Unlocking Customer Purchase Behavior: Market Basket Analysis using Association Rule Mining

Author Info
Niharikareddy Meenigea
Data Analyst
Virginial International University

Abstract

Market basket analysis (MBA) otherwise called affiliation rule learning or liking examination, is an information mining procedure that can be utilized in different fields, like showcasing, bioinformatics, schooling field, atomic science and so forth. The primary point of MBA in advertising is to give the data to the retailer to comprehend the buy conduct of the purchaser, which can help the retailer in right direction. There are different calculations are accessible for performing MBA. The current calculations work on static information and they don't catch changes in information with time. Be that as it may, proposed calculation mine static information as well as gives a better approach to consider changes occurring in information. This paper talks about the information mining method with the end goal that affiliation rule mining and give another calculation which may accommodating to look at the client conduct and helps with expanding the deals.

Keywords: Data Mining, Business Analysis, KDD, MBA, Bunching, and Grouping.

1. Presentation

Today, the huge measure of information is being kept up with in the data sets in different fields like retail showcases, banking area, clinical field and so on. Yet, it isn't required that the entire data is valuable for the client. That is the reason, extricating the valuable data from enormous measure of data is vital. This course of extricating helpful information is known as information mining or An Information Disclosure and Information (KDD) process. The general course of finding and deciphering designs from information includes many advances like determination, pre handling, change, information mining and understanding. Information mining helps in the business for advertising. Crafted by involving market bushel examination in administration research has been performed by A virtuoso et al. 1 Market container investigation is otherwise called affiliation rule mining. It helps the advertising investigator to comprehend the way of behaving of clients for example which items are being purchased together.

There are different procedures and calculations that are accessible to perform information mining.

1.1. Procedures of Information Mining

There are numerous information mining strategies and calculations are accessible to find significant example and rules.

There are various methods are:

Grouping: In characterization, first look at the elements of recently introduced object and dole out it to a predefined class for instance arrange the acknowledge candidates as low, medium or high gamble.

Affiliation: The principal objective of affiliation is to lay out the connection between things which exist on the lookout.

The normal instances of affiliation demonstrating are Market crate Investigation and strategically pitching programs. The apparatuses utilized for affiliation rule mining are apriori calculation and we ka tool stash. Forecast: In this usefulness, expectation of a few obscure or missing characteristic values in light of other Data. For instance: Figure the deal an incentive for the following week in view of accessible data.8,9

Bunching: In this, Information Mining coordinates information into significant sub-gatherings (bunches) to such an extent that focuses inside the gathering are like one another, and as various as conceivable from the places in different gatherings. It is an unaided order . A powerful unique solo grouping algorithmic methodology for market bin examination has been proposed.

Anomaly Examination: In this, Information Mining is finished to recognize and make sense of exemptions. For instance, in the event of Market Bin Information Examination, anomaly can be some exchange which happens strangely.

1.2. Affiliation Rule Mining

The Fascinating connections can be Affiliation rule digging is valuable for finding fascinating connections concealed in huge informational indexes. In the accompanying model, there are a few exchanges of the shop have been taken as displayed in Table 1. addressed as affiliation rules as displayed underneath:

Milk Æ Spread

The above decide shows that there is major areas of strength for an among milk and spread. It shows that numerous clients purchase milk and spread together. These standards can be useful for retailers to grasp purchasing nature of clients.

Perhaps of the most well known datum mining approaches is to find continuous thing sets from an exchange dataset and infer affiliation rules. 7 The study on affiliation rule mining has been performed by Zhao et al. 11. In this study, various sorts of mining, for example, affiliation rule mining, characterization, grouping and different strategies have been examined. Further two essential measures have been talked about for affiliation rules for example backing and certainty. In this exploration 11 information has given about Apriori series draws near, AIS calculation, Apriori Calculation,

FP-Tree Calculation (Successive Example Tree Calculation), RARM (Fast Affiliation rule Mining) Calculation. Yet, from this multitude of calculations, Apriori is the greatest improvement from past calculations and simple to carry out. Crafted by market crate examination with information mining techniques has been proposed by Andrej Market bushel investigation had been carried out in light of Six Sigma approach. The point of this study was to work on the outcome and change the sigma execution level of the interaction. Overall guideline enlistment (GRI) calculation was utilized in this review to lay out the affiliation rules. Hilage et al.13 has proposed a use of information mining procedures to a s chose business association with exceptional reference for purchasing conduct . The outcome was analyzed in the wake of applying affiliation rule mining strategy, rule acceptance method and apriori calculation. In this way the aftereffects of these three procedures were joined and endeavors were made to figure out the right purchasing conduct of the client. Crafted by separating information utilizing market bushel examination has been proposed by Roaorane et al. Affiliation rule information mining method was utilized. For this they utilized the dataset of store and investigate the everyday exchanges of the market. The fundamental reason for this study was to orchestrate the results of general store in such a manner with the goal that the benefit of grocery store might increment.

