

```
pip install mlxtend
```

```
import numpy as np
import pandas as pd
from mlxtend.frequent_patterns import apriori, association_rules
```

Dataset: You can find the dataset in UCI Machine Learning repository. Link :
<http://archive.ics.uci.edu/ml/datasets/Online+Retail>

```
# Loading the dataset
```

```
data = pd.read_excel('/content/drive/MyDrive/Online Retail.xlsx')
```

```
# Showing the data
```

```
data.head()
```

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID
0	536365	85123A	WHITE HANGING HEART T- LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850.C
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.C

```
# The columns in the data
```

```
data.columns
```

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

```
# The shape of the data
```

```
data.shape
```

```
(541909, 8)
```

```
# Checkign whether there is any null values of not
```

```
data.isnull().values.any()
```

```
True
```

```
# As the previous cell told us that there are some null values. So, let's find them!
```

```
data.isnull().sum()
```

```
InvoiceNo      0
StockCode      0
Description    1454
Quantity      0
InvoiceDate    0
UnitPrice      0
CustomerID    135080
Country        0
dtype: int64
```

Data Preprocessing

```
# Stripping extra spaces in the description
data['Description'] = data['Description'].str.strip()
```

```
# Dropping the rows without any invoice number
data.dropna(axis = 0, subset=['InvoiceNo'], inplace = True)
data['InvoiceNo'] = data['InvoiceNo'].astype('str')
```

```
# Dropping all transactions which were done on credit
data = data[~data['InvoiceNo'].str.contains('C')]
```

```
# Let's see the countries in our dataset
```

```
data.Country.unique()
```

```
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)
```

```
# Splitting the data according to the region of transaction
# Transactions done in France
```

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

```
# Defining the hot encoding function to make the data suitable
```

```
def hot_encode(x):
    if(x<= 0):
        return 0
```

```

if(x>= 1):
    return 1

```

Suppose we want to analyze the market trend to France!

Applying one hot encoding

```

basket_encoded = basket_France.applymap(hot_encode)
basket_France = basket_encoded

```

```

basket_France.head()

```

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

```

# Building the model
frq_items = apriori(basket_France, min_support = 0.1, use_colnames = True)

# Collecting the inferred rules in a dataframe
rules = association_rules(frq_items, metric = "lift", min_threshold = 1)
rules = rules.sort_values(['confidence', 'lift'], ascending =[False, False])

rules

```

	antecedents	consequents	antecedent support	consequent support	support	confidence]
41	(SET/6 RED SPOTTY PAPER PLATES)	(SET/6 RED SPOTTY PAPER CUPS)	0.127551	0.137755	0.122449	0.960000	6.968
44	(POSTAGE, SET/6 RED SPOTTY PAPER PLATES)	(SET/6 RED SPOTTY PAPER CUPS)	0.107143	0.137755	0.102041	0.952381	6.913
35	(STRAWBERRY LUNCH BOX WITH CUTLERY)	(POSTAGE)	0.122449	0.765306	0.114796	0.937500	1.225
26	(ROUND SNACK BOXES SET OF4 WOODLAND)	(POSTAGE)	0.158163	0.765306	0.147959	0.935484	1.222
40	(SET/6 RED SPOTTY PAPER CUPS)	(SET/6 RED SPOTTY PAPER PLATES)	0.137755	0.127551	0.122449	0.888889	6.968
10	(PLASTERS IN TIN CIRCUS PARADE)	(POSTAGE)	0.168367	0.765306	0.147959	0.878788	1.148
19	(RABBIT NIGHT LIGHT)	(POSTAGE)	0.188776	0.765306	0.165816	0.878378	1.147
22	(RED TOADSTOOL LED NIGHT LIGHT)	(POSTAGE)	0.181122	0.765306	0.158163	0.873239	1.141
42	(SET/6 RED SPOTTY PAPER CUPS, POSTAGE)	(SET/6 RED SPOTTY PAPER PLATES)	0.117347	0.127551	0.102041	0.869565	6.817
5	(LUNCH BAG WOODLAND)	(POSTAGE)	0.117347	0.765306	0.102041	0.869565	1.136
30	(SET/6 RED SPOTTY PAPER CUPS)	(POSTAGE)	0.137755	0.765306	0.117347	0.851852	1.113
33	(SET/6 RED SPOTTY PAPER PLATES)	(POSTAGE)	0.127551	0.765306	0.107143	0.840000	1.097
1	(LUNCH BAG APPLE DESIGN)	(POSTAGE)	0.125000	0.765306	0.104592	0.836735	1.093
	(SET/6 RED SPOTTY PAPER CUPS)						

25	(REGENCY CAKESTAND 3 TIER)	(POSTAGE)	0.125000	0.765306	0.104592	0.836735	1.093
14	(PLASTERS IN TIN SPACEBOY)	(POSTAGE)	0.137755	0.765306	0.114796	0.833333	1.088
21	(RED RETROSPOT MINI CASES)	(POSTAGE)	0.137755	0.765306	0.114796	0.833333	1.088
43	(SET/6 RED SPOTTY PAPER CUPS, SET/6 RED SPOTTY...	(POSTAGE)	0.122449	0.765306	0.102041	0.833333	1.088
29	(SET/20 RED RETROSPOT PAPER NAPKINS)	(POSTAGE)	0.132653	0.765306	0.109694	0.826923	1.080
17	(PLASTERS IN TIN WOODLAND ANIMALS)	(POSTAGE)	0.170918	0.765306	0.137755	0.805970	1.053
7	(LUNCH BOX WITH CUTLERY RETROSPOT)	(POSTAGE)	0.142857	0.765306	0.114796	0.803571	1.050
47	(SET/6 RED SPOTTY PAPER PLATES)	(SET/6 RED SPOTTY PAPER CUPS, POSTAGE)	0.127551	0.117347	0.102041	0.800000	6.817
39	(SET/6 RED SPOTTY PAPER PLATES)	(SET/20 RED RETROSPOT PAPER NAPKINS)	0.127551	0.132653	0.102041	0.800000	6.030
2	(LUNCH BAG RED RETROSPOT)	(POSTAGE)	0.153061	0.765306	0.122449	0.800000	1.045
38	(SET/20 RED RETROSPOT PAPER NAPKINS)	(SET/6 RED SPOTTY PAPER PLATES)	0.132653	0.127551	0.102041	0.769231	6.030
	(SET/20 RED	(SET/6 RED					

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.

12	TIN	WOODLAND	0.137755	0.170918	0.104592	0.759259	4.442
----	-----	----------	----------	----------	----------	----------	-------

✓ 0s completed at 8:32 AM

