**Algorithm used :**

1. Apriori algorithm :

Apriori algorithm is used to calculate the support value, lift value and confidence value.

* 1. Support Value : Support measures the percentage of transactions containing a particular combination of items relative to the total number of transactions.

Formula : (Number of transactions with coffee and toast)/(Total transactions)

* 1. Lift value : Lift is a metric to determine how much the purchase of antecedent influences the purchase of consequent.

Formula : (Purchased item(toast intersection with coffee))/(Purchased item(toast))\*(Purchased item(coffee))

* 1. Confidence Value : Confidence measures the probability of finding a particular combination of items whenever antecedent is bought.

Formula : (Number of transaction with Coffee and Toast)/(Number of transaction with Coffee)

1. Market based analysis :

Market basket analysis (MBA), also known as association-rule mining, is a useful method of discovering customer purchasing patterns by extracting associations or co-occurrences from stores' transactional databases

Typically we can extract the relationship between products in the form of a rule, an example of association rule:

IF {Coffee} THEN {Toast}.

The set of items a customer buys is known as an itemset, and MBA tries to identify relationships from the purchases of itemset. The output of MBA consists of a series of product association rules. From the transaction data extracted from the shopping carts of online retailers or the point of sales system of retail stores, we can use MBA to extract interesting association rules between products.

**Dataset :**

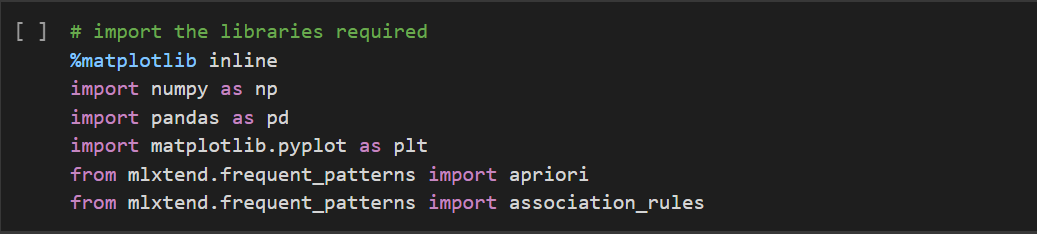
The dataset that is used in this project is publicly available from Kaggle which contains the Transactions data from a bakery from 30/10/2016 to 09/04/2017. The data belongs to a bakery called "The Bread Basket" that serves coffee, bread, muffin, cookies and so on. It is located in the historic center of Edinburgh.

**Tool Used :**

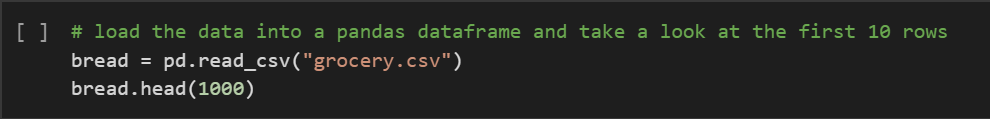
For compiling python code we have used Google colab to execute the programs

**Implentation :**

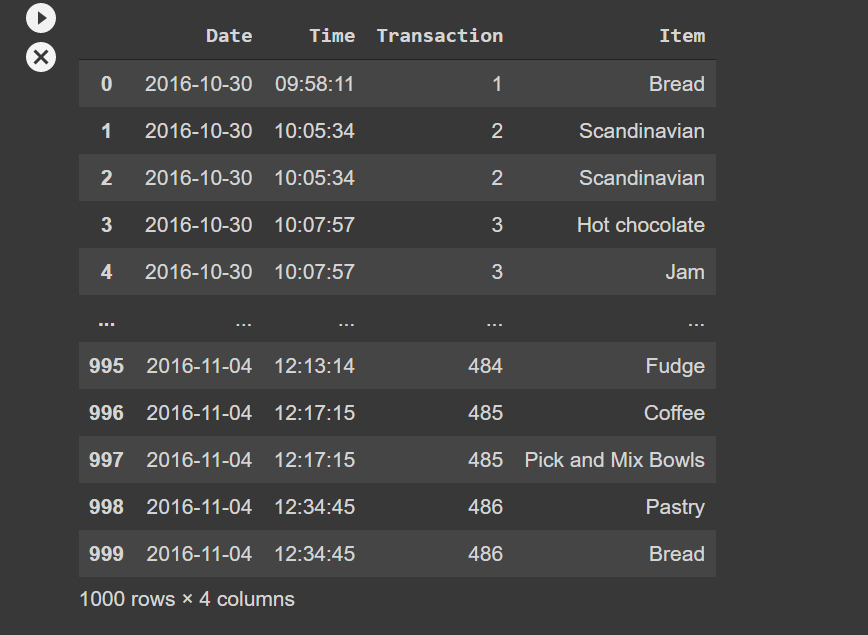
We have imported library numpy which defines numeric python to calculate the numberic values ,pandas for data manipulation calculation, matplotlib to draw the graph of sales and mlxtend frequent patterns which imports the apriori algorithm.



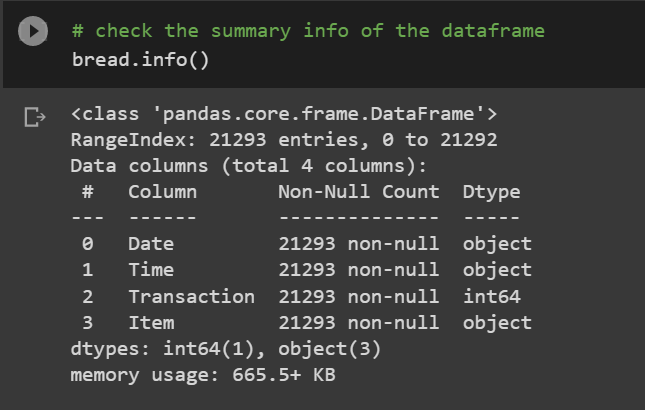
After importing we have imported the data from Kaggle website into the code.



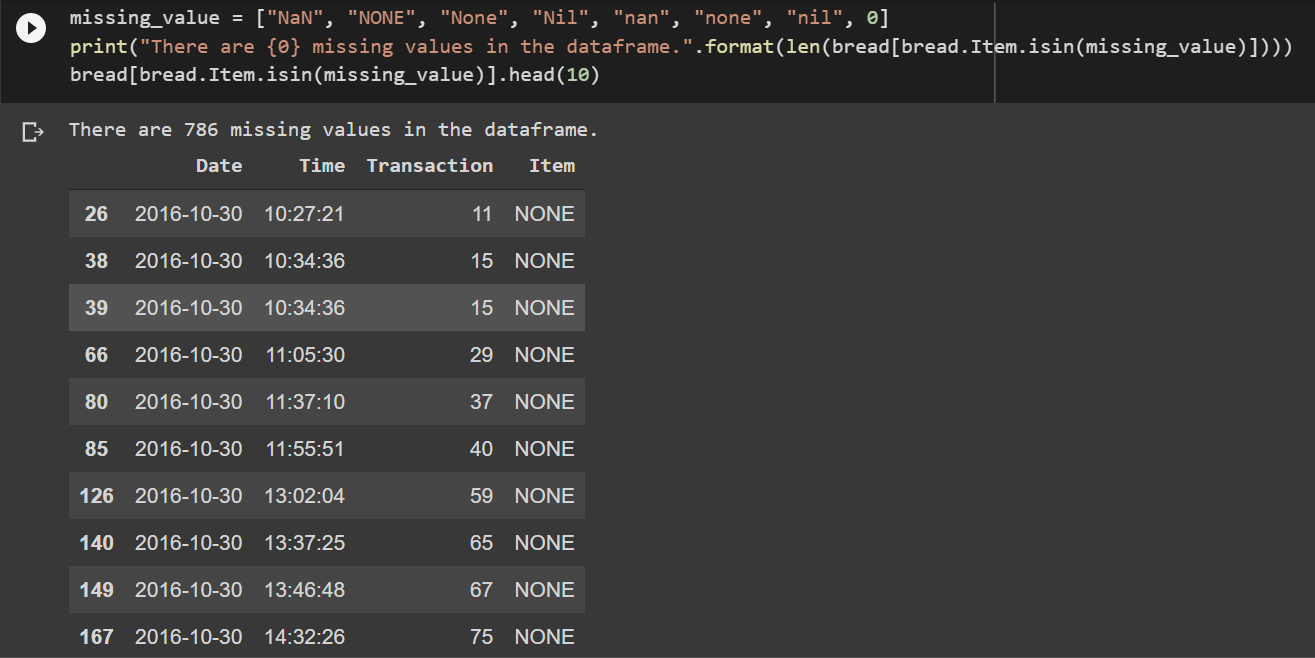
Here are 1000 dataset from as mentioned above in the code.