The current work for affiliation rule mining in market container examination is MBA in Huge Data set Organization,

MBA in Numerous Store climate, MBA utilizing Quick Calculation. 14

1.3. Anomaly Discovery

As indicated by Hawkins10 who characterizes "an exception is a perception which strays such a great amount from different perceptions as to stimulate doubts that it was created by an alternate system".

Crafted by FP-Exception Incessant example based anomaly identification has performed by He et al. Another strategy for exception recognition by finding incessant example from the informational collection was proposed. An action called FPOF (Successive example exception factor) to identify the anomaly exchanges has characterized and proposed the Track down FPOF calculation to find anomalies. Crafted by exception recognition of Business Insight utilizing information mining procedure has executed by Khan et al. Before this work, the primary focal point of scientists was to establish design from huge datasets which might help uncertainty making. Anyway exception location was not the really engaged area of examination. Consequently this work was the headway in exception location.

In spite of the fact that information mining has become famous as an arising method, still there are a few issues to be made plans to make it valuable in different spaces. A portion of the issues looked by information mining are nature of information, between operability, security and protection and so on. The significant issue with the information mining is its absence of considering the investigation of ongoing information. To pursue the taking an alternate route of information, occasional mining come in presence. Occasional mining alludes to play out the information mining after fixed time span. For instance, a departmental store mines for affiliation controls each quarter to find current buy conduct of clients.

2. Existing Calculation

There are numerous calculations are accessible for affiliation rule mining. Existing calculations work on the static information. They find the great affiliation rules on premise of different measurements, for example, support, certainty, lift and so on. In these calculations, when next time they perform information mining, then, at that point, calculation naturally doesn't catch the progressions in information. For that reason they utilize a one more correlation calculation to follow the adjustment of information.

3. Proposed Calculation

Our proposed calculation additionally performs affiliation rule mining. It chips away at change demonstrating idea. Fundamentally, change displaying is utilized to comprehend the elements of information age process by looking at changes that have occurred in found designs. It chips away at the powerful information and performs occasional mining. Intermittent Mining is really the full grown use of KDD process.

3.1. ARM-Indicator Calculation

This calculation is attempting to catch the changing patterns of exchanges in Market Bushel Examination. It depends on the essential thought of teaming up Affiliation Rule Excavator, Changes in Affiliation Rule Indicator in view of a rationale to get areas of strength for the between the different qualities (i.e the merchandise set in market). The primary purpose is on tracking down the relationship between different things in exchanges. We follow along on the things which are related with high certainty (i.e X→Y, then, at that point, certainty = n(X∩Y)/n(x)). So aftereffect of this calculation will be two arrangements of affiliation rules:

1. Affiliation rules which are profoundly unsurprising for future windows.

2. Exceptions (Affiliation Rules which are least plausible to come in next windows).

Input: Set of Exchanges Result: Anticipated Affiliation Rules, Obsolete Affiliation Rules

3.1.1. Definitions and Particulars

Support (X): Backing of thing is the times a thing happens in exchanges in a data set.

Certainty: Certainty is a term related with affiliation rule, It is characterized numerically as : Certainty =Support(X∩Y)/Support(X)Score (X→Y): It is the worth which is doled out to qualities of relationship based on certainty of that affiliation rule as displayed in Table 2.

Table 2:Score assignment based on their confidence

Confidence in %age	Score Assigned
<=10	0
>10 and <=20	1
>20 and <=30	2
>30 and <=40	4
>40 and <=60	6
>60 and <=90	8
>90	9

3.1.2. Informational collection

For this calculation to run, the informational collection had been taken from Expanded bread shop datasets and store it in 4 windows and calculation work on 2000 exchanges in every window and 26 things, things can be reached out up to n. (Connection to site: https://wiki.csc.calpoly.edu/datasets/wiki/apriori)

3.1.3. Phases of Calculation

Stage 1 : In the main stage, we are having with us double datasets of 4 windows with particulars as made sense of in past segment.

Apriori Calculation : In this part we just run the apriori calculation on the parallel datasets of the multitude of windows and found regular thing sets and further affiliation rules from them.

Stage 2 : This stage can be separated in two sub-parts in which two calculations are run on the other hand.

Section 1 - ARM-Update : This calculation makes Score Table and the construction is displayed in Fig.1(a) and afterward refreshing score table as the information from back to back windows come.

ARM-Update(Windowi, Certainty To ScoreTable, ScoreTable)

{

For (I = beginning of-Windowi ; I < end-of-Windowi ; I ++)

{

N = AssignScore (ith assoction rule, ConfidenceToScoreTable) ;

CreateEntryScoreTable(N,ith affiliation rule);

}

}

where

AssignScore (ith affiliation rule, ConfidenceToScoreTable) : It is a capability which is taking Info some

Affiliation rule and Certainty to ScoreTable and this calculation is utilized with calculation with Section 2 Calculation and supplies Data handled to Section 2 calculation which further cycles the data.

CreateEntryScoreTable (N,ith affiliation rule) : It is a capability which make new passage in the score table on the off chance that some affiliation rule isn't in the ScoreTable or on the other hand on the off chance that present, simply add score N to existing guideline.

Section 2 - ARM-Indicator : This part is pursued we have run ARM-Update Calculation, this calculation find the anomalies based on some edge esteem.

ARM-Exception (ScoreTable) {

for (i=0;i$<$number of months ;i++) {

A = FindUpperRules(Rules above limit);

B = FindLowerRules(Rules underneath limit)//containing exceptions ;

}

FindUpperRules() : It is a capability which is finding set of affiliation rules above edge esteem as displayed underneath in

Table 4.

FindLowerRules() : This calculation find the arrangement of affiliation rules beneath edge esteem as displayed underneath in Table

5. These principles are called as exception.

3.1.4. Exploratory outcomes

3.1.5. (a) Particulars

ARM-Update Calculation

Input : Window, Look-Into Table Result : Score Table

where

Window: It contains affiliation rules for some specific time span

Look-Into Table: It contains Certainty to relating Score Values

Score Table : Affiliation rules along lines and their properties in segments with their scores

ARM-Indicator Calculation

Input : Score Table Result : Exceptions

Where

Score Table : Affiliation rules along lines and their traits in segments with their scores

Anomalies: Set of Affiliation Rules which are above score-edge, Set of Affiliation Rules which are beneath score-limit

3.1.5. (b) Focuses for examining results

The outcomes underneath are displayed all together as follows:

1. In the Score Table as displayed in Fig 1(a), the traits are saved column wise at the top and named them a,b,c and

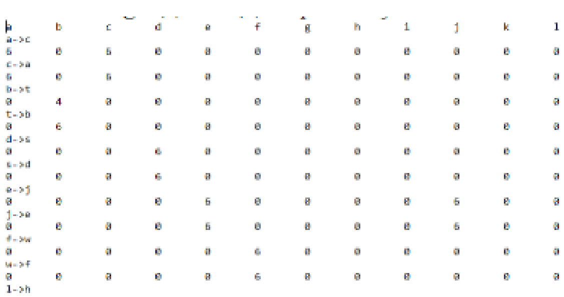
so on for the straightforwardness of the exchange. a,b,c, etc are the things which are kept in market crate. As displayed in the Fig. 1(a), after the top line of all out number of things, there are affiliation rules with their appointed score.

2. Upper Affiliation Rules, significance rules which are above edge are being printed.

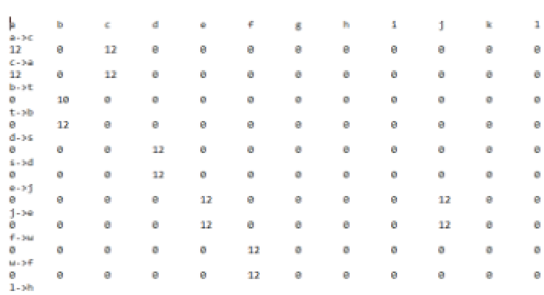
3. Lower Affiliation Rules, meaning the affiliation rules which are beneath limit.

3.1.5. (c) Results with data set

i) Score Table after first month and after second month which follows the change in data from previous month as shown in Fig 1(a) and (b) respectively.



a	b	c	d	e	f	g	h	i	j	k	l
a->c											
6	0	6	0	0	0	0	0	0	0	0	0
c->a											
6	0	6	0	0	0	0	0	0	0	0	0
b->t											
0	4	0	0	0	0	0	0	0	0	0	0
t->b											
0	6	0	0	0	0	0	0	0	0	0	0
d->s											
0	0	0	6	0	0	0	0	0	0	0	0
s->d											
0	0	0	6	0	0	0	0	0	0	0	0
e->j											
0	0	0	0	6	0	0	0	0	6	0	0
j->e											
0	0	0	0	6	0	0	0	0	6	0	0
f->w											
0	0	0	0	0	6	0	0	0	0	0	0
w->f											
0	0	0	0	0	6	0	0	0	0	0	0
l->h											
0	0	0	0	0	0	0	6	0	0	0	6



a	b	c	d	e	f	g	h	i	j	k	l
a->c											
12	0	12	0	0	0	0	0	0	0	0	0
c->a											
12	0	12	0	0	0	0	0	0	0	0	0
b->t											
0	10	0	0	0	0	0	0	0	0	0	0
t->b											
0	12	0	0	0	0	0	0	0	0	0	0
d->s											
0	0	0	12	0	0	0	0	0	0	0	0
s->d											
0	0	0	12	0	0	0	0	0	0	0	0
e->j											
0	0	0	0	12	0	0	0	0	12	0	0
j->e											
0	0	0	0	12	0	0	0	0	12	0	0
f->w											
0	0	0	0	0	12	0	0	0	0	0	0
w->f											
0	0	0	0	0	12	0	0	0	0	0	0
l->h											
0	0	0	0	0	0	0	12	0	0	0	12

Fig. 1 Score table (a) after first month and (b) second month transaction

iii) Score table after third month which follows the change in data from second month and after fourth month which follows the change in data from third month as shown in Fig 2 (a) and (b) respectively.

	a	b	c	d	e	f	g	h	i	j	k	l
a->c	18	0	18	0	0	0	0	0	0	0	0	0
c->a	18	0	18	0	0	0	0	0	0	0	0	0
b->t	0	16	0	0	0	0	0	0	0	0	0	0
t->b	0	18	0	0	0	0	0	0	0	0	0	0
d->s	0	0	0	18	0	0	0	0	0	0	0	0
s->d	0	0	0	18	0	0	0	0	0	0	0	0
e->j	0	0	0	0	18	0	0	0	0	18	0	0
j->e	0	0	0	0	18	0	0	0	0	18	0	0
f->w	0	0	0	0	0	18	0	0	0	0	0	0
w->f	0	0	0	0	0	18	0	0	0	0	0	0
l->h	0	0	0	0	0	0	0	18	0	0	0	18

	a	b	c	d	e	f	g	h	i	j	k	l
a->c	18	0	18	0	0	0	0	0	0	0	0	0
c->a	18	0	18	0	0	0	0	0	0	0	0	0
b->t	0	16	0	0	0	0	0	0	0	0	0	0
t->b	0	24	0	0	0	0	0	0	0	0	0	0
d->s	0	0	0	24	0	0	0	0	0	0	0	0
s->d	0	0	0	24	0	0	0	0	0	0	0	0
e->j	0	0	0	0	24	0	0	0	0	24	0	0
j->e	0	0	0	0	24	0	0	0	0	24	0	0
f->w	0	0	0	0	0	24	0	0	0	0	0	0
w->f	0	0	0	0	0	24	0	0	0	0	0	0
l->h	0	0	0	0	0	0	0	24	0	0	0	24

Fig. 2. Score Table (a) after third month and (b) after fourth month

Outlier Detection

iv) After the fourth month rules as shown in Table 3 we perform outlier detection, at the threshold value = 20 then it divide the rules into two parts upper association rule as shown in Table 4 and lower association rules as shown in Table 5. Lower association rules are known as outliers

Table 3.Fourth month association rules

Association Rules	Score Assigned	
	A	C
a->c	18	18
	C	C
c->a	18	18
	B	T
b->t	16	16
	T	B
t->b	24	24

.

	D	S
d->s	24	24
	S	D
s->d	24	24
	E	J
e->j	24	24
	J	E
j->e	24	24
	F	W
f->w	24	24
	W	F
w->f	24	24
	L	H
l->h	24	24
	P	H

Table 4. Upper Association rules

Association Rules	Score Assigned	
	T	B
t->b	24	24
	D	S
d->s	24	24
	S	D
s->d	24	24
	E	J
e->j	24	24
	J	E
j->e	24	24
	F	W
f->w	24	24
	W	F
w->f	24	24
	L	H
l->h	24	24

Table 5. Lower association rules (Outliers)

Association Rules	Score Assigned	
	A	C
a->c	18	18
	C	C
c->a	18	18
	B	T
b->t	16	16
	X	Y
x->y	0	0

Code and outputs

We have provided the dataset and market basket analysis project code that will be required in this project. We will require a csv file & code for this project. Please download the dataset and source code from the following link: **Market Basket Analysis Project**

