3) implement apriori algorithm on online retail datset ans discuss result

In [ ]:
```
import pandas as pd
import numpy as np

import seaborn as sns
import matplotlib as mlp
import matplotlib.pyplot as plt

from mlxtend.frequent_patterns import apriori
from mlxtend.frequent_patterns import association_rules
```

In [1]:
```
import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))
```

```
/kaggle/input/online-retail-ii-uci/online_retail_II.csv
```

In [2]:
```
transaction_df= pd.read_csv("../input/online-retail-ii-uci/online_retail_II.csv")
transaction_df
```

Out[2]:

	Invoice	StockCode	Description	Quantity	InvoiceDate	Price	Customer ID	Country
0	489434	85048	15CM CHRISTMAS GLASS BALL 20 LIGHTS	12	2009-12-01 07:45:00	6.95	13085.0	United Kingdom
1	489434	79323P	PINK CHERRY LIGHTS	12	2009-12-01 07:45:00	6.75	13085.0	United Kingdom
2	489434	79323W	WHITE CHERRY LIGHTS	12	2009-12-01 07:45:00	6.75	13085.0	United Kingdom
3	489434	22041	RECORD FRAME 7" SINGLE SIZE	48	2009-12-01 07:45:00	2.10	13085.0	United Kingdom
4	489434	21232	STRAWBERRY CERAMIC TRINKET BOX	24	2009-12-01 07:45:00	1.25	13085.0	United Kingdom
...	...	...	...	...	...	...	...	...
1067366	581587	22899	CHILDREN'S APRON DOLLY GIRL	6	2011-12-09 12:50:00	2.10	12680.0	France
1067367	581587	23254	CHILDRENS CUTLERY DOLLY GIRL	4	2011-12-09 12:50:00	4.15	12680.0	France
1067368	581587	23255	CHILDRENS CUTLERY CIRCUS PARADE	4	2011-12-09 12:50:00	4.15	12680.0	France
1067369	581587	22138	BAKING SET 9 PIECE RETROSPOT	3	2011-12-09 12:50:00	4.95	12680.0	France
1067370	581587	POST	POSTAGE	1	2011-12-09 12:50:00	18.00	12680.0	France

1067371 rows × 8 columns

In [3]:
```
transaction_df = transaction_df[transaction_df.Country=='France']
```

In [4]:
```
transaction_filtered = transaction_df[['Invoice','Description','Quantity']].copy()
transaction_filtered
```

Out[4]:

	Invoice	Description	Quantity
71	489439	CHRISTMAS PUDDING TRINKET POT	12
72	489439	BAKING SET 9 PIECE RETROSPOT	9
73	489439	RETRO SPOT TEA SET CERAMIC 11 PC	9
74	489439	LUNCHBOX WITH CUTLERY RETROSPOT	12
75	489439	BLACK/BLUE DOTS RUFFLED UMBRELLA	3
...	...	...	...
1067366	581587	CHILDREN'S APRON DOLLY GIRL	6
1067367	581587	CHILDRENS CUTLERY DOLLY GIRL	4
1067368	581587	CHILDRENS CUTLERY CIRCUS PARADE	4
1067369	581587	BAKING SET 9 PIECE RETROSPOT	3
1067370	581587	POSTAGE	1

14330 rows × 3 columns

In [5]:
```
transaction_filtered.sort_values(by='Quantity', ascending=True)
```

Out[5]:

	Invoice	Description	Quantity
359669	C524235	SET/6 FRUIT SALAD PAPER CUPS	-7128
359670	C524235	SET/6 FRUIT SALAD PAPER PLATES	-7008
359630	C524235	POP ART PEN CASE & PENS	-5184
359636	C524235	MULTICOLOUR SPRING FLOWER MUG	-4992
359653	C524235	BLACK SILVER FLOWER T-LIGHT HOLDER	-4752
...	...	...	...
298916	518505	BLACK SILVER FLOWER T-LIGHT HOLDER	4752
298933	518505	MULTICOLOUR SPRING FLOWER MUG	4992
298953	518505	POP ART PEN CASE & PENS	5184
298941	518505	SET/6 FRUIT SALAD PAPER PLATES	7008
298942	518505	SET/6 FRUIT SALAD PAPER CUPS	7128

14330 rows × 3 columns

In [6]:
```
transaction_filtered = transaction_filtered[transaction_filtered.Quantity > 0 ]
transaction_filtered.sort_values(by='Quantity', ascending=True)
```

Out[6]:

	Invoice	Description	Quantity
1067370	581587	POSTAGE	1
565638	539727	POSTAGE	1
571557	540351	WOODEN SKITTLES GARDEN SET	1
571558	540351	WOODEN CROQUET GARDEN SET	1
571561	540351	TOOL BOX SOFT TOY	1
...	...	...	...
298916	518505	BLACK SILVER FLOWER T-LIGHT HOLDER	4752
298933	518505	MULTICOLOUR SPRING FLOWER MUG	4992
298953	518505	POP ART PEN CASE & PENS	5184
298941	518505	SET/6 FRUIT SALAD PAPER PLATES	7008
298942	518505	SET/6 FRUIT SALAD PAPER CUPS	7128

