Guided Project Report

AssociationRuleMiningMarketBasketAnalysis

Name: Shruti Verma
Course: AI and ML
(Batch 4)
Duration: 10 months

Problem Statement: Build a relation between the products bought by the customers using store data

Prerequisites

What things you need to install the software and how to install them:

Python 3.8 or higher versions This setup requires that your machine has latest version of python. The following url https://www.python.org/downloads/ can be referred to download python. Once you have python downloaded and installed, you will need to setup PATH variables (if you want to run python program directly, detail instructions are below in how to run software section). To do that check this: https://www.pythoncentral.io/add-python-to-path-python-is-not- recognized-as-an-internal-or-external- command/. Setting up PATH variable is optional as you can also run program without it and more instruction are given below on this topic.

Second and easier option is to download anaconda and use its anaconda prompt to run the commands. To install anaconda check this url https://www.anaconda.com/download/ You will also need to download and install below 3 packages after you install either python or anaconda from the steps above Sklearn (scikit-learn) numpy scipy if you have chosen to install python 3.8 then run below commands in command prompt/terminal to install these packages pip install -U scikit-learn pip install numpy pip install scipy if you have chosen to install anaconda then run below commands in anaconda prompt to install these packages conda install -c scikit-learn conda install -c anaconda numpy conda install -c anaconda scipy . Install the mlxtend using the command ! pip install mlxtend

Dataset used

The data source is store data having close to 7500 entries for 20 food items using the below link

https://drive.google.com/file/d/1y5DYn0dGoSbC22xowBq2d4po6h1JxcTQ/view?usp=sharing