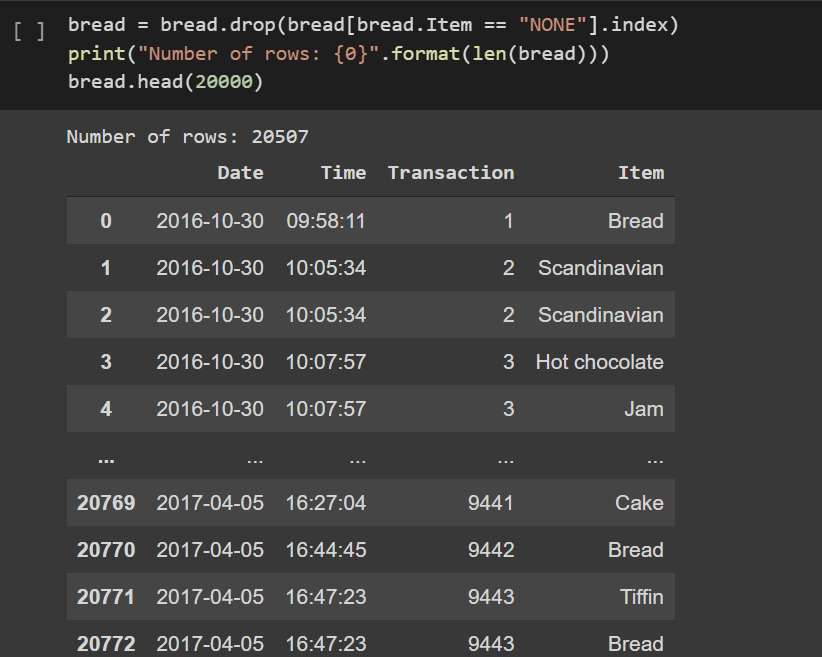
The summary info of the dataset are shown below as a info structure.



