Topic: Association Rules

Prepare rules for the all the data sets

1) Try different values of support and confidence. Observe the change in number of rules for different support, confidence values

2) Change the minimum length in apriori algorithm

3) Visualize the obtained rules using different plots

1.) Books.csv

	ChildBks	YouthBks	CookBks	DoItYBks	RefBks	ArtBks	GeogBks	ItalCook	ItalAtlas	ItalArt	Florence
1	0	1	0	1	0	0	1	0	0	0	0
2	1	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0
4	1	1	1	0	1	0	1	0	0	0	0
5	0	0	1	0	0	0	1	0	0	0	0
6	1	0	0	0	0	1	0	0	0	0	1
7	0	1	0	0	0	0	0	0	0	0	0
8	0	1	0	0	1	0	0	0	0	0	0
9	1	0	0	1	0	0	0	0	0	0	0
10	1	1	1	0	0	0	1	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0
12	0	0	1	0	0	0	1	0	0	0	0
13	1	0	0	0	0	1	0	0	0	0	1

2.) Groceries.csv

	citrus.fruit	semi.finished.bread	margarine	ready.soups
1	tropical fruit	yogurt	coffee	
2	whole milk			
3	pip fruit	yogurt	cream cheese	meat spreads
4	other vegetables	whole milk	condensed milk	long life bakery product
5	whole milk	butter	yogurt	rice
6	abrasive cleaner			
7	rolls/buns			
8	other vegetables	UHT-milk	rolls/buns	bottled beer
9	liquor (appetizer)			
10	pot plants			
11	whole milk	cereals		
12	tropical fruit	other vegetables	white bread	bottled water

3.) my_movies.csv

	V1	V2	V3	V4	V5	Sixth.Sense	Gladiator	LOTR1	Harry.Potter1	Patriot	LOTR2	Harry.Potter2	LOTR	Braveheart	Green.Mile
1	Sixth Sense	LOTR1	Harry Potter1	Green Mile	LOTR2	1	0	1	1	0	1	0	0	0	1
2	Gladiator	Patriot	Braveheart			0	1	0	0	1	0	0	0	1	0
3	LOTR1	LOTR2				0	0	1	0	0	1	0	0	0	0
4	Gladiator	Patriot	Sixth Sense			1	1	0	0	1	0	0	0	0	0
5	Gladiator	Patriot	Sixth Sense			1	1	0	0	1	0	0	0	0	0
6	Gladiator	Patriot	Sixth Sense			1	1	0	0	1	0	0	0	0	0
7	Harry Potter1	Harry Potter2				0	0	0	1	0	0	1	0	0	0
8	Gladiator	Patriot				0	1	0	0	1	0	0	0	0	0
9	Gladiator	Patriot	Sixth Sense			1	1	0	0	1	0	0	0	0	0
10	Sixth Sense	LOTR	Gladiator	Green Mile		1	1	0	0	0	0	0	1	0	1

4.) myphonedata.csv

	V1	V2	V3	red	white	green	yellow	orange	blue
1	red	white	green	1	1	1	0	0	0
2	white	orange		0	1	0	0	1	0
3	white	blue		0	1	0	0	0	1
4	red	white	orange	1	1	0	0	1	0
5	red	blue		1	0	0	0	0	1
6	white	blue		0	1	0	0	0	1
7	red	blue		1	0	0	0	0	1
8	red	white	blue	1	1	0	0	0	1
9	green			0	0	1	0	0	0
10	red	white	blue	1	1	0	0	0	1
11	yellow			0	0	0	1	0	0

5.) transaction_retail.csv

	X.HANGING.	X.HEART.	X.HOLDER.	X.T.LIGHT.	X.WHITE.	NA.
1	'LANTERN'	'METAL'	'WHITE'	*NA*	*NA*	*NA*
2	'COAT'	'CREAM'	'CUPID'	'HANGER'	'HEARTS'	*NA*
3	'BOTTLE'	'FLAG'	'HOT'	'KNITTED'	'UNION'	'WATER'
4	'HEART.'	'HOTTIE'	'RED'	'WHITE'	'WOOLLY'	*NA*
5	'7'	'BABUSHKA'	'BOXES'	'NESTING'	'SET'	*NA*
6	'FROSTED'	'GLASS'	'HOLDER'	'STAR'	'T-LIGHT'	*NA*
7	'HAND'	'JACK'	'UNION'	'WARMER'	*NA*	*NA*
8	'DOT'	'HAND'	'POLKA'	'RED'	'WARMER'	*NA*
9	'ASSORTED'	'BIRD'	'COLOUR'	'ORNAMENT'	*NA*	*NA*
10	'BEDROOM'	'PLAYHOUSE'	'POPPY''S'	*NA*	*NA*	*NA*
11	'KITCHEN'	'PLAYHOUSE'	'POPPY''S'	*NA*	*NA*	*NA*
12	'CHARLOTTE'	'DOLL'	'FELTCRAFT'	'PRINCESS'	*NA*	*NA*

Hints:

1. Business Problem

    1.1. Objective
    1.2. Constraints (if any)

2. Data Pre-processing

    2.1 Data cleaning, Feature Engineering, EDA etc.

3. Model Building

    3.1 Partition the dataset
    3.2 Model(s) - Reasons to choose any algorithm
    3.3 Model(s) Improvement steps
    3.4 Model Evaluation
    3.5 Python and R codes

4. Deployment

    4.1 Deploy solutions using R shiny and Python Flask.

5. Result Share the benefits/impact of the solution - how or in what way the business (client) gets benefit from the solution provided.

Note:

1. For each assignment the solution should be submitted in the format

2. Research and Perform all possible steps for improving the model(s) accuracy

    Ex: Feature Engineering, Hyper Parameter tuning etc.

3. All the codes (executable programs) are running without errors

4. Documentation of the module should be submitted along with R & Python codes, elaborating on every step mentioned here