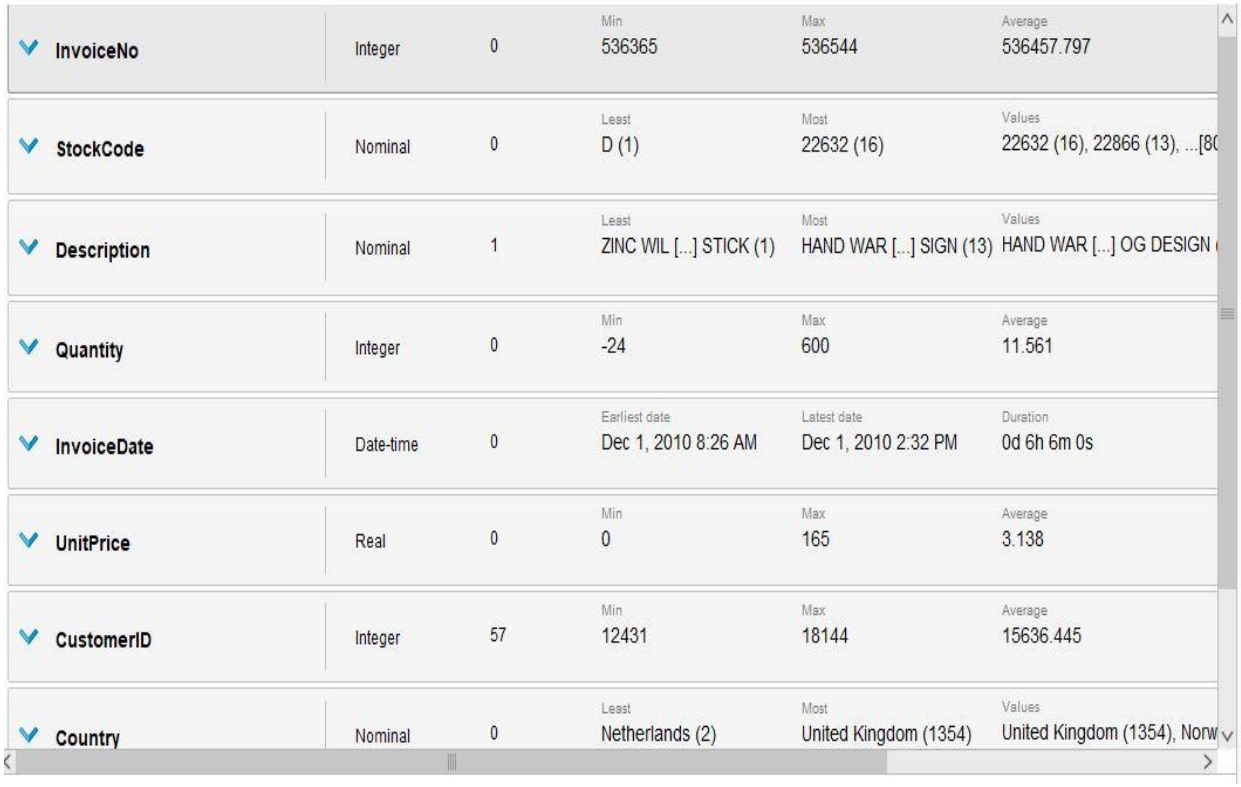
Assignment 2 - Association Rules:

The following are the steps by which the analysis is performed on the RapidMiner:-



InvoiceNo	Integer	0	Min 536365	Max 536544	Average 536457.797
StockCode	Nominal	0	Least D (1)	Most 22632 (16)	Values 22632 (16), 22866 (13), ...[8
Description	Nominal	1	Least ZINC WIL [...] STICK (1)	Most HAND WAR [...] SIGN (13)	Values HAND WAR [...] OG DESIGN
Quantity	Integer	0	Min -24	Max 600	Average 11.561
InvoiceDate	Date-time	0	Earliest date Dec 1, 2010 8:26 AM	Latest date Dec 1, 2010 2:32 PM	Duration 0d 6h 6m 0s
UnitPrice	Real	0	Min 0	Max 165	Average 3.138
CustomerID	Integer	57	Min 12431	Max 18144	Average 15636.445
Country	Nominal	0	Least Netherlands (2)	Most United Kingdom (1354)	Values United Kingdom (1354), Norw

(Figure 1.1)

The above figure is the dataset's fields.