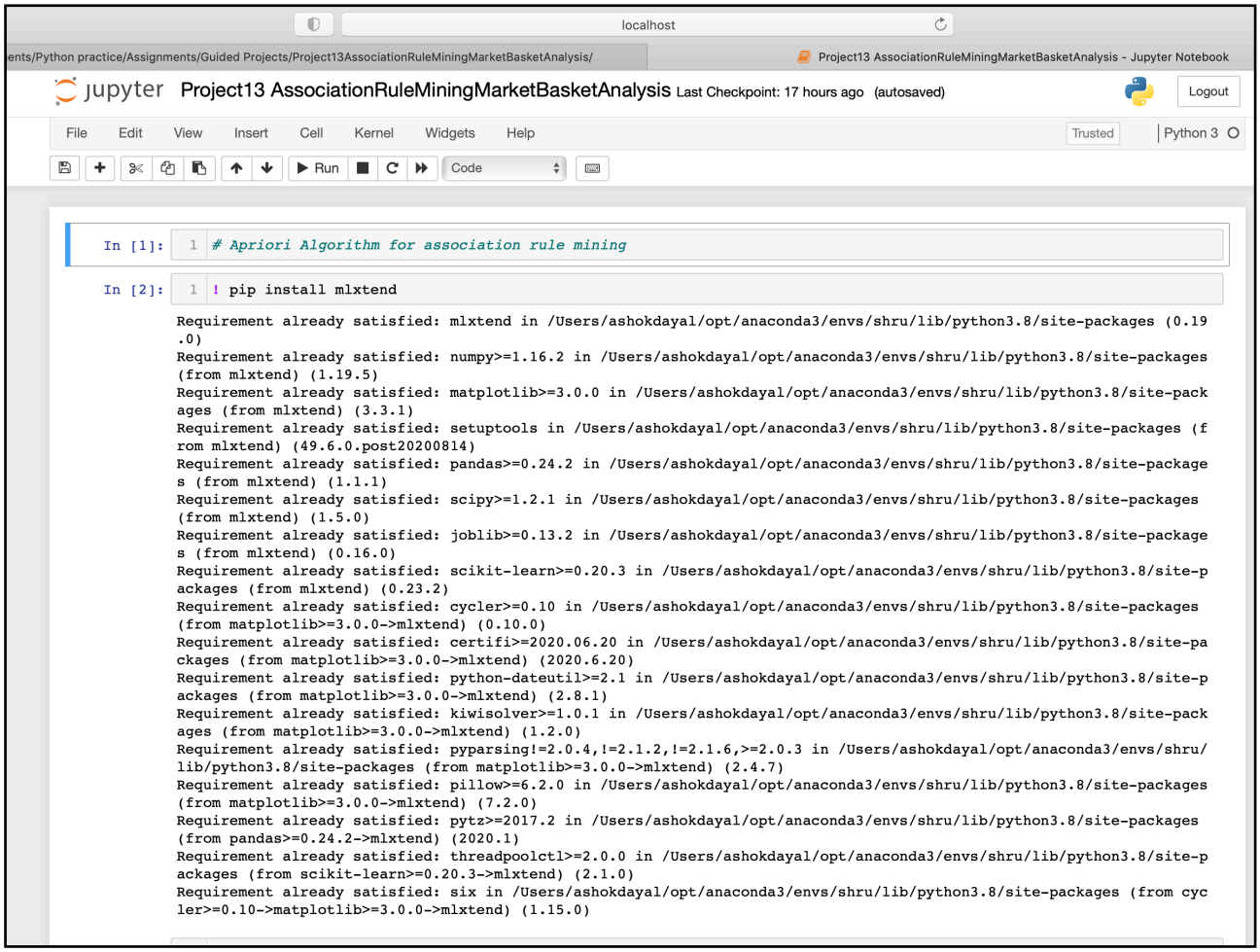
Method used for detection

Apriori Algorithm for association rule mining

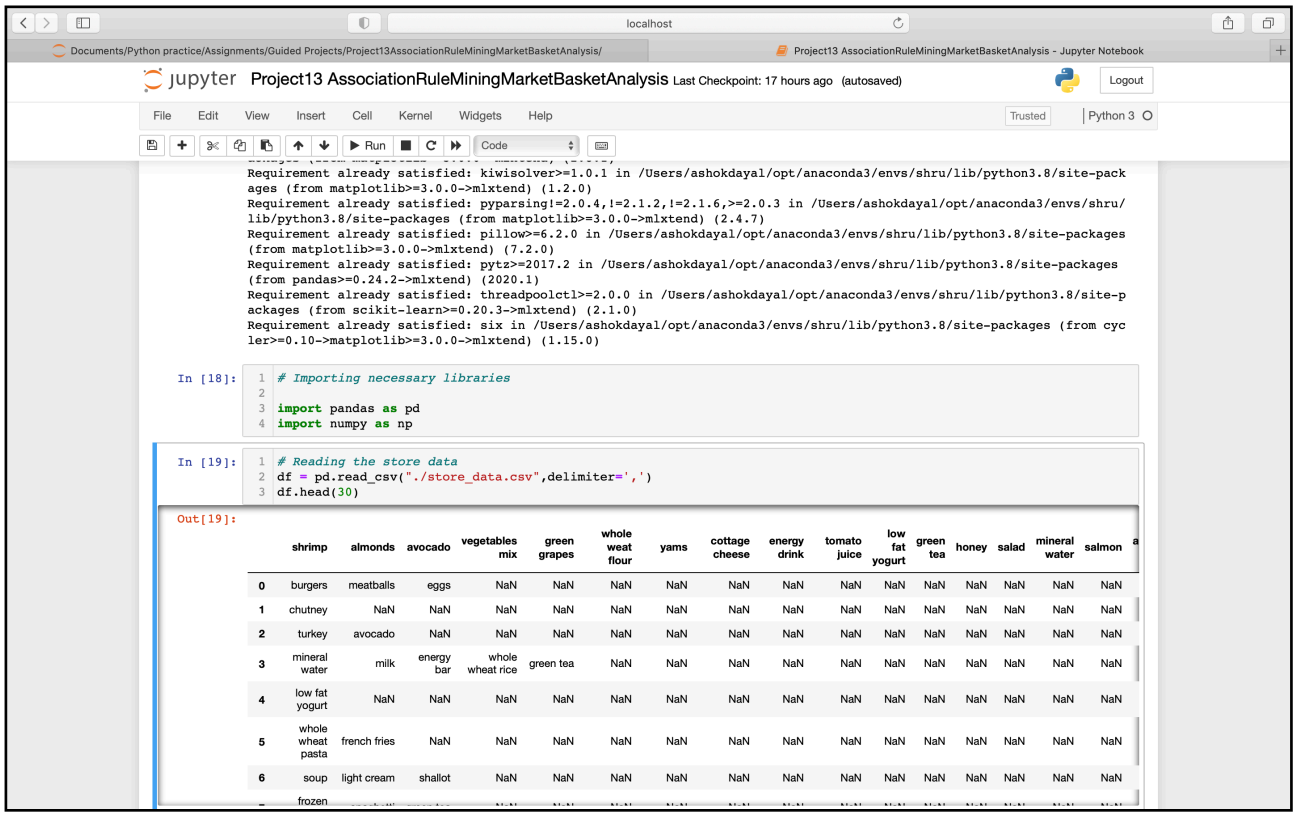
- Set a minimum value for support and confidence. This means that we are only interested in finding rules for the items that have certain default existence (e.g. support) and have a minimum value for co-occurrence with other items (e.g. confidence).
- Extract all the subsets having higher value of support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

Importing the libraries and capturing images:

Installing mlxtend

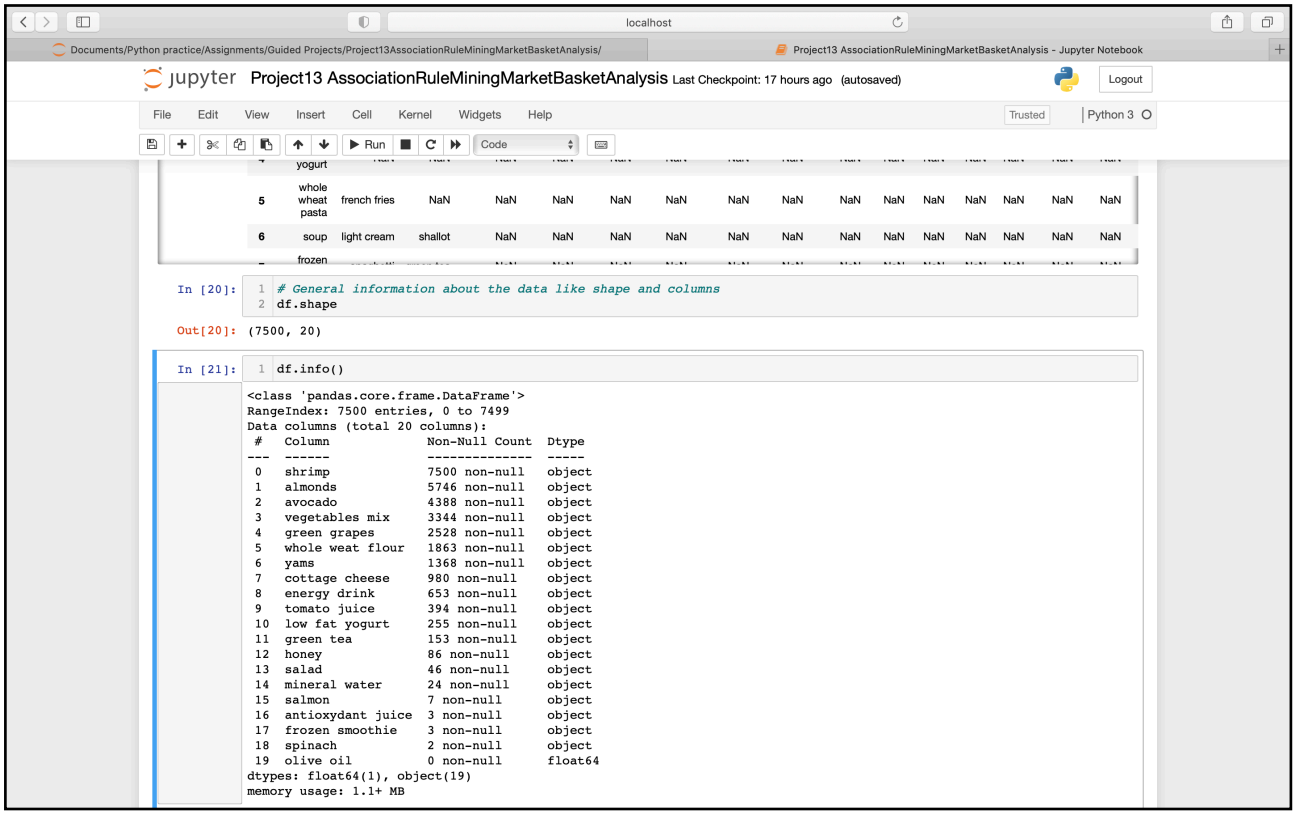


Importing necessary libraries and reading the store data

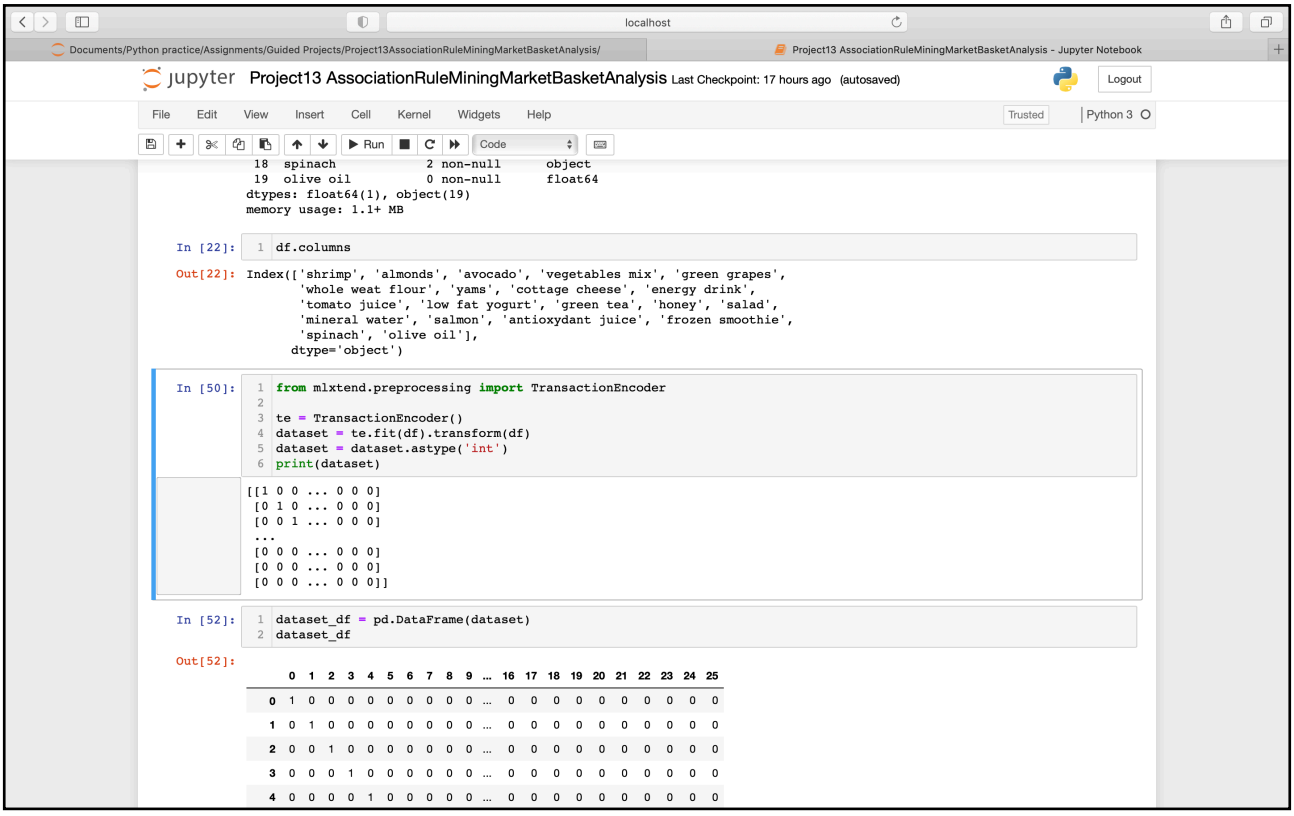


	shrimp	almonds	avocado	vegetables mix	green grapes	whole weat flour	yams	cottage cheese	energy drink	tomato juice	low fat yogurt	green tea	honey	salad	mineral water	salmon
0	burgers	meatballs	eggs	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	chutney	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	turkey	avocado	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	mineral water	milk	energy bar	whole wheat rice	green tea	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	low fat yogurt	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5	whole wheat pasta	french fries	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
6	soup	light cream	shallot	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

General Information about the data like 7500 records with 20 columns

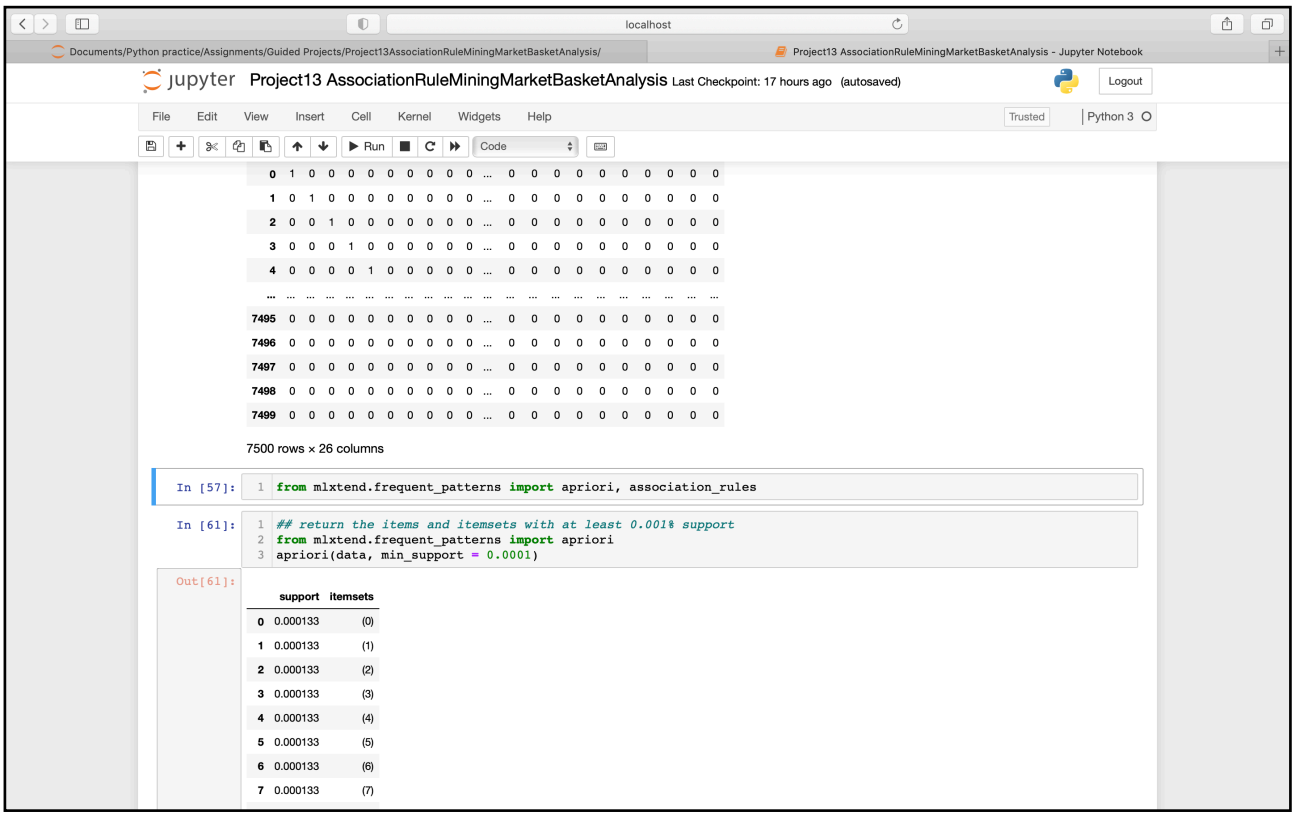


Applying transaction encoder to get the binary code for each record of transaction and converting it into dataframe

