In [ ]:
```python
# Library
import pandas as pd
from mlxtend.frequent_patterns import apriori, association_rules
```

In [ ]:
```python
# Import data year 2009-2010
df_2009 = pd.read_excel('online_retail_II.xlsx', sheet_name='Year 2009-2010', engine

# Import data year 2010-2011
df_2010 = pd.read_excel('online_retail_II.xlsx', sheet_name='Year 2010-2011', engine
```

In [ ]:
```python
df = pd.concat([df_2009,df_2010]) # Menggabungkan sheets
df.drop_duplicates(inplace = True)  # Drop duplicate
df.reset_index(inplace = True) # Reset index
df.drop('index', axis = 1, inplace = True) # Drop index
```

In [ ]:
```python
df['Country'].unique() # Melihat negara yang ada di dataset
```

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

In [ ]:
```python
df['Description'] = df['Description'].str.strip() #Hapus semua spasi di description
df.dropna(axis=0, subset=['Invoice'], inplace=True) #Lepas invoice tanpa angka
df['Invoice'] = df['Invoice'].astype('str') #Ubah invoice ke string
df = df[~df['Invoice'].str.contains('C')] #Hapus kode c dari  invoices
```

In [ ]:
```python
# Buat dataframe baru dengan kolom invoice dan description dengan filter country = U
basket = (df [df ['Country'] =="USA"] #Filter negara USA
          .groupby(['Invoice', 'Description'])['Quantity'] #Group by invoice dan des
          .sum().unstack().reset_index().fillna(0) #Reset index dan isi dengan 0
          .set_index('Invoice')) #Set index invoice
```

In [ ]:
```python
# Buat function untuk mengubah nilai menjadi 0 atau 1
def encode_units(x):
    if x <= 0:
        return 0
    if x >= 1:
        return 1
```

In [ ]:
```python
basket_USA = basket.applymap(encode_units) #Apply function ke dataframe
basket_USA.drop('POSTAGE', inplace=True, axis=1) #Hapus postage dari dataframe
frequent_itemsets = apriori(basket_USA, min_support=0.07, use_colnames=True) #Buat f
rules = association_rules(frequent_itemsets, metric="lift", min_threshold=1) #Buat a
rules = rules.sort_values(['confidence', 'lift'], ascending =[False, False]) #Sort r
rules.head() #Tampilkan 5 data pertama
```

Out[ ]:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	con
0	(60 CAKE CASES DOLLY GIRL DESIGN)	(MINI PAINT SET VINTAGE)	0.1	0.1	0.1	1.0	10.0	0.09	
1	(MINI PAINT SET VINTAGE)	(60 CAKE CASES DOLLY GIRL DESIGN)	0.1	0.1	0.1	1.0	10.0	0.09	
30	(BASKET OF TOADSTOOLS)	(PAPER CHAIN KIT RETRO SPOT)	0.1	0.1	0.1	1.0	10.0	0.09	
31	(PAPER CHAIN KIT RETRO SPOT)	(BASKET OF TOADSTOOLS)	0.1	0.1	0.1	1.0	10.0	0.09	
36	(BIRD HOUSE HOT WATER BOTTLE)	(HOT WATER BOTTLE TEA AND SYMPATHY)	0.1	0.1	0.1	1.0	10.0	0.09	

In [ ]:
```python
rules[ (rules['lift'] >= 7) &
       (rules['confidence'] >= 0.8) ].head() #Tampilkan rules dengan lift >= 7 dan c
```

Out[ ]:

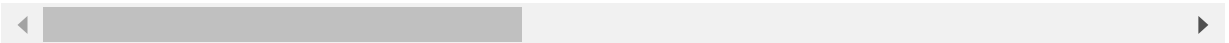|	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	con
0	(60 CAKE CASES DOLLY GIRL DESIGN)	(MINI PAINT SET VINTAGE)	0.1	0.1	0.1	1.0	10.0	0.09	
1	(MINI PAINT SET VINTAGE)	(60 CAKE CASES DOLLY GIRL DESIGN)	0.1	0.1	0.1	1.0	10.0	0.09	
30	(BASKET OF TOADSTOOLS)	(PAPER CHAIN KIT RETRO SPOT)	0.1	0.1	0.1	1.0	10.0	0.09	
31	(PAPER CHAIN KIT RETRO SPOT)	(BASKET OF TOADSTOOLS)	0.1	0.1	0.1	1.0	10.0	0.09	
36	(BIRD HOUSE HOT WATER BOTTLE)	(HOT WATER BOTTLE TEA AND SYMPATHY)	0.1	0.1	0.1	1.0	10.0	0.09	

In [ ]:
```python
basket2 = (df[df['Country'] =="Australia"]
          .groupby(['Invoice', 'Description'])['Quantity'] #Group by invoice dan des
          .sum().unstack().reset_index().fillna(0) #Reset index dan isi dengan 0
          .set_index('Invoice')) #Buat dataframe baru dengan kolom invoice dan descr
basket2.head() #Tampilkan 5 data pertama
```

