**Vidya Jyothi Institute of Technology (Autonomous)**

**(Accredited by NBA, Approved by AICTE New Delhi)**

**Aziz nagar Gate, C.B. Post, Hyderabad-500 075.**

**DEPARTMENT OF ARTIFICIAL INTELLIGENCE.**

**Algorithm name:**

## Association Rule Mining via Apriori Algorithm in Python

**Dataset Download link:**

<https://github.com/rakeshrohan-123/Apriori/blob/main/store_data.csv>

**Algorithm explanation:**

Association rule mining is a technique to identify underlying relations between different items. Take an example of a Super Market where customers can buy variety of items. Usually, there is a pattern in what the customers buy. For instance, mothers with babies buy baby products such as milk and diapers. Damsels may buy makeup items whereas bachelors may buy beers and chips etc. In short, transactions involve a pattern. More profit can be generated if the relationship between the items purchased in different transactions can be identified.

For instance, if item A and B are bought together more frequently then several steps can be taken to increase the profit. For example:

1. A and B can be placed together so that when a customer buys one of the product he doesn't have to go far away to buy the other product.
2. People who buy one of the products can be targeted through an advertisement campaign to buy the other.
3. Collective discounts can be offered on these products if the customer buys both of them.
4. Both A and B can be packaged together.

The process of identifying an associations between products is called [association rule mining](https://en.wikipedia.org/wiki/Association_rule_learning).

#### Theory of Apriori Algorithm

There are three major components of Apriori algorithm:

* Support
* Confidence
* Lift

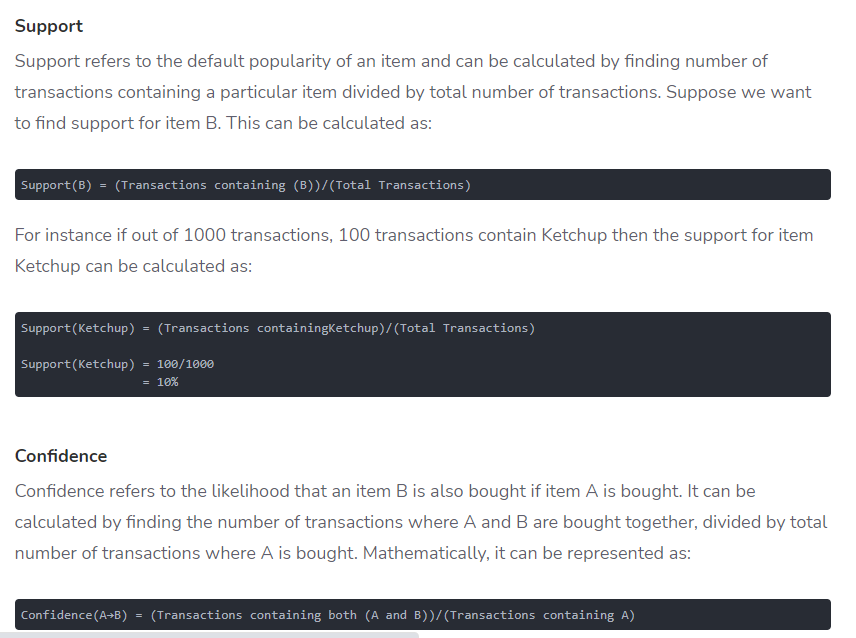
We will explain these three concepts with the help of an example.

Suppose we have a record of 1 thousand customer transactions, and we want to find the Support, Confidence, and Lift for two items e.g. burgers and ketchup. Out of one thousand transactions, 100 contain ketchup while 150 contain a burger. Out of 150 transactions where a burger is purchased, 50 transactions contain ketchup as well. Using this data, we want to find the support, confidence, and lift using mathematical formulas .this formulas are in mathematical part.

Steps involved are:

1. Set a minimum value for support and confidence. This means that we are only interested in finding rules for the items that have certain default existence (e.g. support) and have a minimum value for co-occurrence with other items (e.g. confidence).
2. Extract all the subsets having higher value of support than minimum threshold.
3. Select all the rules from the subsets with confidence value higher than minimum threshold.
4. Order the rules by descending order of Lift.