Steps to Implement

1. Import the modules and the libraries. For this project, we are importing the libraries numpy, pandas, and sklearn. Here we also read our dataset and we are saving it into a variable.

```
import numpy as np #importing the numpy library we will need in this project
import pandas as pd #importing the pandas library we will need in this project
from mlxtend.frequent_patterns #import apriori #importing our apriori algorithm from mlxtend
from mlxtend.frequent_patterns import association_rules #import association from mlxtend.frequent
import time #importing the time function from python library
df = pd.read_csv(‘dataset.csv') #importing our dataset Online RetailShop Germany which is a csv file
df.head() #printing the head of our dataset file
```

2. Installing the mlxtend library

```
!pip install mlxtend
```

3. Here we are reading our dataset. We are also removing the spaces and checking if any value is missing.

```
missing_value = ["NaN", "NONE", "None", "nan", "none", "n/a", "na", " "] #making an array for possible keyword of missing values
df = pd.read_csv('OnlineRetailShopGermany.csv', na_values = missing_value) #read our dataset file and passing our missing value array
print (df.isnull().sum()) #here we are checking if there is any null value in the dataset.
```

4. Here we are using the strip function to strip the text in the description column.

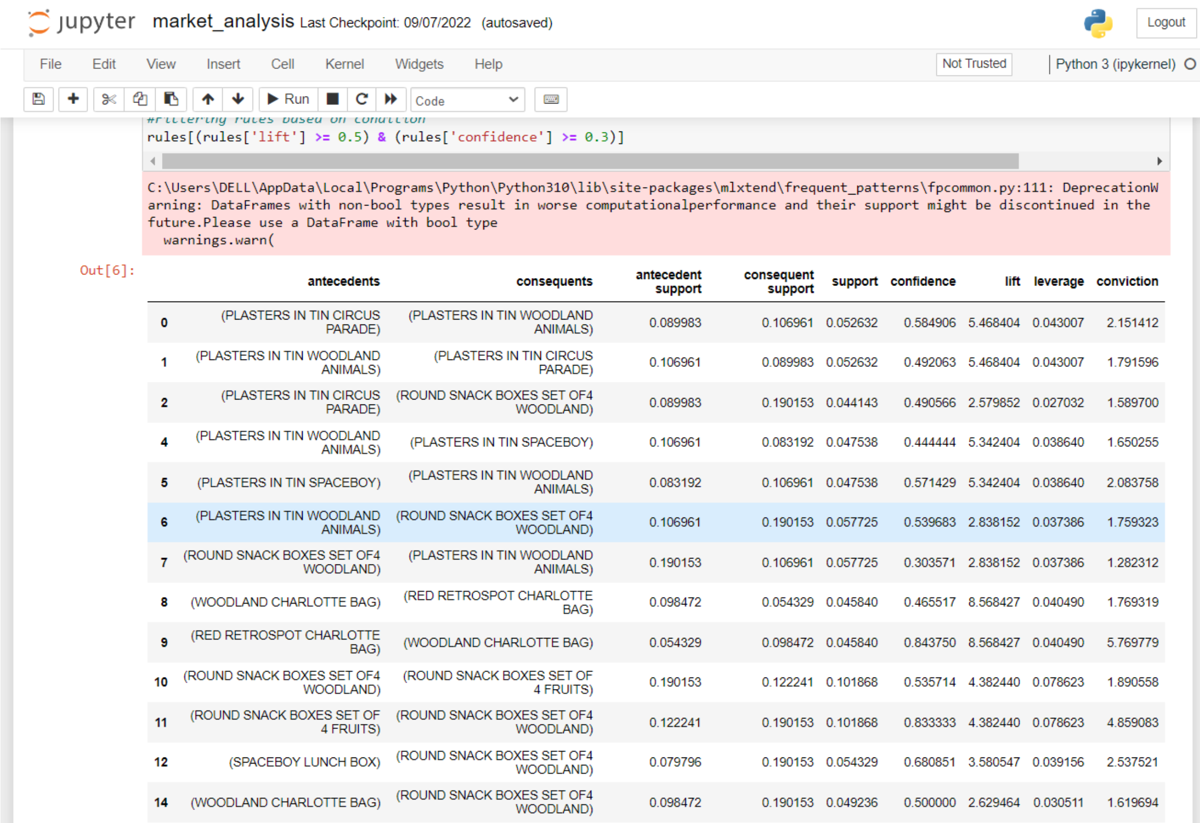
```
df['Description'] = df['Description'].str.strip() #here we are using the strip fnction of string in python to strip the Description column
df.Description.value_counts(normalize=True)[:10] #here we are printing the count of the description value.
```

