"Did you forget"

As discussed in the case study, Bigbasket is considering developing a recommendation system which suggests specific items that might have been forgotten by a customer. At the check-out, this content-based recommender system recommends products that are similar to the items previously purchased by the customer. The goal of this assignment is to develop the "did you forget" feature on a smaller scale. First, you should find the top 20 members who purchased the most items in total and sort this "members list" from the largest to the smallest. There is another dataset called "Current_Basket.csv" which contains the current items available in the basket of each of the 20 customers. By using "association rules mining", for each of these customers, you should find the top 5 (at most) items that they might have forgotten. The criteria of choosing the top items are the "Support" measure. The final output should be a table with the member ID, recommended items, and the support measure for each of the recommended items.

Answer: At first, we have found out the top 20 members who purchased the most items in total and sorted this "members list" from the largest to the smallest. We have used pivot table to find out the 'Member' with the 'total count of order' then sorted the orders from the largest to the smallest.

Sl. No	Member	Count of Order
1	M38622	1438
2	M33064	1318
3	M41747	1131
4	M32409	1106
5	M31966	1102
6	M56368	1021
7	M36432	1001
8	M41781	920
9	M35538	912
10	M33491	874
11	M48101	824
12	M35649	797
13	M45470	777
14	M43831	744
15	M33558	739
16	M32449	735
17	M52629	716
18	M55932	694
19	M43977	682
20	M78720	682

There is another dataset called “Current_Basket.csv” which contains the current items available in the basket of each of the 20 customers. By using “association rules mining”, for each of these customers, you should find the top 5 (at most) items that they might have forgotten. The criteria of choosing the top items are the “Support” measure. The final output should be a table with the member ID, recommended items, and the support measure for each of the recommended items.

We have imported the “Current_Basket.csv” data set into R to analyse the data and thus applying the “association rules mining”. Please refer to the R code attached to this submission.

```
> str(current_basket)
'data.frame':   20 obs. of  6 variables:
 $ Members..position.in.the.list: int  1 2 3 4 5 6 7 8 9 10 ...
 $ Item1                        : Factor w/ 16 levels "Banana","Beans",..: 12 2 10 7 2 11 8 14 13 5 ...
 $ Item2                        : Factor w/ 17 levels "Almonds","Avalakki / Poha",..: 12 16 2 10 8 4 13 16 16 7 ...
 $ Item3                        : Factor w/ 12 levels "Banana","Brinjals",..: NA 5 6 3 11 2 8 4 1 NA ...
 $ Item4                        : Factor w/ 4 levels "Butter & Cream",..: NA 2 NA 4 NA NA 3 NA NA NA ...
 $ Item5                        : Factor w/ 1 level "Agarbatti": NA NA NA NA NA NA 1 NA NA NA ...
```

	Members..position.in.the.list	Item1	Item2	Item3	Item4	Item5
1	1	Raw Rice	Organic F&V	NA	NA	NA
2	2	Beans	Root Vegetables	Namkeen	Other Vegetables	NA
3	3	Organic Rice & Rice Products	Avalakki / Poha	Organic F&V	NA	NA
4	4	Namkeen	Moong Dal	Cream Biscuits	Whole Spices	NA
5	5	Beans	Ground Coffee	Sooji & Rava	NA	NA
6	6	Other Vegetables	Beans	Brinjals	NA	NA
7	7	Oats	Organic Masalas & Spices	Other Vegetables	Raisins	Agarbatti
8	8	Toothpaste	Root Vegetables	Hair Conditioner	NA	NA
9	9	Snacky Nuts	Root Vegetables	Banana	NA	NA
10	10	Cream Biscuits	Exotic Vegetables	NA	NA	NA
11	11	Besan	Other Juices	Other Dals	NA	NA
12	12	Banana	Whole Spices	Raisins	NA	NA
13	13	Namkeen	Chips	NA	NA	NA
14	14	Ghee	Other Dals	NA	NA	NA
15	15	Banana	Instant Pastas	NA	NA	NA
16	16	Veg & Fruit	Diapers & Wipes	NA	NA	NA
17	17	Banana	Namkeen	Toilet Cleaners	NA	NA
18	18	Organic F&V	Banana	Cream Biscuits	Butter & Cream	NA