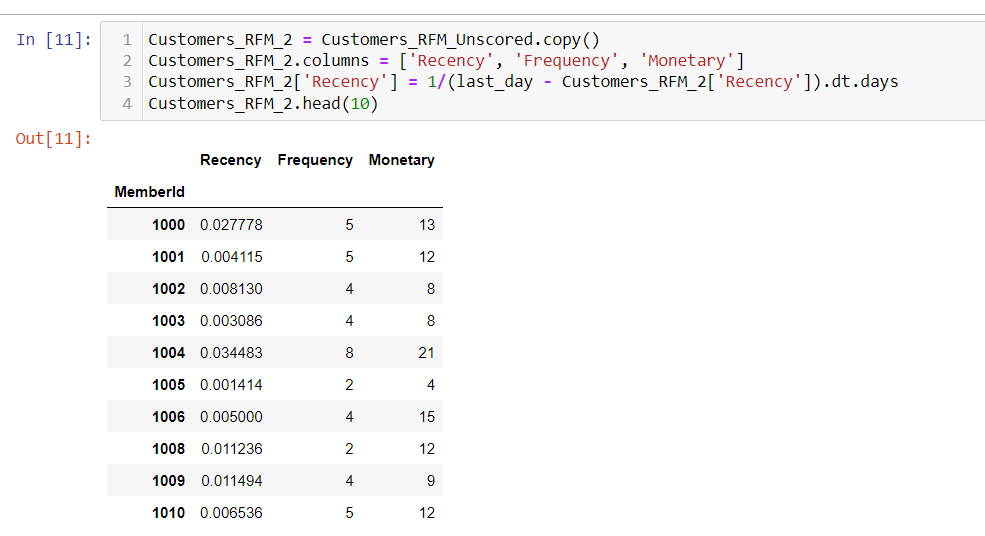
For a RFM analysis Recency, Frequency and Monetary Value should be calculated for each customer.

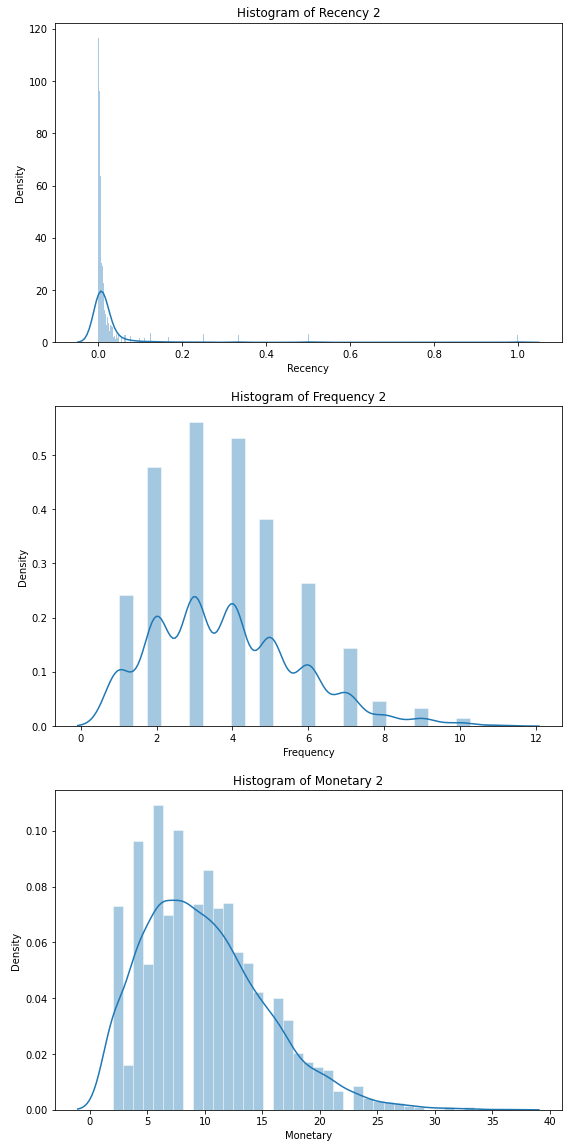
Monetary Value: Since the data does not include any information on the values of items and the money spent by a customer, Monetary Value can not be calculated as it is normally. I decided to use the number of items bought in total by a customer to calculate Monetary Value.

Recency: It is calculated as 1/(last working date + 1 day – Last purchase date)

Frequency: This variable equals the number of days at which a customer has made a transaction.

For calculating these variables, First using the following query, Last day of purchase, Frequency and Monetary Value were calculated.

Using the results, a data frame is created to hold the final RFM values.

Histogram plot is plotted for the RFM values.

**Question 3**

Customer segmentation with k-means