Out[ ]:

Description	10 COLOUR SPACEBOY PEN	12 PENCIL SMALL TUBE WOODLAND	12 PENCILS SMALL TUBE RED SPOTTY	12 PENCILS TALL TUBE POSY	12 PENCILS TALL TUBE RED RETROSPOT	16 PIECE CUTLERY SET PANTRY DESIGN	20 DOLLY PEGS RETROSPOT	3 H(HAN(MA GARI
Invoice								
489450	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
492744	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
497879	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
498550	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
498617	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

5 rows × 832 columns

In [ ]:
```python
basket_AUS = basket2.applymap(encode_units) #Apply function ke dataframe
basket_AUS.drop('POSTAGE', inplace=True, axis=1) #Hapus postage dari dataframe
frequent_itemsets2 = apriori(basket_AUS, min_support=0.07, use_colnames=True) #Buat
rules2 = association_rules(frequent_itemsets2, metric="lift", min_threshold=1) #Buat
rules2 = rules2.sort_values(['confidence', 'lift'], ascending =[False, False])  #Sor
rules2.head() #Tampilkan 5 data pertama
```

Out[ ]:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage
1	(CIRCUS PARADE LUNCH BOX)	(DOLLY GIRL LUNCH BOX)	0.073684	0.084211	0.073684	1.0	11.875000	0.067479
8	(CIRCUS PARADE LUNCH BOX, SPACEBOY LUNCH BOX)	(DOLLY GIRL LUNCH BOX)	0.073684	0.084211	0.073684	1.0	11.875000	0.067479
10	(CIRCUS PARADE LUNCH BOX)	(DOLLY GIRL LUNCH BOX, SPACEBOY LUNCH BOX)	0.073684	0.084211	0.073684	1.0	11.875000	0.067479
2	(CIRCUS PARADE LUNCH BOX)	(SPACEBOY LUNCH BOX)	0.073684	0.094737	0.073684	1.0	10.555556	0.066704
4	(DOLLY GIRL LUNCH BOX)	(SPACEBOY LUNCH BOX)	0.084211	0.094737	0.084211	1.0	10.555556	0.076233

In [ ]:
```python
rules2[ (rules2['lift'] >= 6) &
        (rules2['confidence'] >= 0.8)].head() #Tampilkan rules dengan lift >= 6 dan
```

Out[ ]:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage
1	(CIRCUS PARADE LUNCH BOX)	(DOLLY GIRL LUNCH BOX)	0.073684	0.084211	0.073684	1.0	11.875000	0.067479
8	(CIRCUS PARADE LUNCH BOX, SPACEBOY LUNCH BOX)	(DOLLY GIRL LUNCH BOX)	0.073684	0.084211	0.073684	1.0	11.875000	0.067479
10	(CIRCUS PARADE LUNCH BOX)	(DOLLY GIRL LUNCH BOX, SPACEBOY LUNCH BOX)	0.073684	0.084211	0.073684	1.0	11.875000	0.067479
2	(CIRCUS PARADE LUNCH BOX)	(SPACEBOY LUNCH BOX)	0.073684	0.094737	0.073684	1.0	10.555556	0.066704
4	(DOLLY GIRL LUNCH BOX)	(SPACEBOY LUNCH BOX)	0.084211	0.094737	0.084211	1.0	10.555556	0.076233

```python
print(basket_USA.sum()) #Jumlah barang yang dibeli di USA bertotal 303
print(basket_AUS.sum()) #Jumlah barang yang dibeli di Australia bertotal 831
```

```
Description
12 PENCILS SMALL TUBE RED RETROSPOT    1
3 RAFFIA RIBBONS 50'S CHRISTMAS        1
3 TRADITIONAl BISCUIT CUTTERS  SET     1
36 DOILIES DOLLY GIRL                  1
36 DOILIES SPACEBOY DESIGN             1
                                      ..
WRAP COWBOYS                           1
WRAP DOLLY GIRL                        1
WRAP ENGLISH ROSE                      1
WRAP I LOVE LONDON                     1
WRAP WEDDING DAY                       1
Length: 303, dtype: int64
Description
10 COLOUR SPACEBOY PEN                 1
12 PENCIL SMALL TUBE WOODLAND          1
12 PENCILS SMALL TUBE RED SPOTTY       1
12 PENCILS TALL TUBE POSY              2
12 PENCILS TALL TUBE RED RETROSPOT     1
                                      ..
WRAP RED VINTAGE DOILY                 1
WRAP VINTAGE LEAF DESIGN               2
WRAP WEDDING DAY                       3
WRAP,SUKI AND FRIENDS                  1
YELLOW GIANT GARDEN THERMOMETER        1
Length: 831, dtype: int64
```