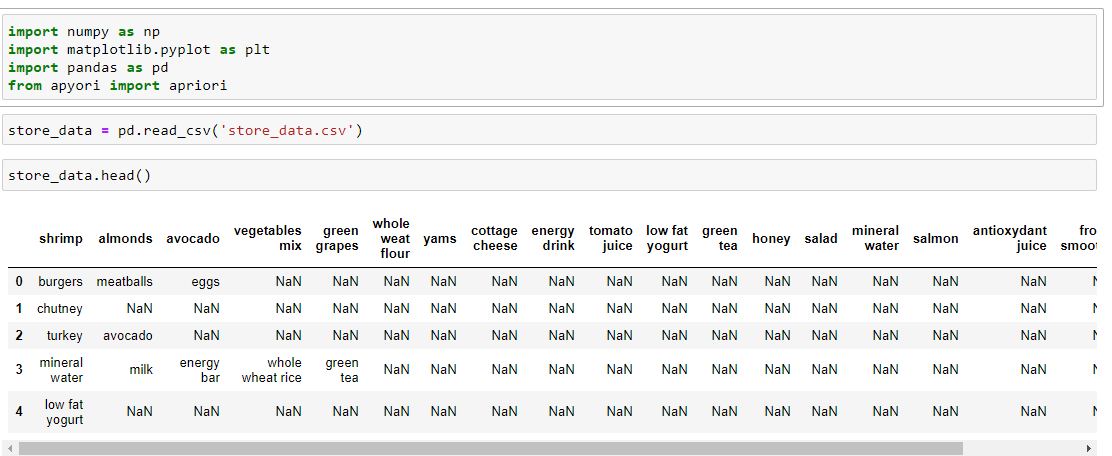
**Mathematical part**



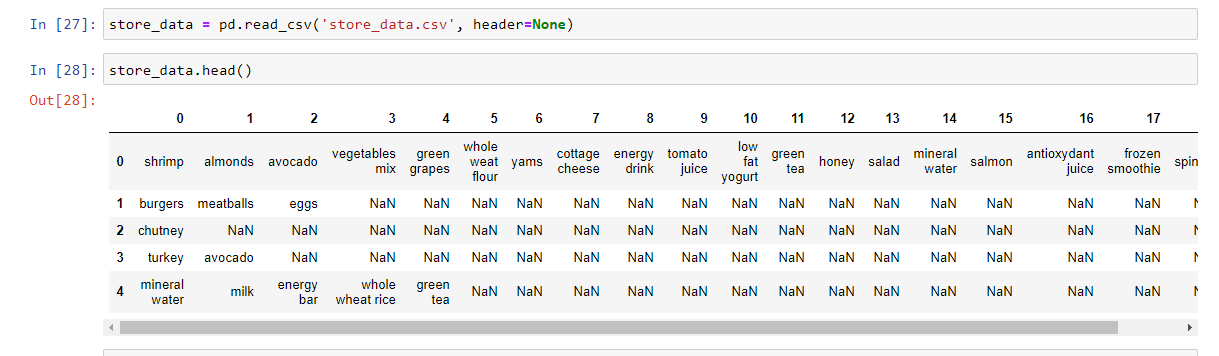
##### Lift

Lift(A -> B) refers to the increase in the ratio of sale of B when A is sold. Lift(A –> B) can be calculated by dividing Confidence(A -> B) divided by Support(B). Mathematically it can be represented as:

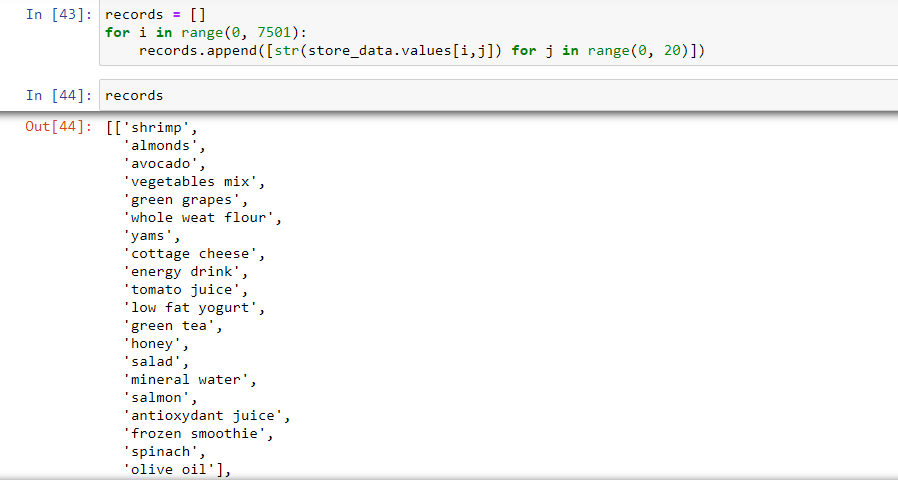
**Code explanation with screen shots:(also try to keep your code in GitHub repository and give link)**