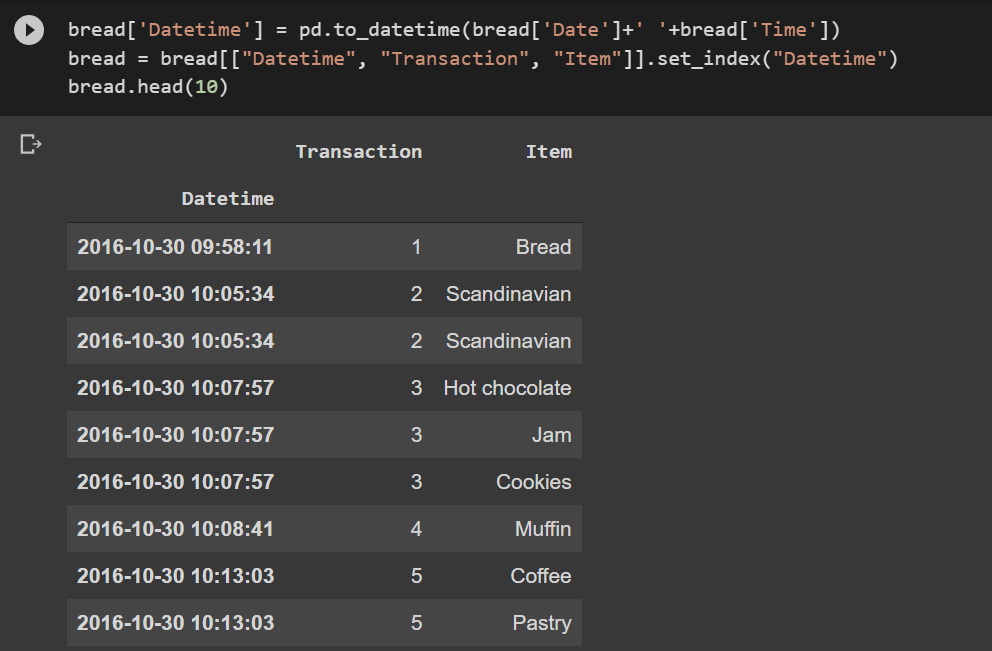
Data preprocessing are shown here here. The following is going to check the None Nan Nil values in the dataset frame. If it occurs then those dataset with the transaction value are displayed below



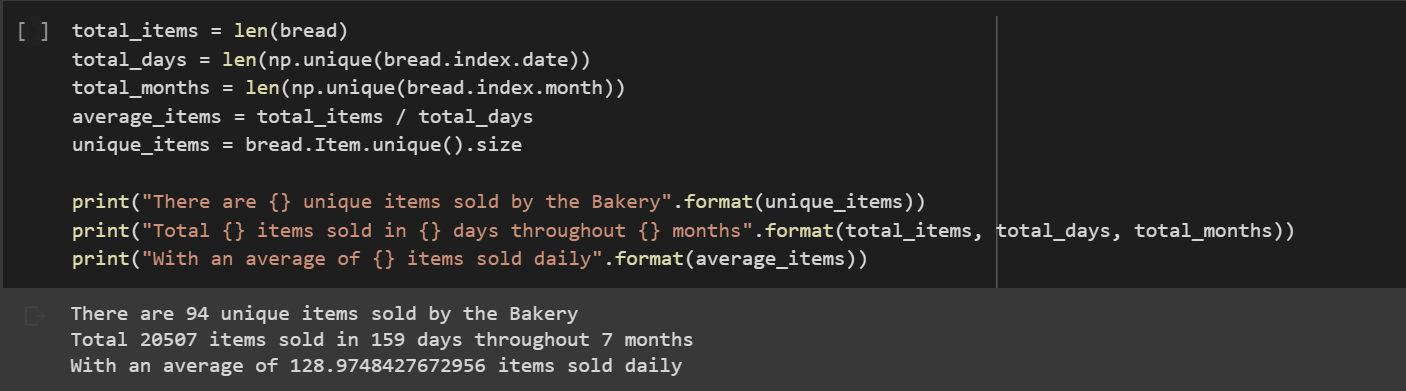
After removing those NONE value we have 20507 dataset value as a preprocessed dataset out of 22543.



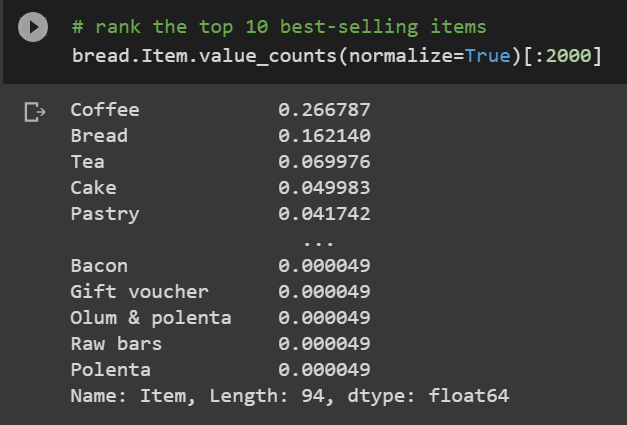
As we see in the above code date and time are shown as different index but now they were merged and shown in single index. It is done because to calculate the MBA process which shows the Graph structure of the sales in the shop.



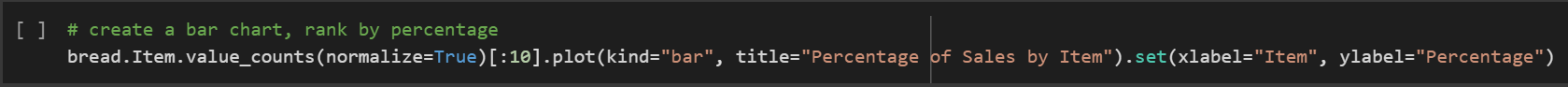
From those 20507 dataset there were 94 unique items, 20507 items are sold in 159 days throughout 7 months with an average of 128.974 items sold daily.

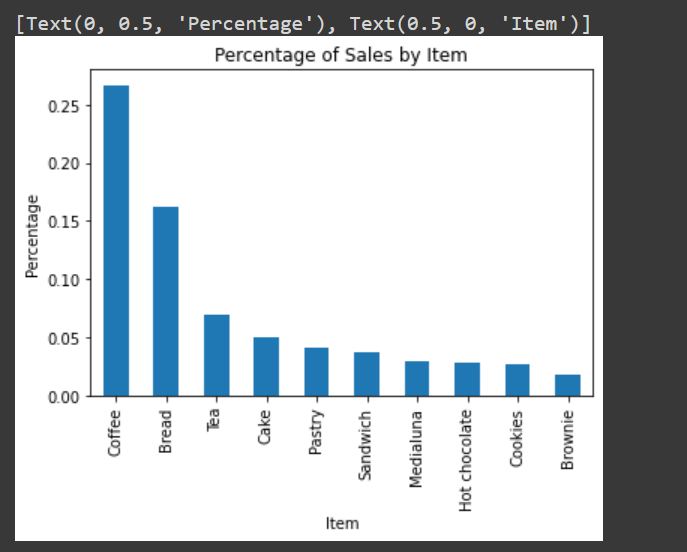


Here are top 10 best selling items in the shop. It is calculated by using count value of the item divided by total number of items in the dataset.

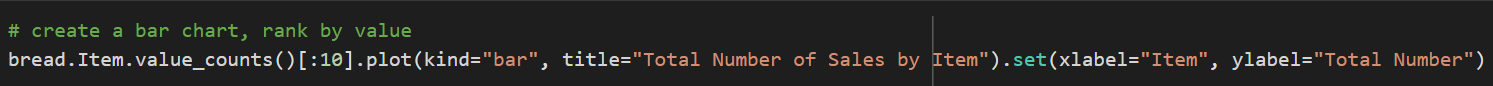


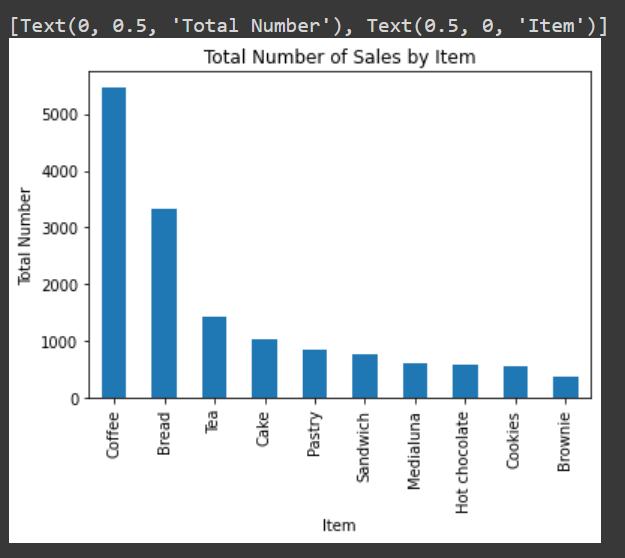
Here is the graph which sows the percentage value of sales by item.