5. Here we are doing pre-processing our dataset. After that, we are selecting duplicate rows excerpt first occurrence based on all columns.

```
df.drop(df[df['Description'] == 'POSTAGE'].index, inplace = True) #here we are dropping the Description column and we are passing the inplace = True
# Select duplicate rows except first occurrence based on all columns
duplicateRows = df[df.duplicated()] #selecting the duplicate columns
# print(duplicateRows.head())
df = df.drop_duplicates() #dropping the duplicates columns in our dataset
```

6. Here we are sorting the dataset according to Quantity using the groupby function.

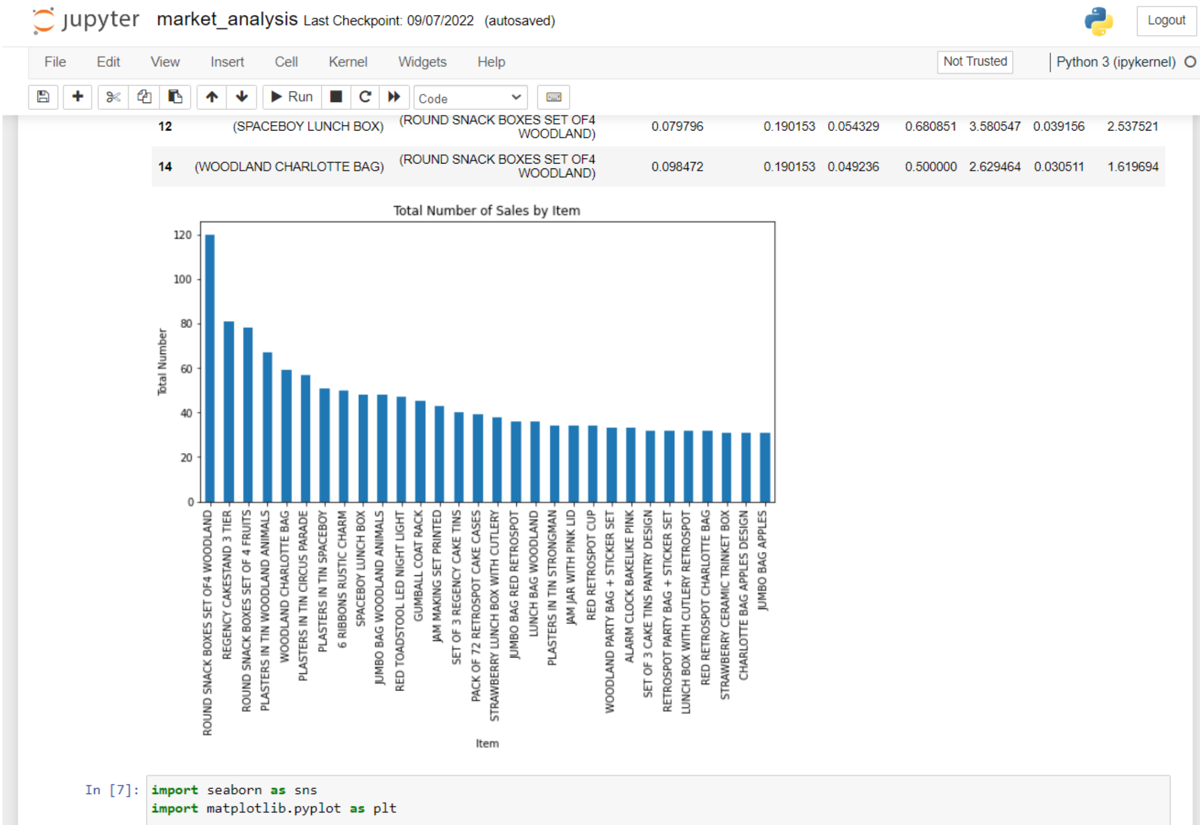
```
df2 = (df.groupby(['InvoiceNo', 'Description'])['Quantity']
.sum().unstack().reset_index().fillna(0)
.set_index('InvoiceNo')) #sorting the dataset according to Quantity and in this we are using the groupby function.
```



	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(PLASTERS IN TIN CIRCUS PARADE)	(PLASTERS IN TIN WOODLAND ANIMALS)	0.089983	0.106961	0.052632	0.584906	5.468404	0.043007	2.151412
1	(PLASTERS IN TIN WOODLAND ANIMALS)	(PLASTERS IN TIN CIRCUS PARADE)	0.106961	0.089983	0.052632	0.492063	5.468404	0.043007	1.791596
2	(PLASTERS IN TIN CIRCUS PARADE)	(ROUND SNACK BOXES SET OF4 WOODLAND)	0.089983	0.190153	0.044143	0.490566	2.579852	0.027032	1.589700
4	(PLASTERS IN TIN WOODLAND ANIMALS)	(PLASTERS IN TIN SPACEBOY)	0.106961	0.083192	0.047538	0.444444	5.342404	0.038640	1.650255
5	(PLASTERS IN TIN SPACEBOY)	(PLASTERS IN TIN WOODLAND ANIMALS)	0.083192	0.106961	0.047538	0.571429	5.342404	0.038640	2.083758
6	(PLASTERS IN TIN WOODLAND ANIMALS)	(ROUND SNACK BOXES SET OF4 WOODLAND)	0.106961	0.190153	0.057725	0.539683	2.838152	0.037386	1.759323
7	(ROUND SNACK BOXES SET OF4 WOODLAND)	(PLASTERS IN TIN WOODLAND ANIMALS)	0.190153	0.106961	0.057725	0.303571	2.838152	0.037386	1.282312
8	(WOODLAND CHARLOTTE BAG)	(RED RETROSPOT CHARLOTTE BAG)	0.098472	0.054329	0.045840	0.465517	8.568427	0.040490	1.769319
9	(RED RETROSPOT CHARLOTTE BAG)	(WOODLAND CHARLOTTE BAG)	0.054329	0.098472	0.045840	0.843750	8.568427	0.040490	5.769779
10	(ROUND SNACK BOXES SET OF4 WOODLAND)	(ROUND SNACK BOXES SET OF 4 FRUITS)	0.190153	0.122241	0.101868	0.535714	4.382440	0.078623	1.890558
11	(ROUND SNACK BOXES SET OF 4 FRUITS)	(ROUND SNACK BOXES SET OF4 WOODLAND)	0.122241	0.190153	0.101868	0.833333	4.382440	0.078623	4.859083
12	(SPACEBOY LUNCH BOX)	(ROUND SNACK BOXES SET OF4 WOODLAND)	0.079796	0.190153	0.054329	0.680851	3.580547	0.039156	2.537521
14	(WOODLAND CHARLOTTE BAG)	(ROUND SNACK BOXES SET OF4 WOODLAND)	0.098472	0.190153	0.049236	0.500000	2.629464	0.030511	1.619694

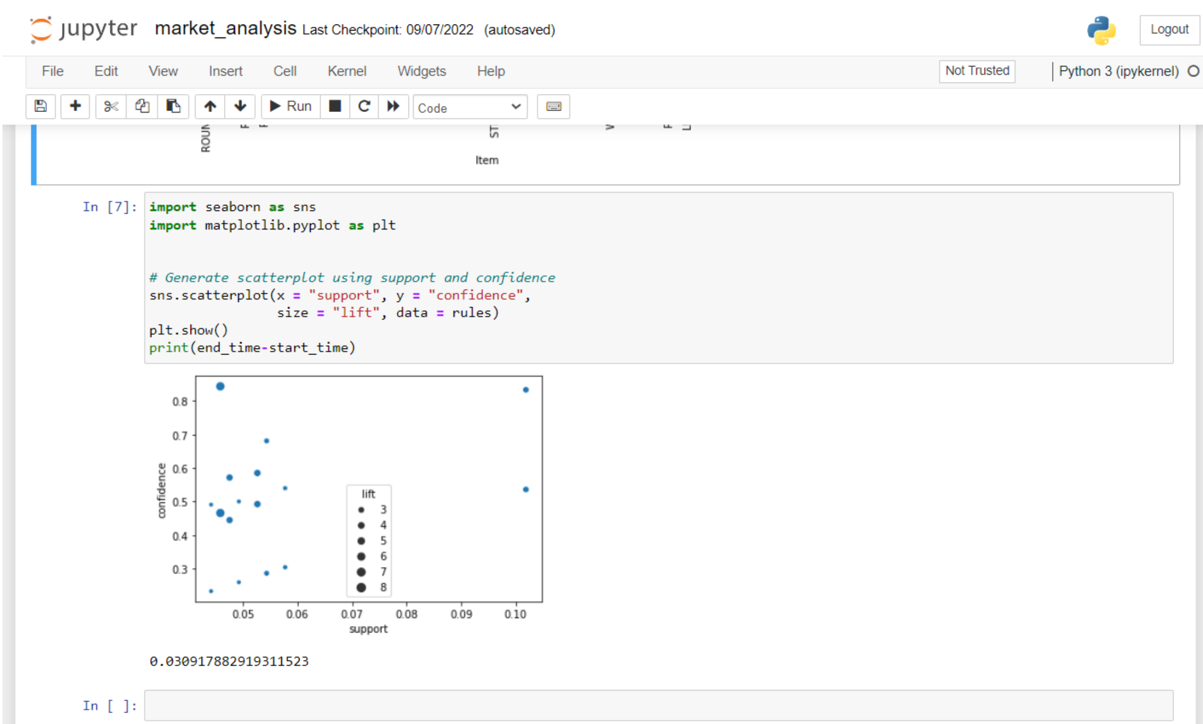
7. Here, we are defining the function convertToZeroOne. In this, we are passing the x values, and we are checking if x<=0 then we are converting it to 0, i.e. converting the negative values to 0 and the non-negative values to 1. Finally, we plot the curve.

```
def convertToZeroOne(x): #defining our converting function
if x <= 0: #checking if x is less than or equal to zero
return 0 #returning
if x >= 1: #if x=1
return 1 # we are returning 1
df3 = df2.applymap(convertToZeroOne) #we are using the function convertToZeroOne
start_time = time.time() #calculating the start time
frequent_itemsets = apriori(df3, min_support=0.04, use_colnames=True)#using the apriori algorithm and passing the dataset
end_time = time.time() #calculating the end time
Frequent_itemsets #printing the dataset.
rules = association_rules(frequent_itemsets, metric="lift", min_threshold=1) #using the association rules function and passing the dataset to it.
rules.head()#printing the rules dataset.
rules[(rules['lift'] >= 0.5) & (rules['confidence'] >= 0.3)] #checking if confidence >=0.3 and lift >=0.5
```