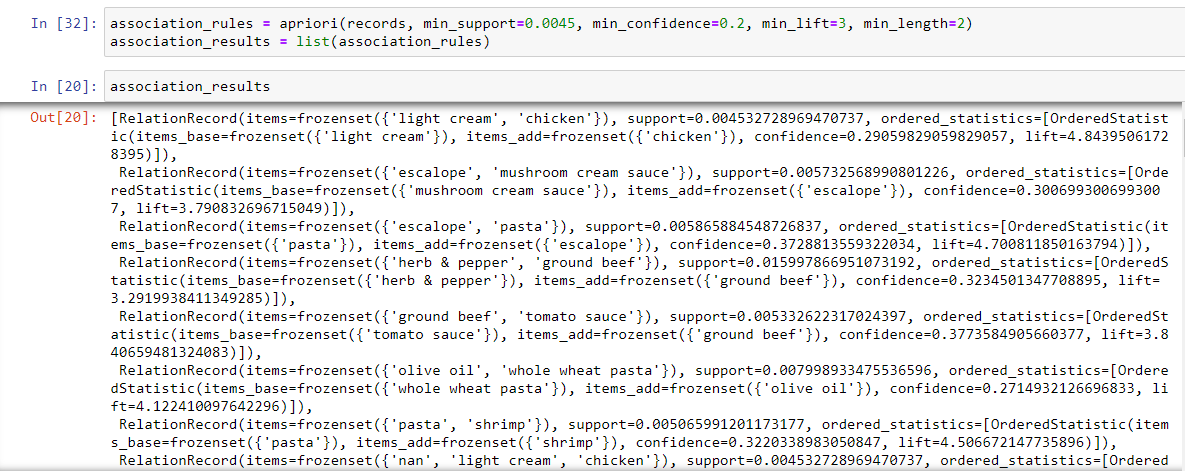
Just imported required packages and loaded dataset in store\_data variable and used head() method to see first five from dataset



As we can see above pic we took header is equal to none because they are not column names

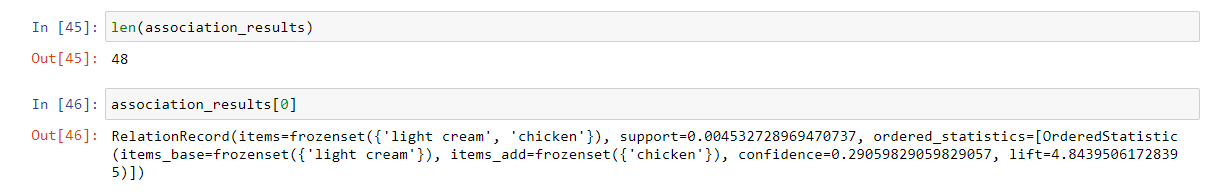


We took empty list as record variable and using for loop we appended all items in to list



The apriori class requires some parameter values to work. The first parameter is the list of list that you want to extract rules from. The second parameter is the min\_support parameter. This parameter is used to select the items with support values greater than the value specified by the parameter. Next, the min\_confidence parameter filters those rules that have confidence greater than the confidence threshold specified by the parameter. Similarly, the min\_lift parameter specifies the minimum lift value for the short listed rules. Finally, the min\_length parameter specifies the minimum number of items that you want in your rules.

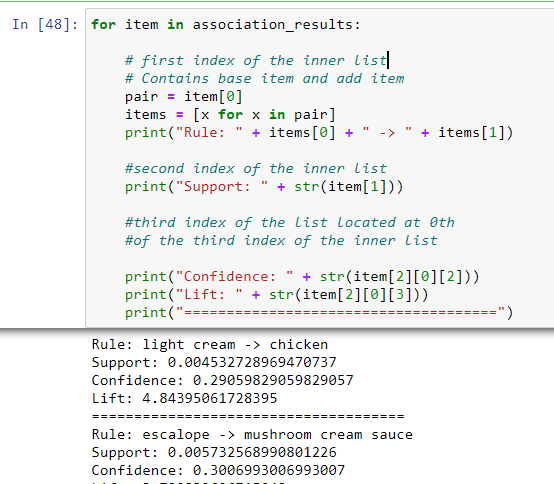
Let's suppose that we want rules for only those items that are purchased at least 5 times a day, or 7 x 5 = 35 times in one week, since our dataset is for a one-week time period. The support for those items can be calculated as 35/7500 = 0.0045. The minimum confidence for the rules is 20% or 0.2. Similarly, we specify the value for lift as 3 and finally min\_length is 2 since we want at least two products in our rules. These values are mostly just arbitrarily chosen, so you can play with these values and see what difference it makes in the rules you get back out.



We just seen length and sawn first element of our list

And then finally to see all relationship between items

The following script displays the rule, the support, the confidence, and lift for each rule in a more clear way:



You can also find full code and dataset in below github url.

<https://github.com/rakeshrohan-123/Apriori/blob/main/Apriori.ipynb>

https://github.com/rakeshrohan-123/Apriori/blob/main/store\_data.csv

**Application or use cases:**

* Association rule mining algorithms such as Apriori are very useful for finding simple associations between our data items. They are easy to implement and have high explain-ability. However for more advanced insights, such those used by Google or Amazon etc., more complex algorithms, such as [recommender systems](https://en.wikipedia.org/wiki/Recommender_system), are used. However, you can probably see that this method is a very simple way to get basic associations if that's all your use-case needs.