13941 rows × 3 columns

In [7]:
```
transaction_filtered['Quantity']= [1]*len(transaction_filtered)
```

```
/opt/conda/lib/python3.7/site-packages/ipykernel_launcher.py:1: SettingWithCopyWarnin
g:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/us
er_guide/indexing.html#returning-a-view-versus-a-copy
  """Entry point for launching an IPython kernel.
```

In [9]:
```
transaction_filtered
```

Out[9]:

	Invoice	Description	Quantity
71	489439	CHRISTMAS PUDDING TRINKET POT	1
72	489439	BAKING SET 9 PIECE RETROSPOT	1
73	489439	RETRO SPOT TEA SET CERAMIC 11 PC	1
74	489439	LUNCHBOX WITH CUTLERY RETROSPOT	1
75	489439	BLACK/BLUE DOTS RUFFLED UMBRELLA	1
...	...	...	...
1067366	581587	CHILDREN'S APRON DOLLY GIRL	1
1067367	581587	CHILDRENS CUTLERY DOLLY GIRL	1
1067368	581587	CHILDRENS CUTLERY CIRCUS PARADE	1
1067369	581587	BAKING SET 9 PIECE RETROSPOT	1
1067370	581587	POSTAGE	1

13941 rows × 3 columns

In [10]:
```
invoice = list(transaction_filtered.Invoice)
index_no = [invoice[index] for index in np.arange(len(invoice)) if not invoice[index].
transaction_filtered[transaction_filtered['Invoice'].isin(index_no)]
```

Out[10]:

Invoice	Description	Quantity

In [11]:
```
index_no
```

Out[11]:
```
[]
```

In [12]:
```
transaction_filtered= transaction_filtered[~transaction_filtered['Invoice'].isin(index
```

In [13]:
```
invoice = list(transaction_filtered.Invoice)
index_no = [index for index in np.arange(len(invoice)) if not invoice[index].isnumeric
transaction_filtered.iloc[index_no,:]
```

Out[13]:

Invoice	Description	Quantity

In [14]:
```
temp_df = transaction_filtered[transaction_filtered.Description != transaction_filtere
temp_df
```

Out[14]:

Invoice	Description	Quantity

In [15]:
```
for invoice in list(temp_df.Invoice):
    if len(transaction_filtered[transaction_filtered.Invoice == invoice]) > 1:
        print((str)(invoice))
        temp = transaction_filtered[transaction_filtered.Invoice == invoice].groupby([
        if len(list(set(temp)))>0 :
            print(temp)
```

In [16]:
```
transaction_filtered.dropna(axis=0, inplace=True)
```

In [17]:
```
transaction_filtered
```

Out[17]:

	Invoice	Description	Quantity
71	489439	CHRISTMAS PUDDING TRINKET POT	1
72	489439	BAKING SET 9 PIECE RETROSPOT	1
73	489439	RETRO SPOT TEA SET CERAMIC 11 PC	1
74	489439	LUNCHBOX WITH CUTLERY RETROSPOT	1
75	489439	BLACK/BLUE DOTS RUFFLED UMBRELLA	1
...	...	...	...
1067366	581587	CHILDREN'S APRON DOLLY GIRL	1
1067367	581587	CHILDRENS CUTLERY DOLLY GIRL	1
1067368	581587	CHILDRENS CUTLERY CIRCUS PARADE	1
1067369	581587	BAKING SET 9 PIECE RETROSPOT	1
1067370	581587	POSTAGE	1

13941 rows × 3 columns

In [18]:
```
def return_one(x):
    return 1
```

In [19]:
```
table = pd.pivot_table(transaction_filtered, values='Quantity', index=['Invoice'],
                       columns=['Description'], aggfunc=return_one, fill_value=0)
table
```

Out[19]:

Description	50'S CHRISTMAS GIFT BAG LARGE	DOLLY GIRL BEAKER	FLAMINGO LIGHTS	I LOVE LONDON MINI BACKPACK	LARGE SKULL WINDMILL	NINE DRAWER OFFICE TIDY	RED/WHITE DOT MINI CASES	SE T TOW I LC LOND
Invoice								
489439	0	0	0	0	0	0	0	
489557	0	0	0	0	0	0	0	
489883	0	0	0	0	0	0	1	
490139	0	0	0	0	0	0	0	
490152	0	0	0	0	0	0	1	
...	...	...	...	...	...	...	...	
580986	0	0	0	0	0	0	0	
581001	0	0	0	0	0	0	0	
581171	0	0	0	0	0	0	0	
581279	0	0	0	0	0	0	0	
581587	0	0	0	0	0	0	0	

622 rows × 2211 columns

In [20]:
```
frequent_itemsets = apriori(table, min_support=0.01, use_colnames=True)
frequent_itemsets
```