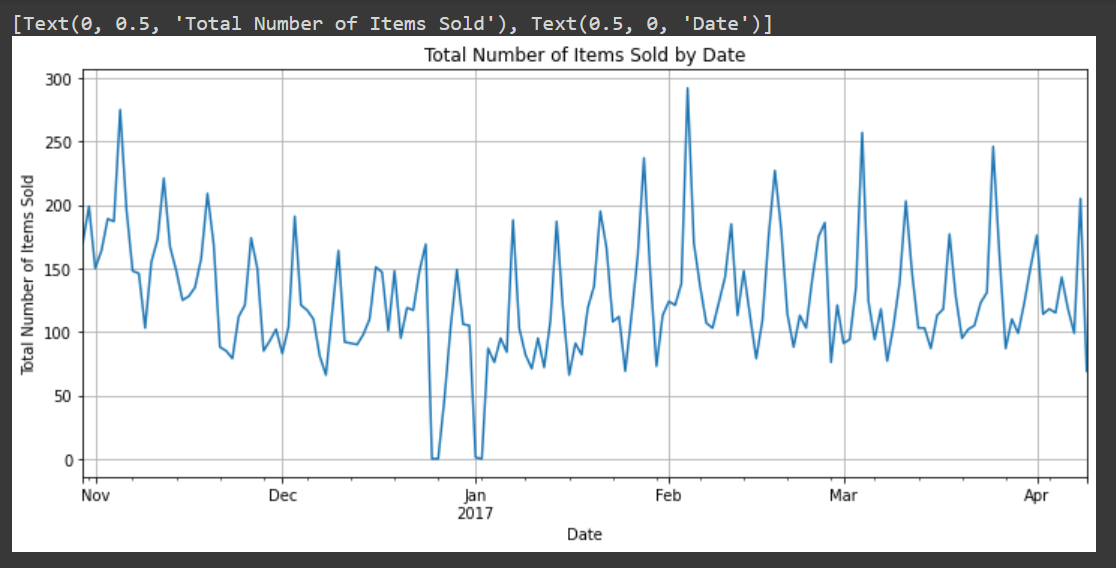
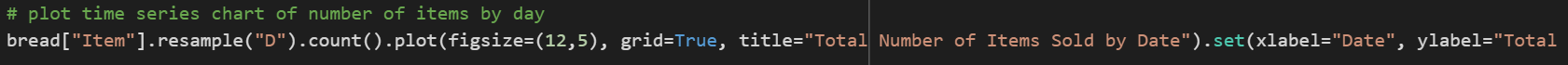


Here is the bar chart which represents the total number of sales of item in the dataset.

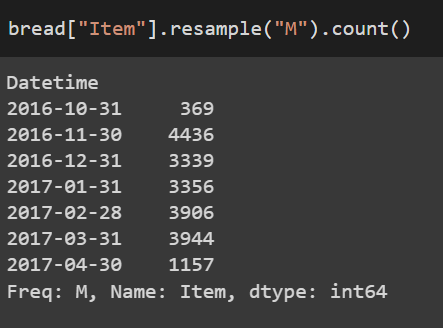




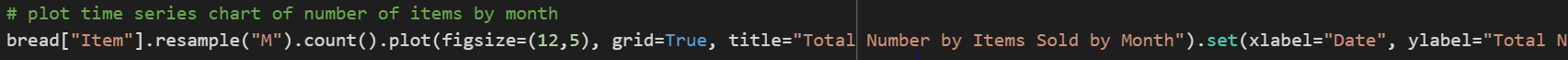
The below graph chart which display the total number of items sold by day to day from the dataset given.

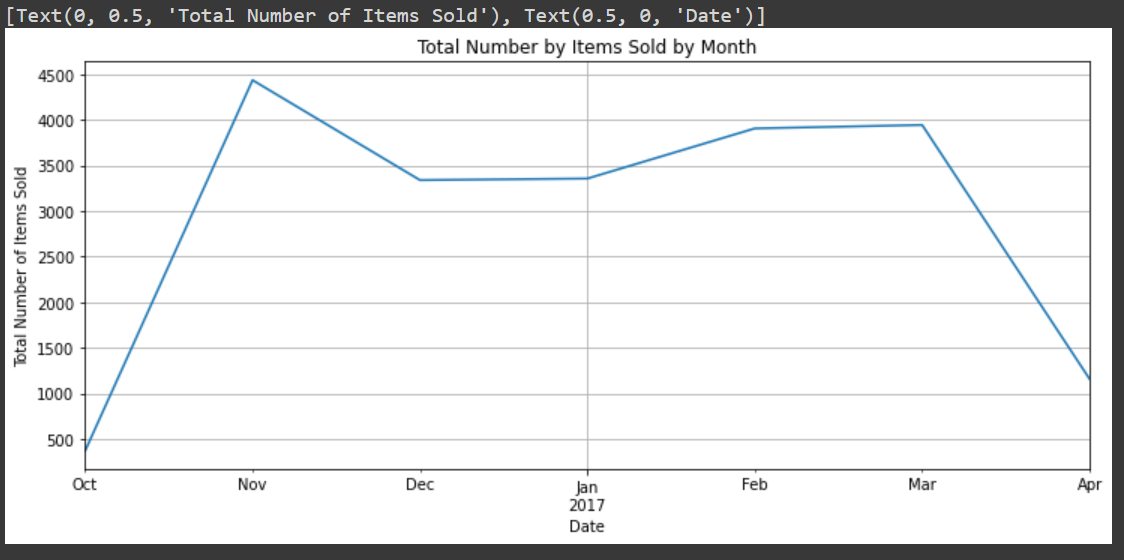


The below snap shows per day sales in the shop

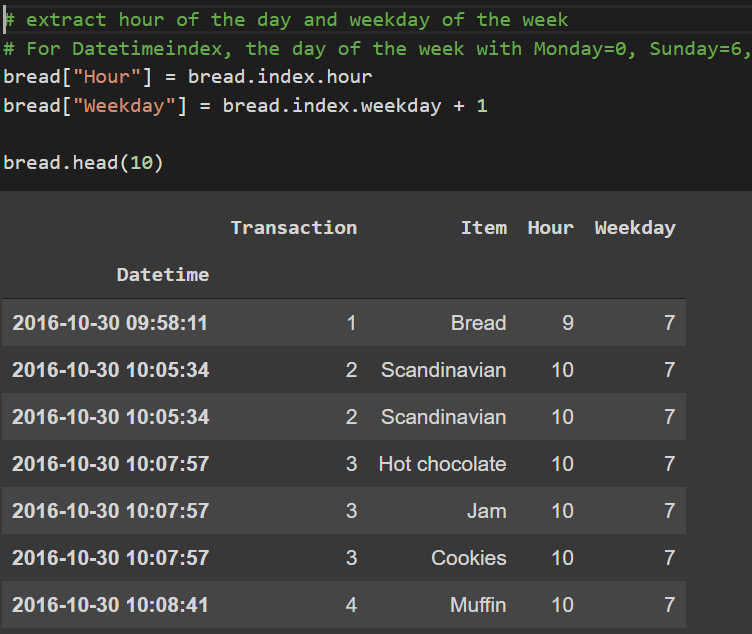


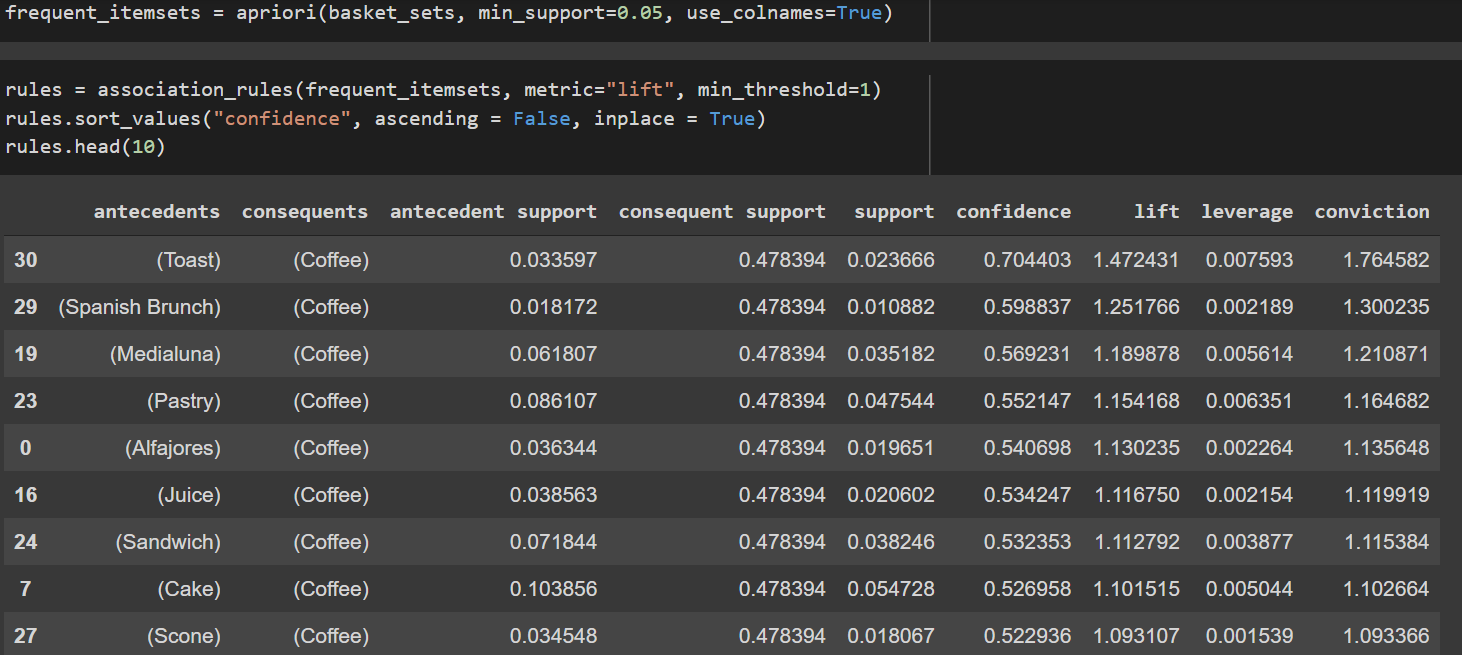
The below graph shows the total number of items sold by month wise from the dataset which we have imported.





From these we have declared the day of the week with Monday=0 and sunday=6. So it keeps on incrementing form monday to sunday as +1 type.



So finally by using Apriori algorithm we have find out the support, confidence and lift value from the given dataset. We set the minimum threshold for lift at 1 and then sort the result by descending confidence value.

The output above shows the Top 10 itemsets sorted by confidence value and all itemsets have support value over 1% and lift value over 1. The first itemset shows the association rule "if Toast then Coffee" with support value at 0.023666 means nearly 2.4% of all transactions have this combination of Toast and Coffee bought together. We also have 70% confidence that Coffee sales happen whenever a Toast is purchased. The lift value of 1.47 (greater than 1) shows that the purchase of Coffee is indeed influenced by the purchase of Toast rather than Coffee's purchase being independent of Toast. The lift value of 1.47 means that Toast's purchase lifts the Coffee's purchase by 1.47 times.

Therefore, we can conclude that there is indeed evidence to suggest that the purchase of Toast leads to the purchase of Coffee. The owner of the bakery "The Bread Basket" should consider bundling Toast and Cofee together as a Breakfast Set or Lunch Set, the staff in the store should also be trained to cross-sell Coffee to customers who purchase Toast, knowing that they are more likely to purchase them together, thereby increasing the store's revenue.