Implementation of Apriory Algorithm

In [1]:
```
import numpy as np
import pandas as pd
from mlxtend.frequent_patterns import apriori, association_rules
```

In [2]:
```
data = pd.read_excel('Online Retail.xlsx')
```

In [3]:
```
data.head()
```

Out[3]:

	InvoiceNo	StockCode	lower	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Cour
0	536365	85123A	white hanging heart t-light holder	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850.0	Uni Kingd
1	536365	71053	white metal lantern	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	Uni Kingd
2	536365	84406B	cream cupid hearts coat hanger	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850.0	Uni Kingd
3	536365	84029G	knitted union flag hot water bottle	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850.0	Uni Kingd
4	536365	84029E	red woolly hottie white heart.	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850.0	Uni Kingd

In [4]:
```
data.columns
```

Out[4]:
```
Index(['InvoiceNo', 'StockCode', 'lower', 'Description', 'Quantity',
       'InvoiceDate', 'UnitPrice', 'CustomerID', 'Country'],
      dtype='object')
```

In [5]:
```
data.shape
```

Out[5]: (541909, 9)

In [6]:
```
data.isnull().values.any()
```

Out[6]: True

In [7]:
```
data.isnull().sum()
```

Out[7]:
```
InvoiceNo         0
StockCode         0
lower        540093
Description    1454
Quantity          0
InvoiceDate       0
UnitPrice         0
CustomerID    135080
Country           0
dtype: int64
```

Data Preprocessing

In [8]:
```
data['Description'] = data['Description'].str.strip()
data.dropna(axis = 0, subset =['InvoiceNo'], inplace = True)
data['InvoiceNo'] = data['InvoiceNo'].astype('str')
data = data[~data['InvoiceNo'].str.contains('C')]
```

In [9]:
```
data.Country.unique()
```

Out[9]:
```
array(['United Kingdom', 'France', 'Australia', 'Netherlands', 'Germany',
       'Norway', 'EIRE', 'Switzerland', 'Spain', 'Poland', 'Portugal',
       'Italy', 'Belgium', 'Lithuania', 'Japan', 'Iceland',
       'Channel Islands', 'Denmark', 'Cyprus', 'Sweden', 'Finland',
       'Austria', 'Bahrain', 'Israel', 'Greece', 'Hong Kong', 'Singapore',
       'Lebanon', 'United Arab Emirates', 'Saudi Arabia',
       'Czech Republic', 'Canada', 'Unspecified', 'Brazil', 'USA',
       'European Community', 'Malta', 'RSA'], dtype=object)
```

In [10]:
```
basket_France = (data[data['Country'] =="France"]
 .groupby(['InvoiceNo', 'Description'])['Quantity']
 .sum().unstack().reset_index().fillna(0)
 .set_index('InvoiceNo'))
```

In [11]:
```
def hot_encode(x):
    if(x<= 0):
      return 0
    if(x>= 1):
      return 1
```

In [12]:
```
basket_encoded = basket_France.applymap(hot_encode)
basket_France = basket_encoded
```

In [13]:
```
basket_France.head()
```

Out[13]:

Description	10 COLOUR SPACEBOY PEN	12 COLOURED PARTY BALLOONS	12 EGG HOUSE PAINTED WOOD	12 MESSAGE CARDS WITH ENVELOPES	12 PENCIL SMALL TUBE WOODLAND	12 PENCILS SMALL TUBE RED RETROSPOT	12 PENCILS SMALL TUBE SKULL
InvoiceNo							
536370	0	0	0	0	0	0	0
536852	0	0	0	0	0	0	0
536974	0	0	0	0	0	0	0
537065	0	0	0	0	0	0	0
537463	0	0	0	0	0	0	0

5 rows × 1563 columns

Building the model

In [14]:
```
frq_items = apriori(basket_France, min_support = 0.1, use_colnames = True)
rules = association_rules(frq_items, metric ="lift", min_threshold = 1)
rules = rules.sort_values(['confidence', 'lift'], ascending =[False, False])
```

In [15]:
```
print(rules.head())
```

```
                        antecedents                 consequents  \
41        (SET/6 RED SPOTTY PAPER PLATES)    (SET/6 RED SPOTTY PAPER CUPS)
43  (POSTAGE, SET/6 RED SPOTTY PAPER PLATES)    (SET/6 RED SPOTTY PAPER CUPS)
35     (STRAWBERRY LUNCH BOX WITH CUTLERY)                 (POSTAGE)
27     (ROUND SNACK BOXES SET OF4 WOODLAND)                 (POSTAGE)
40          (SET/6 RED SPOTTY PAPER CUPS)  (SET/6 RED SPOTTY PAPER PLATES)

    antecedent support  consequent support   support  confidence     lift  \
41            0.127551            0.137755  0.122449    0.960000  6.968889
43            0.107143            0.137755  0.102041    0.952381  6.913580
35            0.122449            0.765306  0.114796    0.937500  1.225000
27            0.158163            0.765306  0.147959    0.935484  1.222366
40            0.137755            0.127551  0.122449    0.888889  6.968889

    leverage  conviction
41  0.104878   21.556122
43  0.087281   18.107143
35  0.021085    3.755102
27  0.026916    3.637755
40  0.104878    7.852041
```

From the above output, it can be seen that paper cups and plates are bought together in France.

This is because the French have a culture of having a get-together with their friends and family atleast once a week

In [ ]: