**04th feb’21 ASSIGNMENT-5**

**TEAM MEMBERS :**

* DEVASENA K - 1934006
* SARIKA M – 1934040

**ASSOCIATION RULE MINING**

The main components of association rules are as,

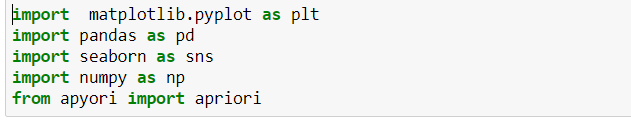
* Support
* Confidence
* Lift

For large sets of data, there can be hundreds of items in hundreds of thousands transactions. The Apriori algorithm tries to extract rules for each possible combination of items

**DATASET DESCRIPTION:**

we will used the Apriori algorithm to find rules that describe associations between different products given 7500 transactions over the course of a week at a French retail store.

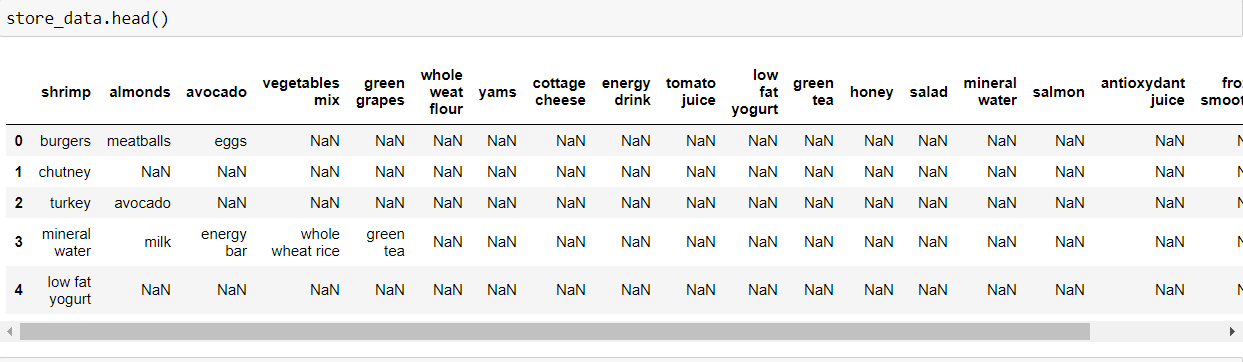
**IMPORTING THE LIBRARIES:**



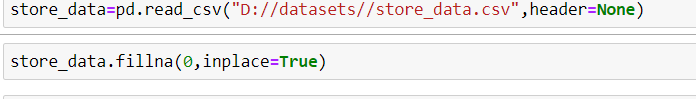
**IMPORTING THE DATASET:**



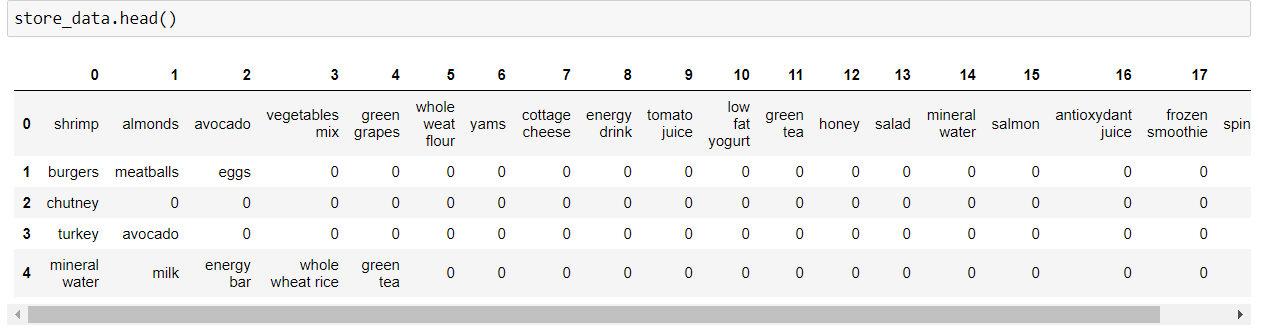
**SAMPLE DATA:**



**TO GET RID OF THE HEADER ROW:**



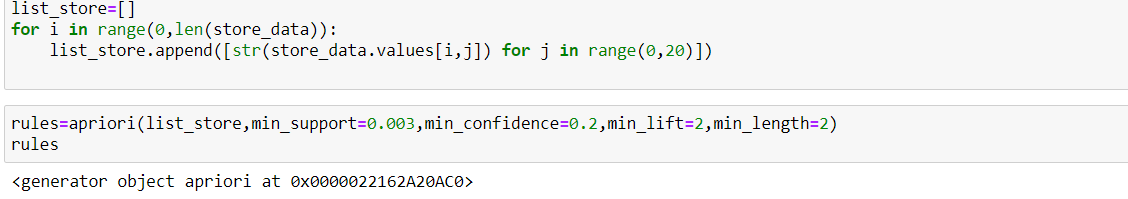
**DATA AFTER REMOVING HEADER :**



Hence the data is ready for applying the APRIORI ALGORITHM to find out which items are commonly sold together, so that owner can place them together in order to increase the profit

**DATA PREPROCESSING:**

Hence for applying apriori algorithm we need to change the pandas dataframe into a list of another list, therefore following code changes it



**APPLY THE APRIORI ALGORITHM ON THE DATASET:**

Algorithm requires some parameters to work on hence

* The first parameter is the list of list that you want to extract rules from.
* **SUPPORT :**

Tells the default popularity of an item and can be calculated by finding number of transactions containing a particular item divided by total number of transactions.

* **CONFIDENCE :**

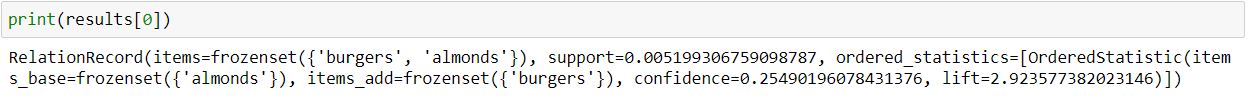
Confidence refers to the likelihood that an item B is also bought if item A is bought

* **LIFT**

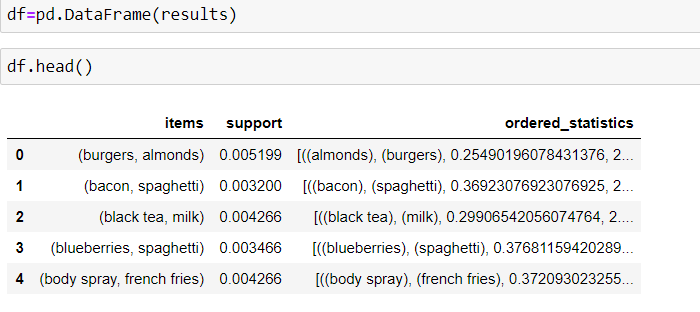
parameter specifies the minimum lift value for the short listed rules.

##### **VIEWING THE RESULTS:**





**TO VIEW IN DATAFRAME:**



**ANALYSIS:**

* The first item in the list is a list itself containing three items. The first item of the list shows the grocery items in the rule.
* For instance from the first item, we can see that **BURGERS** and **ALMONDS** are commonly bought together. This makes sense since people who purchase **BURGERS** are careful about what they eat hence they are more likely to buy **ALMONDS**.
* The support value for the first rule is 0.003. This number is calculated by dividing the number of transactions containing BURGERS divided by total number of transactions.
* The confidence level for the rule is 0.2549 which shows that out of all the transactions that contain BURGER, 25.49% of the transactions also contain ALMONDS.
* Finally, the lift of 2.92 tells us that ALMOND is 2.92 times more likely to be bought by the customers who buy BURGER compared to the default likelihood of the sale of ALMONDS