Performing Data Mining on Retail Data

Importing Pandas for dataframes
matplot and squarify for plotting graphs
datetime to calculate recent transactions date(Recency)

Reading excel file and then removing noice. Dropping all instances with quantity less than 0, As quantity cannot be negative.

Later calculating total price per customer based on quantity and unit price.
And from those values, calculating Recency/Frequency/Monetary as below :

- RecentTransaction(Recency) = current date - Last transaction date
- NumOfTransactions(Frequency) = Total No. of Invoice per customer
- MoneySpent(Monetary) = Sum of Money spent for all transactions by single customer

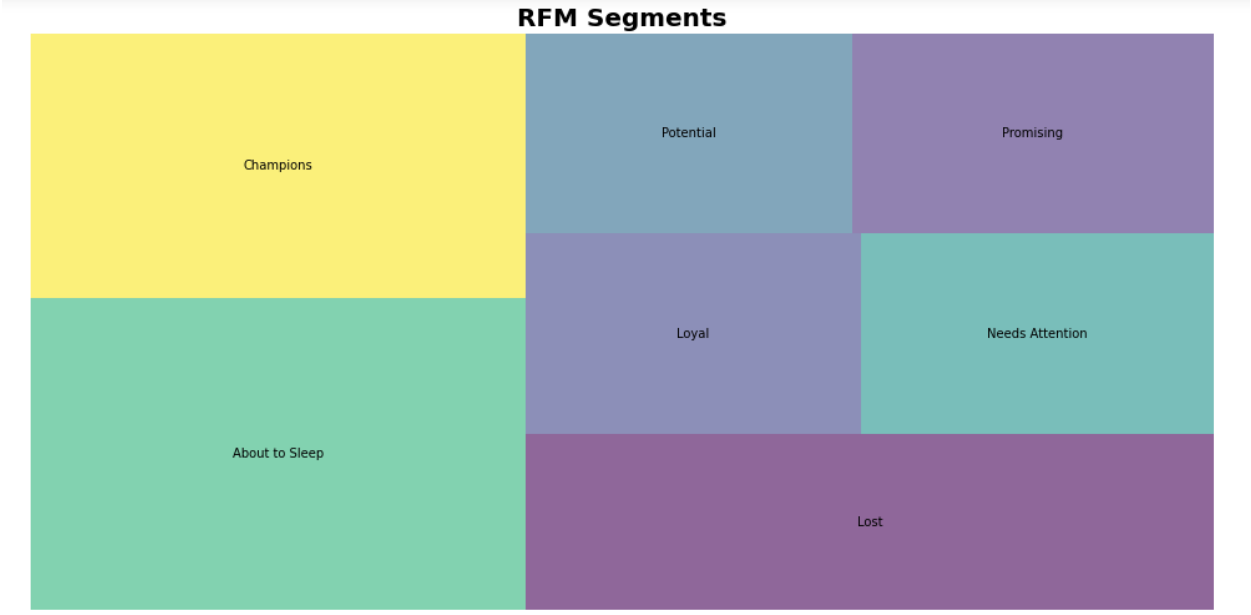
Calculating R F M score as per quartile(percentile) rank 1-4:

- Lesser the value of RecentTransaction, Better Recency
- Greater the value of NumOfTransactions, Better Frequency
- Greater the value of MoneySpent, Better Monetary

Categorizing customers based on RFM_Sum(sum of each R/F/M values):

1. Champions
2. Loyal
3. Potential
4. Promising
5. Needs Attention
6. About to Sleep
7. Lost

Later Calculating mean for each R/F/M within each category and its respective count(density of customers in each category).
Lastly, to visualize the insights we have genareted box plot using squarify library keeping “count” as the unit. The greater the area of each compartment, greater the count of customers in category.
Due to this we can find out, How much % of customers we need to focus on and what are the density of customers within in each segment.


RFM Segments
Champions
Potential
Promising
Loyal
Needs Attention
About to Sleep
Lost