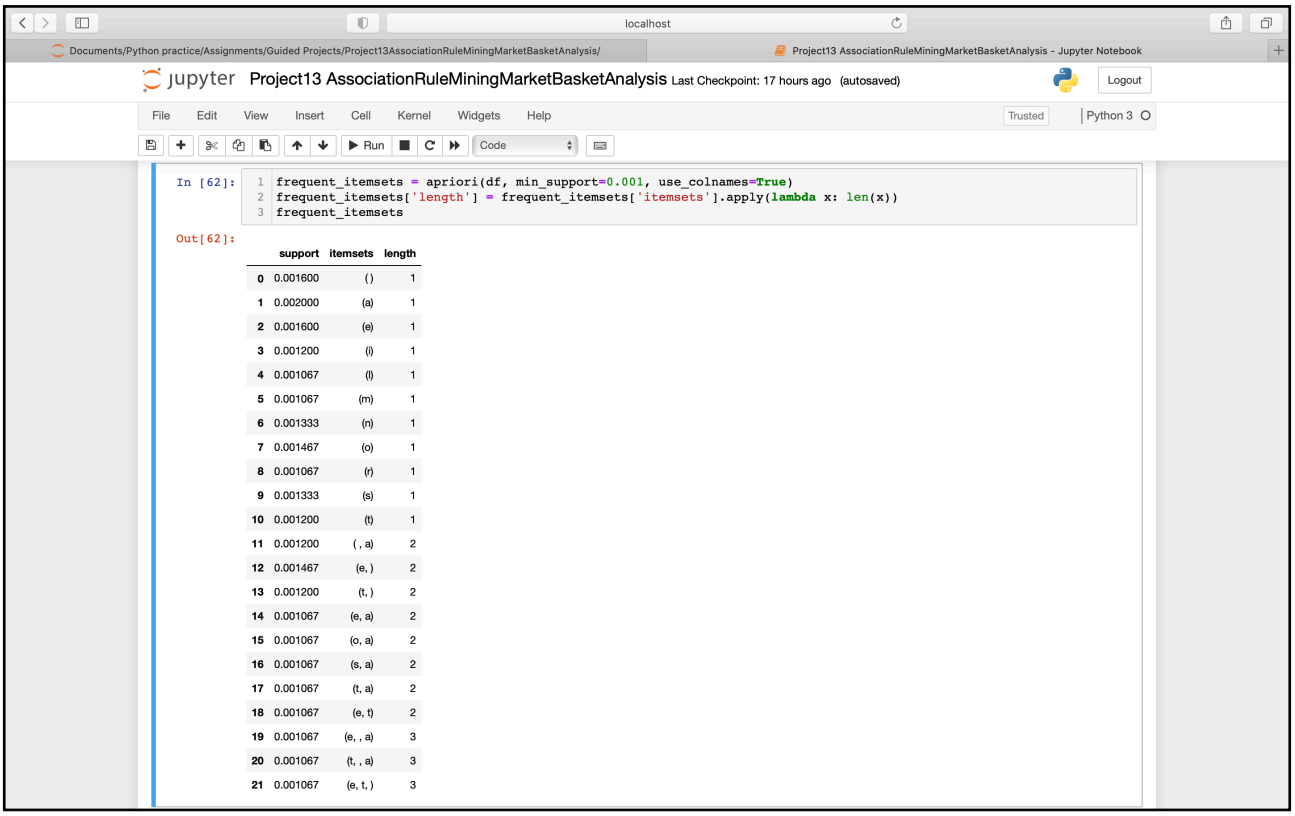


	0	1	2	3	4	5	6	7	8	9	...	16	17	18	19	20	21	22	23	24	25
0	1	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
1	0	1	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
2	0	0	1	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
3	0	0	0	1	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	1	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0

Applying apriori association rules and return the items and item sets with at least 0.001% support