Out[20]:

	support	itemsets
0	0.014469	(DOLLY GIRL BEAKER)
1	0.027331	(RED/WHITE DOT MINI CASES)
2	0.025723	(SET 2 TEA TOWELS I LOVE LONDON)
3	0.025723	(SPACEBOY BABY GIFT SET)
4	0.020900	(10 COLOUR SPACEBOY PEN)
...	...	...
8174	0.011254	(SET/6 RED SPOTTY PAPER PLATES, POSTAGE, PACK ...
8175	0.017685	(SET/6 RED SPOTTY PAPER PLATES, POSTAGE, PACK ...
8176	0.011254	(SET/6 RED SPOTTY PAPER PLATES, PACK OF 20 SKU...
8177	0.011254	(SET/6 RED SPOTTY PAPER PLATES, POSTAGE, PACK ...
8178	0.011254	(SET/6 RED SPOTTY PAPER PLATES, POSTAGE, SET O...

8179 rows × 2 columns

In [21]:
```
rules = association_rules(frequent_itemsets, metric="lift", min_threshold=1)
rules
```

Out[21]:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage
0	(DOLLY GIRL BEAKER)	(DOLLY GIRL CHILDRENS BOWL)	0.014469	0.028939	0.011254	0.777778	26.876543	0.010835
1	(DOLLY GIRL CHILDRENS BOWL)	(DOLLY GIRL BEAKER)	0.028939	0.014469	0.011254	0.388889	26.876543	0.010835
2	(DOLLY GIRL BEAKER)	(POSTAGE)	0.014469	0.749196	0.011254	0.777778	1.038150	0.000414
3	(POSTAGE)	(DOLLY GIRL BEAKER)	0.749196	0.014469	0.011254	0.015021	1.038150	0.000414
4	(DOLLY GIRL BEAKER)	(SPACEBOY CHILDRENS BOWL)	0.014469	0.032154	0.011254	0.777778	24.188889	0.010789
...	...	...	...	...	...	...	...	...
68519	(SET OF 9 BLACK SKULL BALLOONS)	(SET/6 RED SPOTTY PAPER PLATES, POSTAGE, SET/2...	0.057878	0.019293	0.011254	0.194444	10.078704	0.010137
68520	(SET/20 RED RETROSPOT PAPER NAPKINS)	(SET/6 RED SPOTTY PAPER PLATES, POSTAGE, SET O...	0.094855	0.016077	0.011254	0.118644	7.379661	0.009729
68521	(PACK OF 6 SKULL PAPER PLATES)	(SET/6 RED SPOTTY PAPER PLATES, POSTAGE, SET O...	0.048232	0.011254	0.011254	0.233333	20.733333	0.010711
68522	(PACK OF 6 SKULL PAPER CUPS)	(SET/6 RED SPOTTY PAPER PLATES, POSTAGE, SET O...	0.057878	0.011254	0.011254	0.194444	17.277778	0.010603
68523	(SET/6 RED SPOTTY PAPER CUPS)	(SET/6 RED SPOTTY PAPER PLATES, POSTAGE, SET O...	0.136656	0.011254	0.011254	0.082353	7.317647	0.009716

68524 rows × 9 columns

In [22]:
```
rules.sort_values(by=['support','confidence'], ascending=False)
```

Out[22]:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage
3654	(RED TOADSTOOL LED NIGHT LIGHT)	(POSTAGE)	0.212219	0.749196	0.184887	0.871212	1.162863	0.025894
3655	(POSTAGE)	(RED TOADSTOOL LED NIGHT LIGHT)	0.749196	0.212219	0.184887	0.246781	1.162863	0.025894
3684	(ROUND SNACK BOXES SET OF4 WOODLAND)	(POSTAGE)	0.173633	0.749196	0.157556	0.907407	1.211175	0.027471
3685	(POSTAGE)	(ROUND SNACK BOXES SET OF4 WOODLAND)	0.749196	0.173633	0.157556	0.210300	1.211175	0.027471
3258	(PLASTERS IN TIN CIRCUS PARADE)	(POSTAGE)	0.167203	0.749196	0.136656	0.817308	1.090913	0.011388
...	...	...	...	...	...	...	...	...
67513	(POSTAGE)	(PLASTERS IN TIN STRONGMAN, STRAWBERRY LUNCH B...	0.749196	0.011254	0.011254	0.015021	1.334764	0.002823
67636	(POSTAGE)	(RED RETROSPOT MINI CASES, RED TOADSTOOL LED N...	0.749196	0.011254	0.011254	0.015021	1.334764	0.002823
68014	(POSTAGE)	(SET/6 RED SPOTTY PAPER PLATES, PACK OF 20 SKU...	0.749196	0.011254	0.011254	0.015021	1.334764	0.002823
68392	(POSTAGE)	(SET/6 RED SPOTTY PAPER PLATES, PACK OF 20 SKU...	0.749196	0.011254	0.011254	0.015021	1.334764	0.002823
68518	(POSTAGE)	(SET/6 RED SPOTTY PAPER PLATES, SET OF 9 BLACK...	0.749196	0.012862	0.011254	0.015021	1.167918	0.001618

68524 rows × 9 columns