Name: Sagar Soni
Intern Id: 180044

Project Activity-1

Instructions:

Bakery Dataset Description

The data set contains 21293 observations and the following columns

1. **Date.** Categorical variable that tells us the date of the transactions (YYYY-MM-DD format). The column includes dates from 30/10/2016 to 09/04/2017.
2. **Time.** Categorical variable that tells us the time of the transactions (HH:MM:SS format).
3. **Transaction.** Quantitative variable that allows us to differentiate the transactions. The rows that share the same value in this field belong to the same transaction, that's why the data set has less transactions than observations.
4. **Item.** Categorical variable containing the products.

Transaction ID ranges from 1 through 9684. However, there are some skipped numbers in Transaction IDs. Also, there are duplicated entries. Download Dataset Here. dataset

Submission instruction:

- Submission on GitHub and link for your GitHub on Moodle.
- Coding conventions are properly used.
- Visualize the result of each question

GitHub Link:

Section: B (Market Basket Analysis) (50 Marks)

1. **Give an example where you can apply the Apriori algorithm.**

Apriori algorithm is a classical algorithm in data mining. It is used for mining frequent item sets and relevant association rules. It is devised to operate on a database containing a lot of transactions, for instance, items brought by customers in a store.

It helps the customers buy their items with ease, and enhances the sales performance of the departmental store.

This algorithm has utility in the field of healthcare as it can help in detecting adverse drug reactions (ADR) by producing association rules to indicate the combination of medications and patient characteristics that could lead to ADRs.

2. **What happens when we decrease the support level? Why?**

Support represents the popularity of that product of all the product transactions. Support of the product is calculated as the ratio of the number of transactions includes that product and the total number of transactions.

Support of the product = (Number of transactions includes that product)/ (Total number of transactions)

When we decrease the support level, we decrease the amount of popularity we need.

3. **What happens when we increase the confidence level? Why?**

This explains how likely Y is purchased when X is purchased. This defines association between two items.
For example, when a person buys milk they are more likely to buy bread as well or vice versa. This is measured by the proportion of transactions with item X, in which item Y also appears. Expressed as {X -> Y}. Calculated by the proportion of the number of transactions in which both (X & Y) occur to support item X. When we increase the confidence level, we tend to increase the probability of that item to be transacted with the compared item.

4. **What recommendations would you give to the owner of the bakery?**

✔ There is a 70% chance that he/she will buy coffee.
✔ Never recommend your customers to buy brownies with bread.