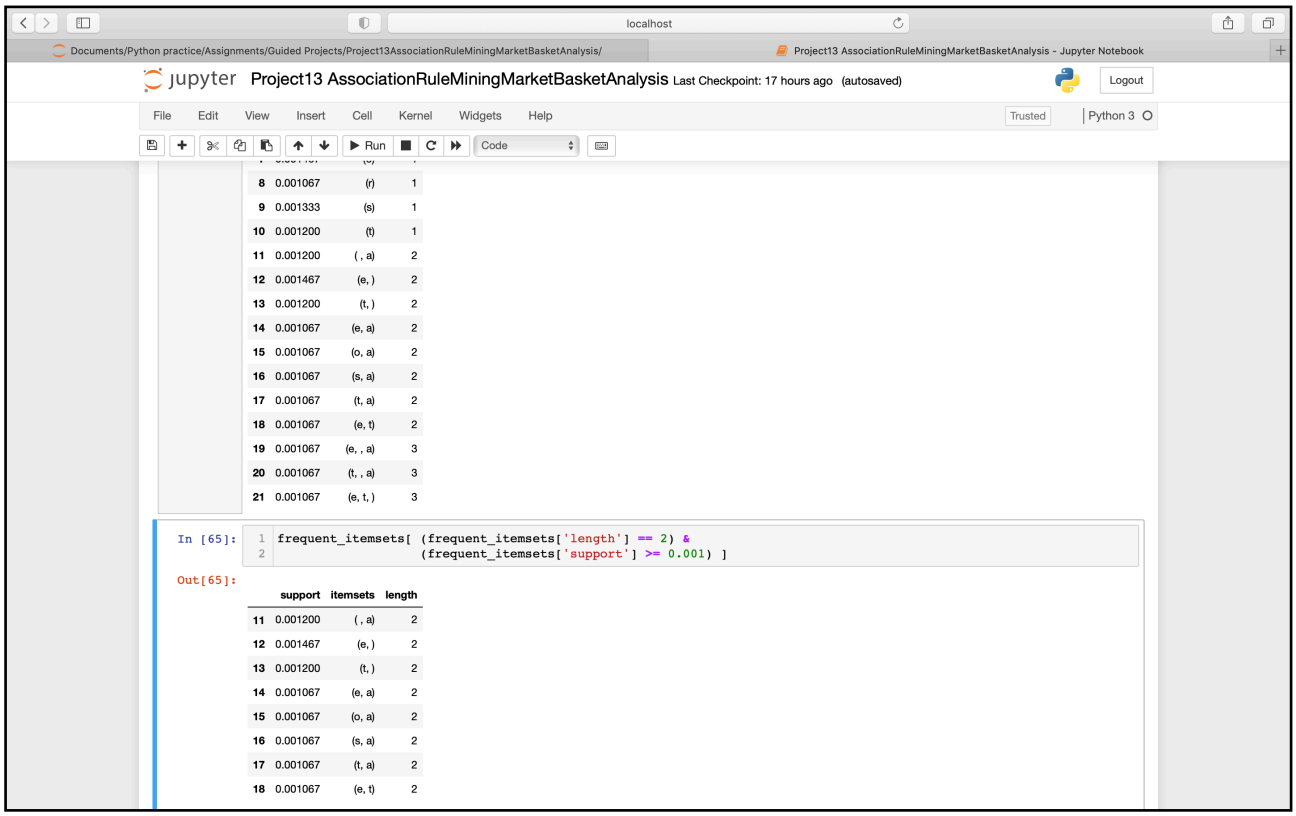


	support	itemsets
0	0.000133	(0)
1	0.000133	(1)
2	0.000133	(2)
3	0.000133	(3)
4	0.000133	(4)
5	0.000133	(5)
6	0.000133	(6)
7	0.000133	(7)



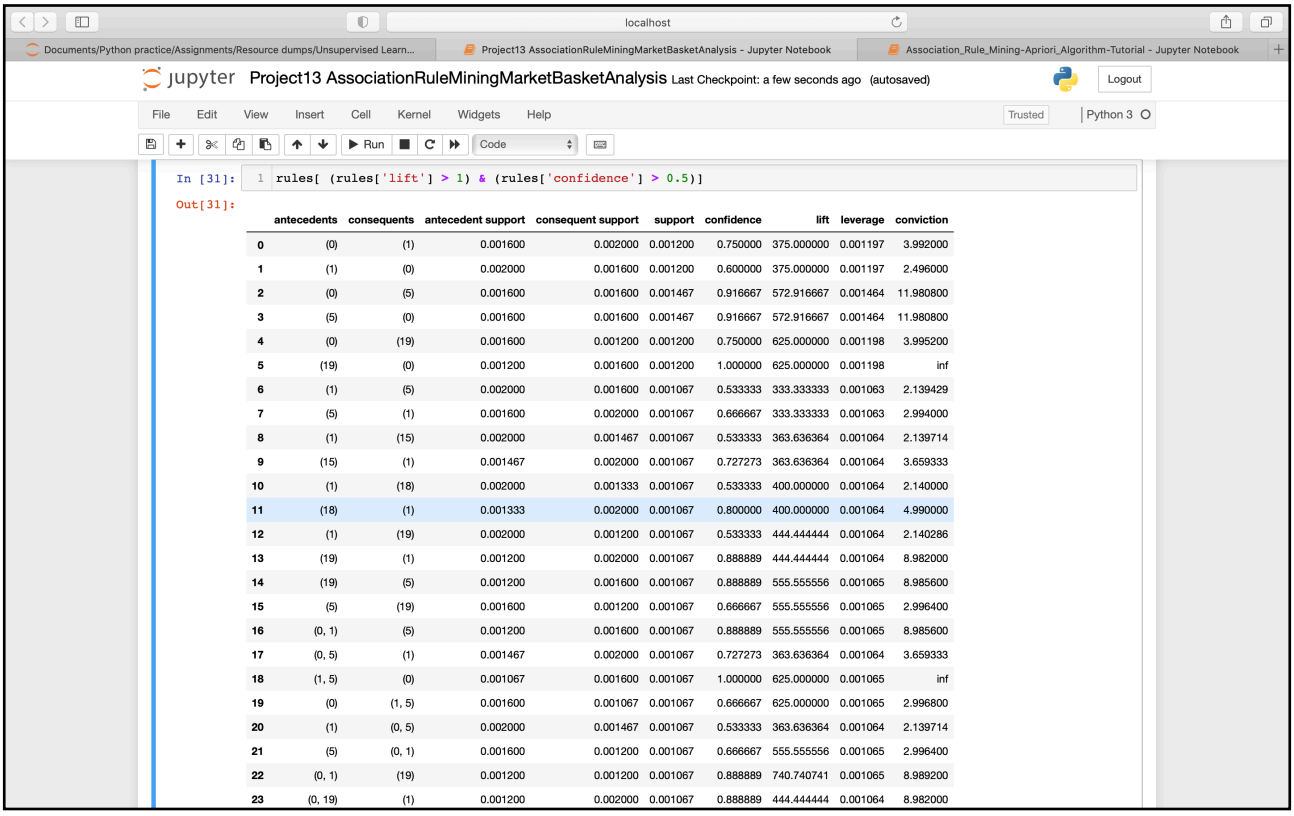
	support	itemsets	length
0	0.001600	( )	1
1	0.002000	(a)	1
2	0.001600	(e)	1
3	0.001200	(i)	1
4	0.001067	(l)	1
5	0.001067	(m)	1
6	0.001333	(n)	1
7	0.001467	(o)	1
8	0.001067	(r)	1
9	0.001333	(s)	1
10	0.001200	(t)	1
11	0.001200	(, a)	2
12	0.001467	(e,)	2
13	0.001200	(t,)	2
14	0.001067	(e, a)	2
15	0.001067	(o, a)	2
16	0.001067	(s, a)	2
17	0.001067	(t, a)	2
18	0.001067	(e, t)	2
19	0.001067	(e, , a)	3
20	0.001067	(t, , a)	3
21	0.001067	(e, t,)	3