8. Importing the seaborn library and matplot library to plot a scatter plot graph.

import seaborn as sns#importing the seaborn library
import matplotlib.pyplot as plt#importing the matplotlib
sns.scatterplot(x = "support", y = "confidence",
size = "lift", data = rules)#using sns to plot a scatter plot curve
plt.show() #plotting the graph



4. Conclusion

At present many data mining algorithms have been developed and applied on variety of practical problems. However periodic mining is a new approach in data mining which has gained its significance these days. This field is evolving due to needs in different applications and limitations of data mining. This would enhance the power of existing data mining techniques. Finding out the patterns due to changes in data is in itself an interesting area to be explored. It may helpful in x Find out interesting patterns from large amount of data.

x Automatically track the changes in facts from previous data; due to this feature it may be helpful in fraud detection.

x Predicting future association rules as well as gives us right methodology to find out outliers.

Authors suggested that, some areas are still there which need to be focused on. Firstly, results have influenced greatly by the manual threshold values for score, so it is needed to automate the threshold values for better recognition of outliers. Secondly, this approach is specifically targeted at Market Basket Data, it may perhaps be extended to other areas.

5. References

1. Raorane AA, Kulkarni RV, Jitkar BD. Association Rule – Extracting Knowledge Using Market Basket Analysis Research Journal of Recent Sciences 2012:1(2):19-27.

2. Verma Sheenu, Bhatnagar Sakshi. An Effective Dynamic Unsupervised Clustering Algorithmic Approach for Market Basket Analysis International Journal of Enterprise Computing and Business Systems 2014:4(2).

3. Maurizio Marek. Data Mining Concepts and Techniques. E-Commerce Winter 2011.

4. Herman A, Forcum LE, Joo Harry. Using Market Basket Analysis in Management Research Journal of Management 2013:39(7):1799-1824.

5. Saurkar Anand V, Bhujade V, Bhagat P, Khaparde A. A Review Paper on various Data Mining Techniques. International Journal of Advanced Research in Computer Science and Software Engineering 2014:4(4):98-101.

6. Kaur Paramjit, Attwal Kanwalpreet S. Data Mining: Review. International Journal of Computer Science and Information Technologies 2014:5(5):6225-6228.

7. Wu X, Kumar V, Quilan JR., Ghosh J, Yang Q, Motoda H. Top 10 Algorithms in Data Mining.Springer-Verlay London Limited 2007:14:1-37.

8. Ngai EWT, Xiu Li, Chau DCK. Application of Data Mining Techniques in Customer Relationship Managemen t :A Literature Review and Classification Elsevier-Expert Systems with Applications 2009:36:2592-2602.

9. Ramageri Bharati M. Data Mining Techniques and Applications. Indian Journal of Computer Science and Engineering:1(4):301-305.

10. Hawkins D. Identification of Outliers Chapman and Hall 1980.

11. Zhao Quiankun, Bhowmick Sourav. Association Rule Mining: A Survey Technical Report CAIS Nanyan Technological University, Singapore 2003:1-20.

12. Trnka Andrej. Market Basket Analysis with Data Mining Methods International Conference on Networking and Information Technology 2010:446-450.

13. Hilage Tejaswini A, Kulkarni RV. Application of data mining techniques to a selected business organization with special reference to buying Behaviour International Journal of Database Management Systems 2011:3(4):169-181.