2)

Justification for Using **Apriori** to mine association Rules:

We need combos of items which would appear together frequently. The best way of finding this would be to implement Association Rule Mining with one set of items as antecedent and another set as consequent.

There were two strategies given to form combo meals ->

The first strategy was to club together one high rated item with a lower rated one. This was achieved by first doing some initial pre-processing of filtering out the unnecessary fields from the given data. . In the given information, one transaction is uniquely determined by its Bill No and Transaction Date. So these fields were used as the aggregate fields to get the presence absence table of all the items. Then we found out the average ratings for the items, and binned them into two equal partitions. We then use the higher rated category as antecedent and the other category as a consequent for the Apriori Node to find the possible combos (sets of Antecedent, Consequent pairs). This entire process was done for all the 4 months individually and by comparison between the rule sets obtained by the 4 months, we find out the required rules so as to limit the loss% under Z%. The support and confidence percentages were manually set to accommodate the constraint.

The Second Strategy asks us to club the less frequently bought items (i.e. the items with low total quantity) for a combo. For this, we initially did some basic pre-processing, then separated the items whose total quantity does not exceed 500, then performed Association Rule Mining on it using Apriori to find out which among these items were frequently bought together. We understand that this does not affect the total change in revenue much as the number of transactions are comparatively very less. But, on implementing this, we found out that in different months, the association rules found were completely different and there wasn’t a single rule consistently available in all the four months with high support implying that there was no constant association amongst the lower frequently bought items at ANC. Hence, we decided to discard the rules obtained by this approach.

The rules which were finally derived were applied on the December dataset and individual decrease in revenue were found on each of them and finally summed up using a python script (2.py). The results are as follows:-

- The total number of combos generated = 121
- Total number of contributing rules in December = 66
- Total decrease in revenue in the month of December using the generated combos = 79710
- %change in revenue in month of December after applying the combos =6.98%

Some of the most contributing combos were (Butter Naan, Butter Chicken), (Kadhai Panner, Butter Naan), (Plain Maggi, Cheese Toast).

Strengths: Particularly high difference in revenue, indicating that many of the combos were used by the people.

Weakness: Out of the 121 Rules generated, 66 were only contributing. There were 55 non-contributing combos.

The final output is present in the output2.csv file.