Getting the result for support greater than 0.001 and length as 2.



	support	itemsets	length
11	0.001200	(, a)	2
12	0.001467	(e,)	2
13	0.001200	(t,)	2
14	0.001067	(e, a)	2
15	0.001067	(o, a)	2
16	0.001067	(s, a)	2
17	0.001067	(t, a)	2
18	0.001067	(e, t)	2

Applying the rule of life greater than one and confidence greater than 0.4



	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(0)	(1)	0.001600	0.002000	0.001200	0.750000	375.000000	0.001197	3.992000
1	(1)	(0)	0.002000	0.001600	0.001200	0.600000	375.000000	0.001197	2.496000
2	(0)	(5)	0.001600	0.001600	0.001467	0.916667	572.916667	0.001464	11.980800
3	(5)	(0)	0.001600	0.001600	0.001467	0.916667	572.916667	0.001464	11.980800
4	(0)	(19)	0.001600	0.001200	0.001200	0.750000	625.000000	0.001198	3.995200
5	(19)	(0)	0.001200	0.001600	0.001200	1.000000	625.000000	0.001198	inf
6	(1)	(5)	0.002000	0.001600	0.001067	0.533333	333.333333	0.001063	2.139429
7	(5)	(1)	0.001600	0.002000	0.001067	0.666667	333.333333	0.001063	2.994000
8	(1)	(15)	0.002000	0.001467	0.001067	0.533333	363.636364	0.001064	2.139714
9	(15)	(1)	0.001467	0.002000	0.001067	0.727273	363.636364	0.001064	3.659333
10	(1)	(18)	0.002000	0.001333	0.001067	0.533333	400.000000	0.001064	2.140000
11	(18)	(1)	0.001333	0.002000	0.001067	0.800000	400.000000	0.001064	4.990000
12	(1)	(19)	0.002000	0.001200	0.001067	0.533333	444.444444	0.001064	2.140286
13	(19)	(1)	0.001200	0.002000	0.001067	0.888889	444.444444	0.001064	8.982000
14	(19)	(5)	0.001200	0.001600	0.001067	0.888889	555.555556	0.001065	8.985600
15	(5)	(19)	0.001600	0.001200	0.001067	0.666667	555.555556	0.001065	2.996400
16	(0, 1)	(5)	0.001200	0.001600	0.001067	0.888889	555.555556	0.001065	8.985600
17	(0, 5)	(1)	0.001467	0.002000	0.001067	0.727273	363.636364	0.001064	3.659333
18	(1, 5)	(0)	0.001067	0.001600	0.001067	1.000000	625.000000	0.001065	inf
19	(0)	(1, 5)	0.001600	0.001067	0.001067	0.666667	625.000000	0.001065	2.996800
20	(1)	(0, 5)	0.002000	0.001467	0.001067	0.533333	363.636364	0.001064	2.139714
21	(5)	(0, 1)	0.001600	0.001200	0.001067	0.666667	555.555556	0.001065	2.996400
22	(0, 1)	(19)	0.001200	0.001200	0.001067	0.888889	740.740741	0.001065	8.989200
23	(0, 19)	(1)	0.001200	0.002000	0.001067	0.888889	444.444444	0.001064	8.982000