Big Data & Data Analytics-2

Sagar Soni
180044

Project Activity - 1

Section – B

Ans 1:

Apriori algorithm is a classical algorithm in data mining. It is used for mining frequent item sets and relevant association rules. It is devised to operate on a database containing a lot of transactions, for instance, items brought by customers in a store.

It helps the customers buy their items with ease, and enhances the sales performance of the departmental store.

This algorithm has utility in the field of healthcare as it can help in detecting adverse drug reactions (ADR) by producing association rules to indicate the combination of medications and patient characteristics that could lead to ADRs.

Ans 2:

Support represents the popularity of that product of all the product transactions. Support of the product is calculated as the ratio of the number of transactions includes that product and the total number of transactions.

Support of the product = (Number of transactions includes that product)/ (Total number of transactions)

When we decrease the support level, we decrease the amount of popularity we need.

Ans 3:

This explains how likely Y is purchased when X is purchased. This defines association between two items.

For example, when a person buys milk they are more likely to buy bread as well or vice versa. This is measured by the proportion of transactions with item X, in which item Y also appears. Expressed as {X -> Y}. Calculated by the proportion of the number of transactions in which both (X & Y) occur to support item X. When we increase the confidence level, we tend to increase the probability of that item to be transacted with the compared item.

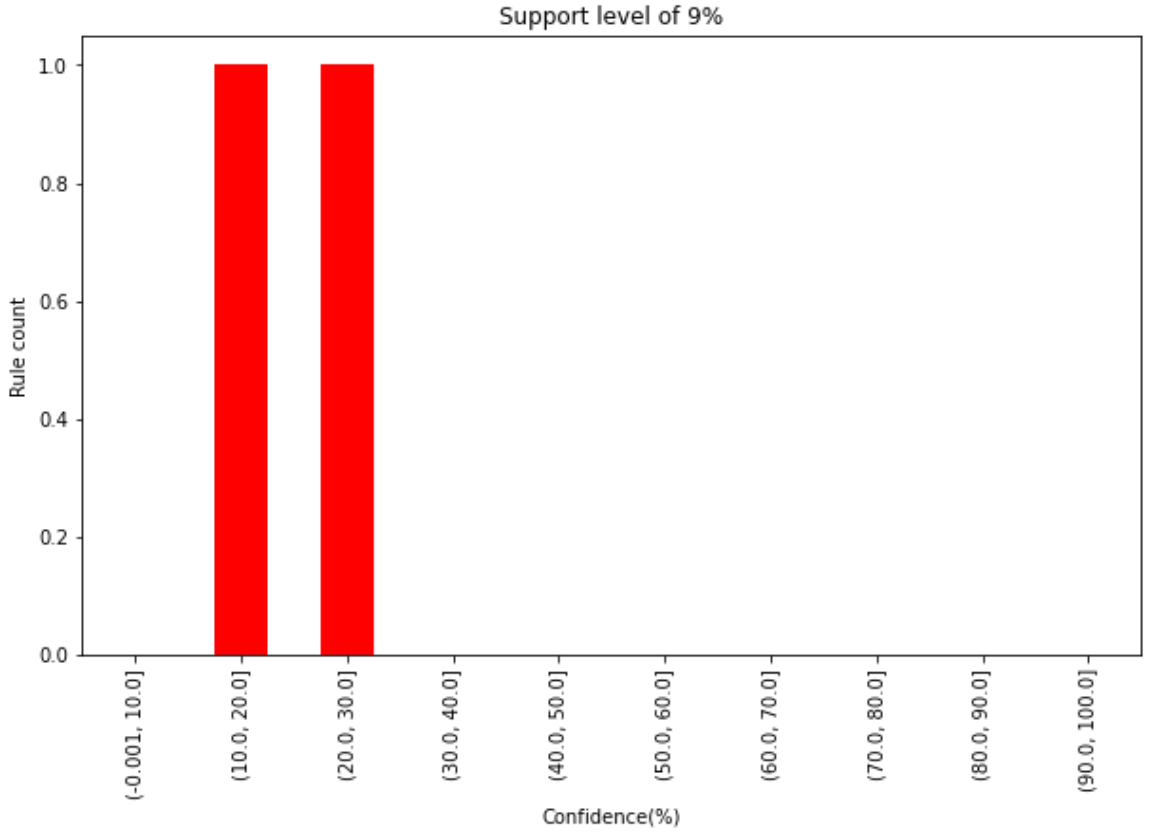
Big Data & Data Analytics-2

Ans 4:

There were no rules for support (10%) so showing min support of 9 %:

Confidence	count
(-0.001, 10.0]	0
(10.0, 20.0]	1
(20.0, 30.0]	1
(30.0, 40.0]	0
(40.0, 50.0]	0
(50.0, 60.0]	0
(60.0, 70.0]	0
(70.0, 80.0]	0
(80.0, 90.0]	0
(90.0, 100.0]	0

Name: confidence, dtype: int64



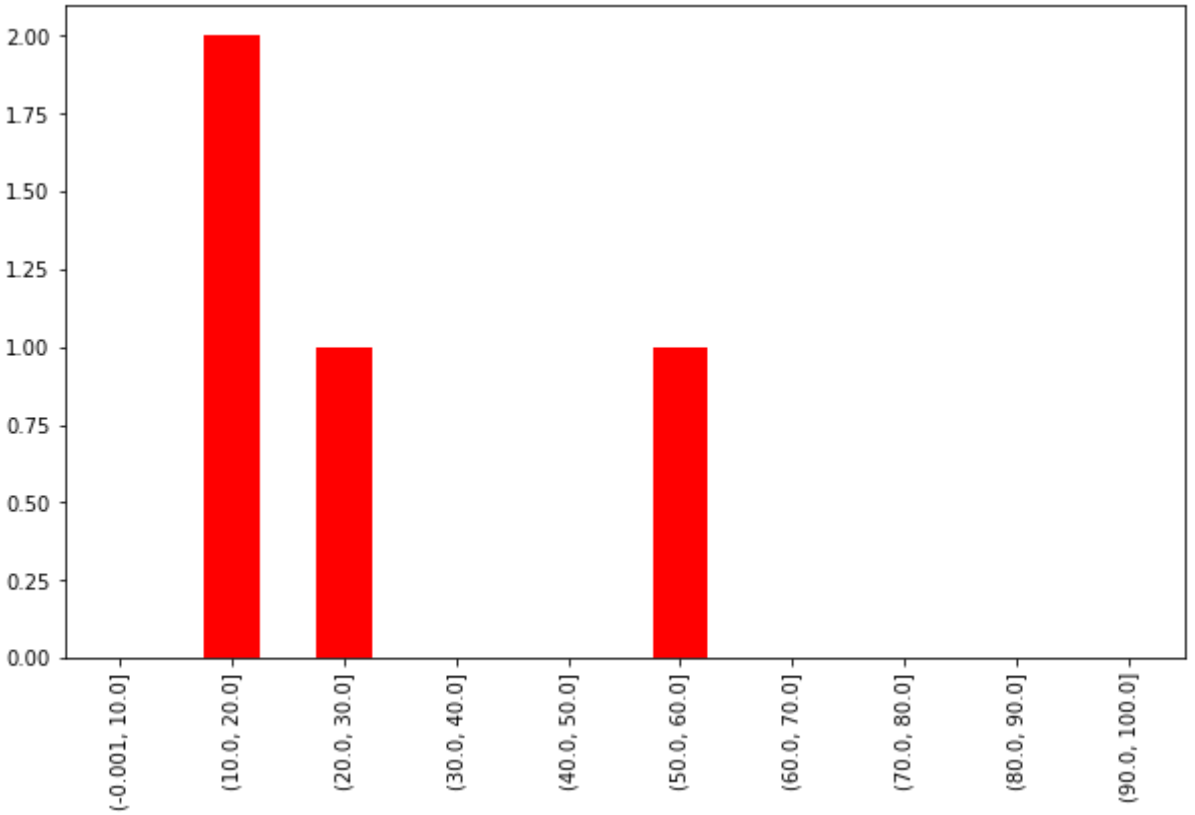
Support Level of 5%:

Confidence	count
(-0.001, 10.0]	0

Big Data & Data Analytics-2

Ans 5:

```
(10.0, 20.0]      2
(20.0, 30.0]      1
(30.0, 40.0]      0
(40.0, 50.0]      0
(50.0, 60.0]      1
(60.0, 70.0]      0
(70.0, 80.0]      0
(80.0, 90.0]      0
(90.0, 100.0]     0
Name: confidence, dtype: int64
```



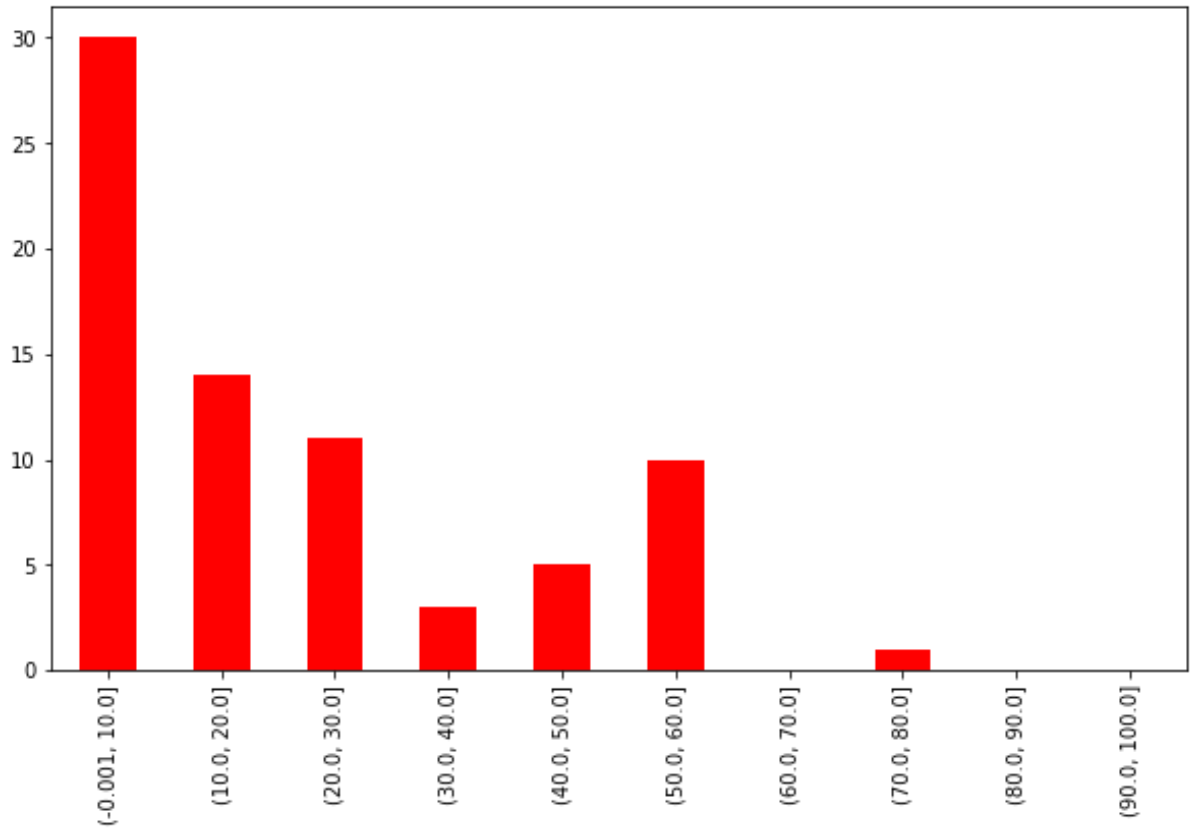
Big Data & Data Analytics-2

Ans 6:

Support Level of 1%:

Confidence	count
(-0.001, 10.0]	30
(10.0, 20.0]	14
(20.0, 30.0]	11
(30.0, 40.0]	3
(40.0, 50.0]	5
(50.0, 60.0]	10
(60.0, 70.0]	0
(70.0, 80.0]	1
(80.0, 90.0]	0
(90.0, 100.0]	0

Name: confidence, dtype: int64



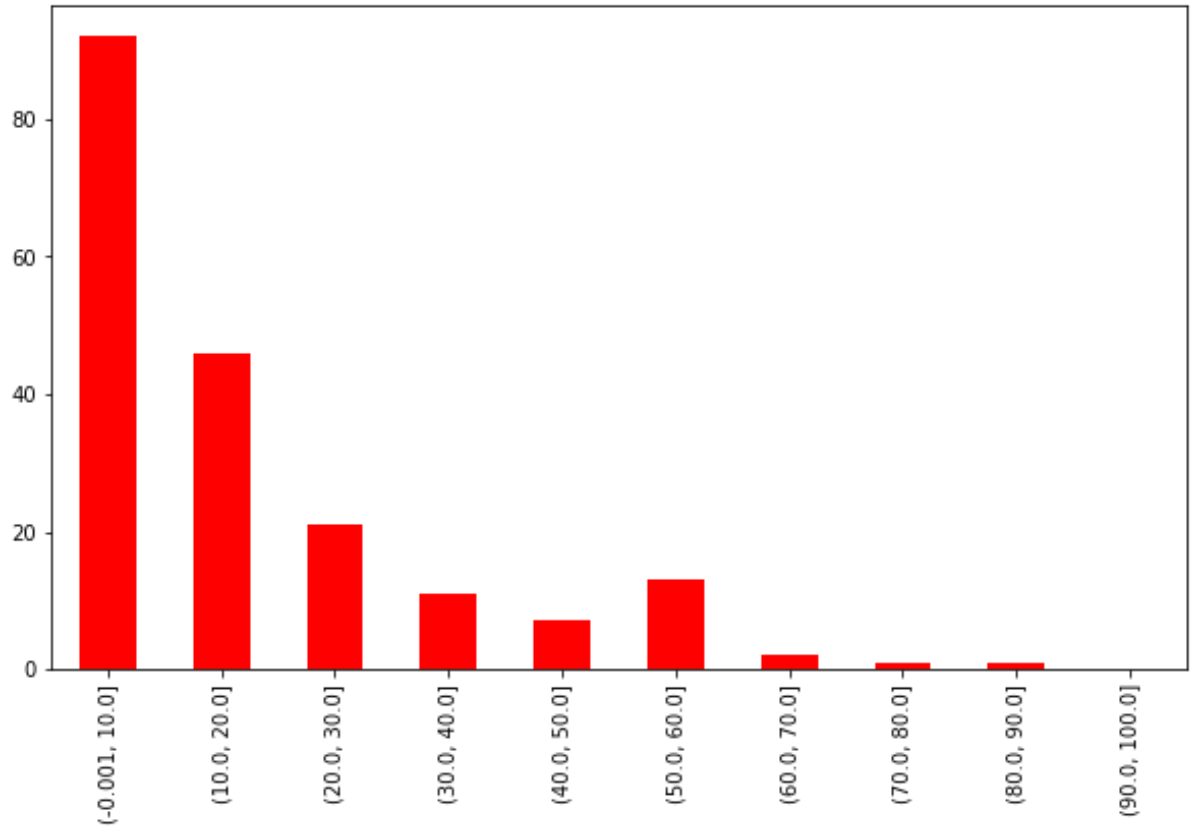
Big Data & Data Analytics-2

Ans 7:

Support Level of 0.5%:

Confidence	count
(-0.001, 10.0]	92
(10.0, 20.0]	46
(20.0, 30.0]	21
(30.0, 40.0]	11
(40.0, 50.0]	7
(50.0, 60.0]	13
(60.0, 70.0]	2
(70.0, 80.0]	1
(80.0, 90.0]	1
(90.0, 100.0]	0

Name: confidence, dtype: int64



Big Data & Data Analytics-2

Ans 9:

```
Rules:     antecedents  consequents  antecedent support  consequent
support   support  \
1     (Bread)   (Alfajores)            0.327205            0.036344
      0.010354
2     (Alfajores)      (Bread)            0.036344
      0.327205  0.010354
3     (Alfajores)     (Coffee)            0.036344
      0.478394  0.019651
4     (Coffee)   (Alfajores)            0.478394            0.036344
      0.019651
5     (Brownie)      (Bread)            0.040042            0.327205
0.010777        confidence      lift  leverage
conviction    0    0.031644  0.870657 -0.001538
0.995145
1     0.284884  0.870657 -0.001538    0.940818
2     0.540698  1.130235  0.002264    1.135648
3     0.041078  1.130235  0.002264    1.004936
4     0.269129  0.822508 -0.002326    0.920538
  pairs with best
confidence:
          antecedents consequents   support  confidence      lift
40           (Toast)    (Coffee)  2.366614   70.440252  1.472431
38  (Spanish Brunch)    (Coffee)  1.088220   59.883721  1.251766
26       (Medialuna)    (Coffee)  3.518225   56.923077  1.189878
30          (Pastry)    (Coffee)  4.754358   55.214724  1.154168 2
      (Alfajores)    (Coffee)  1.965135   54.069767  1.130235
  pairs with best
support:

   antecedents consequents   support  confidence      lift 15
(Coffee)      (Cake)  5.472795   11.439929  1.101515
14      (Cake)    (Coffee)  5.472795   52.695829  1.101515
30    (Pastry)    (Coffee)  4.754358   55.214724  1.154168
31    (Coffee)    (Pastry)  4.754358    9.938163  1.154168
32    (Coffee)  (Sandwich)  3.824617    7.994700  1.112792
```

Big Data & Data Analytics-2

As We can see from the above tables, The best possible pairs would be Cake & Coffee, Paestry & Cake

pairs with worst confidence:

	antecedents	consequents	support	confidence	lift
0	(Bread)	(Alfajores)	1.035394	3.164353	0.870657
51	(Coffee)	(Bread, Pastry)	1.119915	2.340989	0.802807
39	(Coffee)	(Spanish Brunch)	1.088220	2.274735	1.251766
47	(Coffee)	(Bread, Cake)	1.003698	2.098057	0.898557
57	(Coffee)	(Tea, Cake)	1.003698	2.098057	0.882582

pairs with best lift:

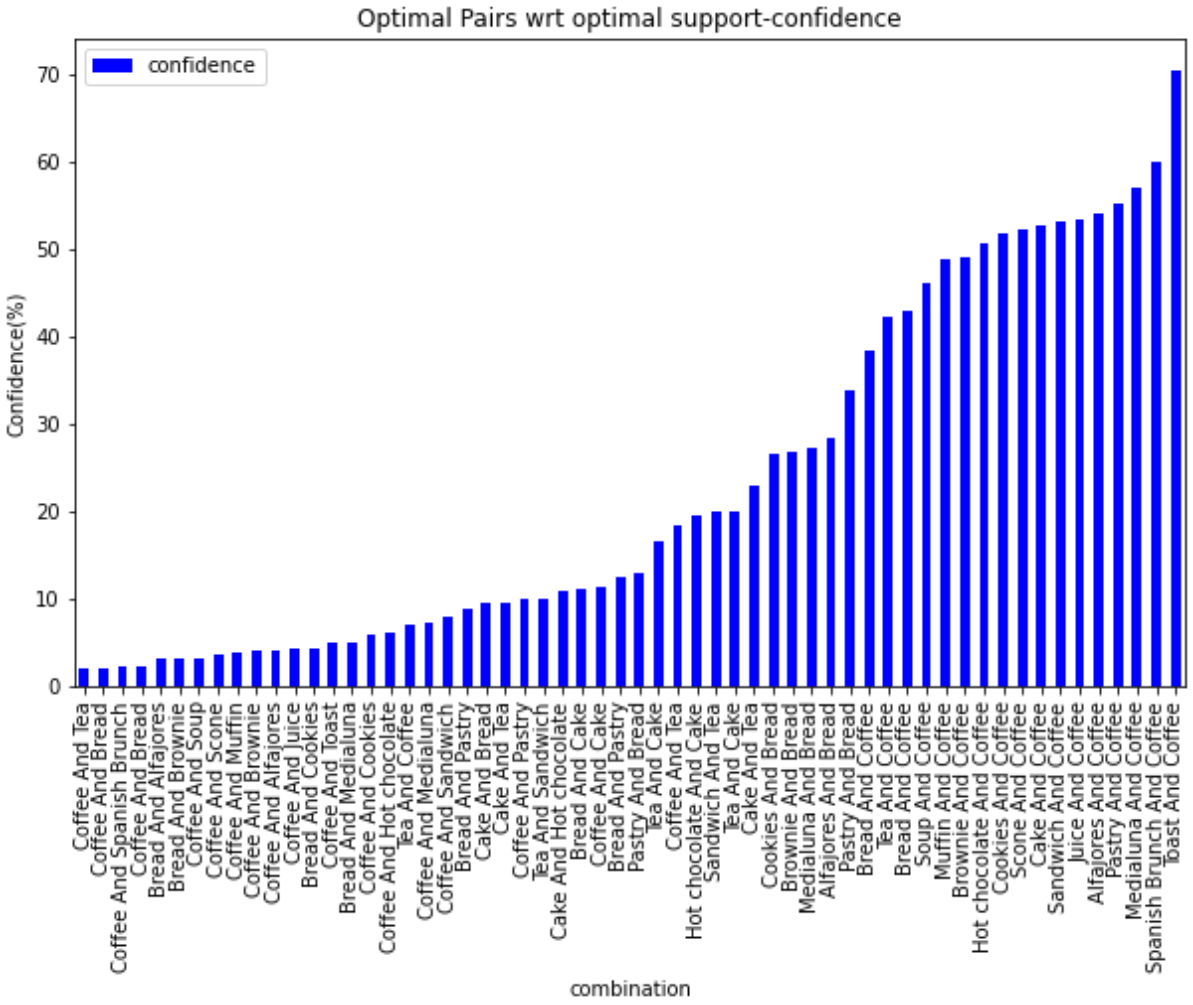
	antecedents	consequents	support	confidence	lift
56	(Cake)	(Tea, Coffee)	1.003698	9.664293	1.937977
53	(Tea, Coffee)	(Cake)	1.003698	20.127119	1.937977
17	(Cake)	(Hot chocolate)	1.141046	10.986775	1.883874
16	(Hot chocolate)	(Cake)	1.141046	19.565217	1.883874
18	(Tea)	(Cake)	2.377179	16.666667	1.604781

pairs with worst lift:

	antecedents	consequents	support	confidence	lift
4	(Brownie)	(Bread)	1.077655	26.912929	0.822508
6	(Cookies)	(Bread)	1.447438	26.601942	0.813004
7	(Bread)	(Cookies)	1.447438	4.423636	0.813004
48	(Bread, Pastry)	(Coffee)	1.119915	38.405797	0.802807
51	(Coffee)	(Bread, Pastry)	1.119915	2.340989	0.802807

All the pairs with minimum support their confidence:

Big Data & Data Analytics-2



Ans 10:

- There is a 70% chance that he/she will buy coffee.
- Never recommend your customers to buy brownies with bread.