Row No.	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	I
1	536365	85123A	WHITE HANG...	6	Dec 1, 2010 ...	2.550	17850	United Kingd...	?
2	536365	71053	WHITE META...	6	Dec 1, 2010 ...	3.390	17850	United Kingd...	?
3	536365	84406B	CREAM CUPI...	8	Dec 1, 2010 ...	2.750	17850	United Kingd...	?
4	536365	84029G	KNITTED UNI...	6	Dec 1, 2010 ...	3.390	17850	United Kingd...	?
5	536365	84029E	RED WOOLL...	6	Dec 1, 2010 ...	3.390	17850	United Kingd...	?
6	536365	22752	SET 7 BABU...	2	Dec 1, 2010 ...	7.650	17850	United Kingd...	?
7	536365	21730	GLASS STAR...	6	Dec 1, 2010 ...	4.250	17850	United Kingd...	?
8	536366	22633	HAND WARM...	6	Dec 1, 2010 ...	1.850	17850	United Kingd...	?
9	536366	22632	HAND WARM...	6	Dec 1, 2010 ...	1.850	17850	United Kingd...	?
10	536367	84879	ASSORTED ...	32	Dec 1, 2010 ...	1.690	13047	United Kingd...	?
11	536367	22745	POPPY'S PLA...	6	Dec 1, 2010 ...	2.100	13047	United Kingd...	?
12	536367	22748	POPPY'S PLA...	6	Dec 1, 2010 ...	2.100	13047	United Kingd...	?
13	536367	22749	FELTCRAFT ...	8	Dec 1, 2010 ...	3.750	13047	United Kingd...	?
14	536367	22310	IVORY KNITT...	6	Dec 1, 2010 ...	1.650	13047	United Kingd...	?
15	536367	84969	BOX OF 6 AS...	6	Dec 1, 2010 ...	4.250	13047	United Kingd...	?

ExampleSet (1,499 examples, 0 special attributes, 9 regular attributes)

(Figure 1.2)

The above figure shows description of its fields.

No. of Sets: 15

Total Max. Size: 4

Min. Size: 1

Max. Size: 4

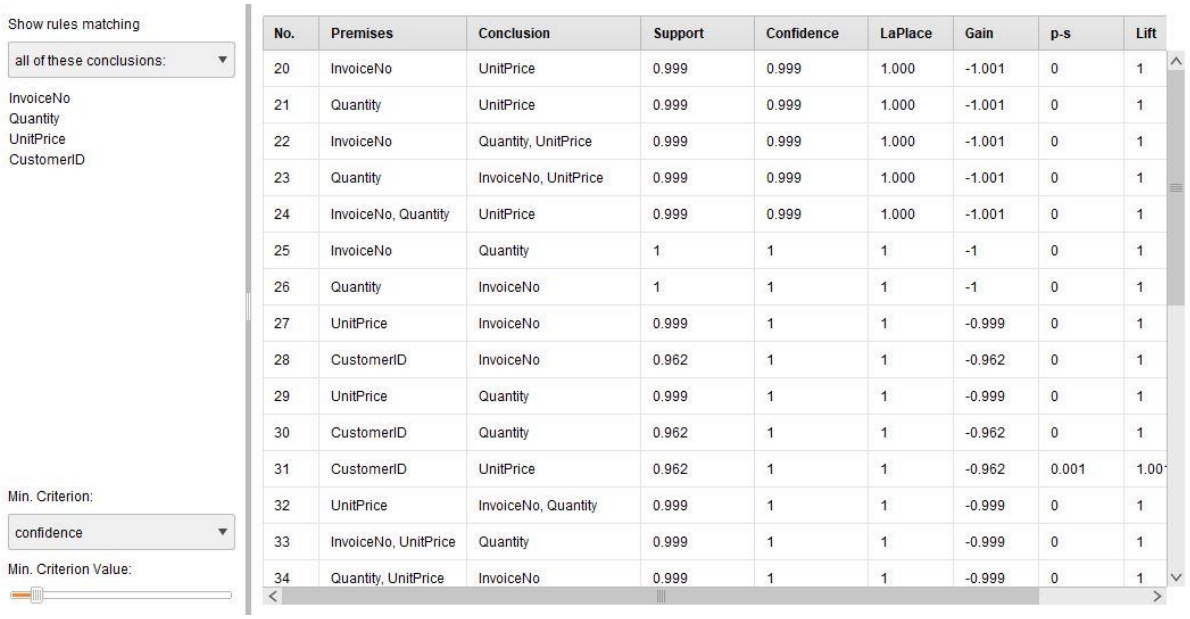
Contains Item:

Update View

Size	Support	Item 1	Item 2	Item 3	Item 4
1	1.000	InvoiceNo			
1	1.000	Quantity			
1	0.999	UnitPrice			
1	0.962	CustomerID			
2	1.000	InvoiceNo	Quantity		
2	0.999	InvoiceNo	UnitPrice		
2	0.962	InvoiceNo	CustomerID		
2	0.999	Quantity	UnitPrice		
2	0.962	Quantity	CustomerID		
2	0.962	UnitPrice	CustomerID		
3	0.999	InvoiceNo	Quantity	UnitPrice	
3	0.962	InvoiceNo	Quantity	CustomerID	
3	0.962	InvoiceNo	UnitPrice	CustomerID	
3	0.962	Quantity	UnitPrice	CustomerID	
4	0.962	InvoiceNo	Quantity	UnitPrice	CustomerID

(Figure 1.3)

The above figure shows the results after the use of FP-Growth operator.
A support of 0.2 is used as a parameter for the better results.



No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain	p-s	Lift
20	InvoiceNo	UnitPrice	0.999	0.999	1.000	-1.001	0	1
21	Quantity	UnitPrice	0.999	0.999	1.000	-1.001	0	1
22	InvoiceNo	Quantity, UnitPrice	0.999	0.999	1.000	-1.001	0	1
23	Quantity	InvoiceNo, UnitPrice	0.999	0.999	1.000	-1.001	0	1
24	InvoiceNo, Quantity	UnitPrice	0.999	0.999	1.000	-1.001	0	1
25	InvoiceNo	Quantity	1	1	1	-1	0	1
26	Quantity	InvoiceNo	1	1	1	-1	0	1
27	UnitPrice	InvoiceNo	0.999	1	1	-0.999	0	1
28	CustomerID	InvoiceNo	0.962	1	1	-0.962	0	1
29	UnitPrice	Quantity	0.999	1	1	-0.999	0	1
30	CustomerID	Quantity	0.962	1	1	-0.962	0	1
31	CustomerID	UnitPrice	0.962	1	1	-0.962	0.001	1.00
32	UnitPrice	InvoiceNo, Quantity	0.999	1	1	-0.999	0	1
33	InvoiceNo, UnitPrice	Quantity	0.999	1	1	-0.999	0	1
34	Quantity, UnitPrice	InvoiceNo	0.999	1	1	-0.999	0	1

(Figure 1.4 a)

No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain	p-s	Lift
36	InvoiceNo, Custom...	Quantity	0.962	1	1	-0.962	0	1
37	Quantity, Customer...	InvoiceNo	0.962	1	1	-0.962	0	1
38	CustomerID	InvoiceNo, UnitPrice	0.962	1	1	-0.962	0.001	1.00
39	InvoiceNo, Custom...	UnitPrice	0.962	1	1	-0.962	0.001	1.00
40	UnitPrice, Custom...	InvoiceNo	0.962	1	1	-0.962	0	1
41	CustomerID	Quantity, UnitPrice	0.962	1	1	-0.962	0.001	1.00
42	Quantity, Customer...	UnitPrice	0.962	1	1	-0.962	0.001	1.00
43	UnitPrice, Custom...	Quantity	0.962	1	1	-0.962	0	1
44	CustomerID	InvoiceNo, Quantity, Un...	0.962	1	1	-0.962	0.001	1.00
45	InvoiceNo, Custom...	Quantity, UnitPrice	0.962	1	1	-0.962	0.001	1.00
46	Quantity, Customer...	InvoiceNo, UnitPrice	0.962	1	1	-0.962	0.001	1.00
47	InvoiceNo, Quantity...	UnitPrice	0.962	1	1	-0.962	0.001	1.00
48	UnitPrice, Custom...	InvoiceNo, Quantity	0.962	1	1	-0.962	0	1
49	InvoiceNo, UnitPric...	Quantity	0.962	1	1	-0.962	0	1
50	Quantity, UnitPrice,...	InvoiceNo	0.962	1	1	-0.962	0	1

(Figure 1.4 b)

The above figures show the results of Association Rule's implication on the dataset. Here we can see the results of 'Support' and 'Confidence' thoroughly. The minimum confidence level was 0.5 at this level in the operator's parameters.

REPORT:

What rules did you find?

We found 31 association rules, i.e.; Invoice No / CustomerID, Quantity / CustomerID, CustomerID, UnitPrice / CustomerID etc

What attributes are most strongly associated with one another?

These attributes are the InvoiceNo, Quantity, UnitPrice, InvoiceNo / Quantity, CustomerID.

Are there products that are frequently connected that surprise you? Why do you think this might be?

No, all the products are according to my understanding.

How much did you have to test different support and confidence values before you found some associations rules?

I have tested 3 times just, first testing was with the default values and the next two comes with some changes.

Were any of your association rules good enough that you would base decisions on them? Why or why not?

I will go on with all the rules, because all the rules found indicate that the products are related to shopping baskets and given a value for the confidence and support that is good enough for our research/analysis.

References:

Dataset used

https://www.kaggle.com/