In [1]:
```
!pip install pyspark
```

```
Requirement already satisfied: pyspark in c:\users\dhruv\anaconda3\lib\site-packag
es (3.2.1)
Requirement already satisfied: py4j==0.10.9.3 in c:\users\dhruv\anaconda3\lib\site
-packages (from pyspark) (0.10.9.3)
```

In [2]:
```
from pyspark.sql import SparkSession
import pandas as pd
from pyspark.sql.functions import col, isnan, when, count, isnull
import matplotlib.pyplot as plt
import seaborn as sns
from pyspark.ml.feature import StringIndexer
from pyspark.sql import SQLContext, SparkSession
from pyspark.sql.functions import *
import numpy as np
```

In [3]:
```
spark = SparkSession.builder.appName("Apriori").getOrCreate()
df = spark.read.csv('D:\DAIICT\SEMESTER_2\BDP\dataset\Assignment-1_Data.csv', sep=
df.printSchema()
```

```
root
 |-- BillNo: string (nullable = true)
 |-- Itemname: string (nullable = true)
 |-- Quantity: string (nullable = true)
 |-- Date: string (nullable = true)
 |-- Price: string (nullable = true)
 |-- CustomerID: string (nullable = true)
 |-- Country: string (nullable = true)
```

In [4]:
```
df.show()
```

```
+------+--------------------+--------+--------------+-----+----------+----------
----+
|BillNo|            Itemname|Quantity|          Date|Price|CustomerID|       Cou
ntry|
+------+--------------------+--------+--------------+-----+----------+----------
----+
|536365|WHITE HANGING HEA...|       6|01.12.2010 08:26| 2,55|     17850|United Kin
gdom|
|536365| WHITE METAL LANTERN|       6|01.12.2010 08:26| 3,39|     17850|United Kin
gdom|
|536365|CREAM CUPID HEART...|       8|01.12.2010 08:26| 2,75|     17850|United Kin
gdom|
|536365|KNITTED UNION FLA...|       6|01.12.2010 08:26| 3,39|     17850|United Kin
gdom|
|536365|RED WOOLLY HOTTIE...|       6|01.12.2010 08:26| 3,39|     17850|United Kin
gdom|
|536365|SET 7 BABUSHKA NE...|       2|01.12.2010 08:26| 7,65|     17850|United Kin
gdom|
|536365|GLASS STAR FROSTE...|       6|01.12.2010 08:26| 4,25|     17850|United Kin
gdom|
|536366|HAND WARMER UNION...|       6|01.12.2010 08:28| 1,85|     17850|United Kin
gdom|
|536366|HAND WARMER RED P...|       6|01.12.2010 08:28| 1,85|     17850|United Kin
gdom|
|536367|ASSORTED COLOUR B...|      32|01.12.2010 08:34| 1,69|     13047|United Kin
gdom|
|536367|POPPY'S PLAYHOUSE...|       6|01.12.2010 08:34|  2,1|     13047|United Kin
gdom|
|536367|POPPY'S PLAYHOUSE...|       6|01.12.2010 08:34|  2,1|     13047|United Kin
gdom|
|536367|FELTCRAFT PRINCES...|       8|01.12.2010 08:34| 3,75|     13047|United Kin
gdom|
|536367|IVORY KNITTED MUG...|       6|01.12.2010 08:34| 1,65|     13047|United Kin
gdom|
|536367|BOX OF 6 ASSORTED...|       6|01.12.2010 08:34| 4,25|     13047|United Kin
gdom|
|536367|BOX OF VINTAGE JI...|       3|01.12.2010 08:34| 4,95|     13047|United Kin
gdom|
|536367|BOX OF VINTAGE AL...|       2|01.12.2010 08:34| 9,95|     13047|United Kin
gdom|
|536367|HOME BUILDING BLO...|       3|01.12.2010 08:34| 5,95|     13047|United Kin
gdom|
|536367|LOVE BUILDING BLO...|       3|01.12.2010 08:34| 5,95|     13047|United Kin
gdom|
|536367|RECIPE BOX WITH M...|       4|01.12.2010 08:34| 7,95|     13047|United Kin
gdom|
+------+--------------------+--------+--------------+-----+----------+----------
----+
only showing top 20 rows
```

In [5]:
```
summary1 = df.summary().toPandas()
summary1
```

Out[5]:

	summary	BillNo	Itemname	Quantity	Date	Price
0	count	522064	520609	522064	522064	522064
1	mean	559950.7852856276	None	10.090435272303779	None	15.576812289966394
2	stddev	13452.750899837123	None	161.11052518229036	None	72.62169390055651
3	min	536365	"ASSORTED FLOWER COLOUR ""LEIS"""	-1	01.02.2011 08:23	-11062,06
4	25%	547892.0	None	1.0	None	0.0
5	50%	560601.0	None	3.0	None	0.0
6	75%	571895.0	None	10.0	None	18.0
7	max	A563187	wrongly sold sets	992	31.10.2011 17:19	99,96

Missing Values

In [6]:
```
missing_counts = df.select([count(when(isnull(c) | isnan(c), c)).alias(c) for c in
missing_counts.show()
```

```
+------+--------+--------+----+-----+----------+-------+
|BillNo|Itemname|Quantity|Date|Price|CustomerID|Country|
+------+--------+--------+----+-----+----------+-------+
|     0|    1455|       0|   0|    0|    134041|      0|
+------+--------+--------+----+-----+----------+-------+
```

Filling Null values in CustomerID with 999999

In [7]:
```
df = df.fillna(999999,subset='CustomerID')
```

In [8]:
```
summary2 = df.describe().toPandas()
summary2
```

Out[8]:

	summary	BillNo	Itemname	Quantity	Date	Price
0	count	522064	520609	522064	522064	522064
1	mean	559950.7852856276	None	10.090435272303779	None	15.576812289966394
2	stddev	13452.750899837	None	161.11052518229036	None	72.62169390055651
3	min	536365	"ASSORTED FLOWER COLOUR ""LEIS"""	-1	01.02.2011 08:23	-11062,06
4	max	A563187	wrongly sold sets	992	31.10.2011 17:19	99,96

We can see that min Quantity is -1 which is not possible.

In [9]:
```
df.filter(df.Quantity<=0).show()
```

```
+------+--------+--------+---------------+-----+----------+--------------+
|BillNo|Itemname|Quantity|           Date|Price|CustomerID|       Country|
+------+--------+--------+---------------+-----+----------+--------------+
|536589|    null|     -10|01.12.2010 16:50|    0|      null|United Kingdom|
|536764|    null|     -38|02.12.2010 14:42|    0|      null|United Kingdom|
|536996|    null|     -20|03.12.2010 15:30|    0|      null|United Kingdom|
|536997|    null|     -20|03.12.2010 15:30|    0|      null|United Kingdom|
|536998|    null|      -6|03.12.2010 15:30|    0|      null|United Kingdom|
|537000|    null|     -22|03.12.2010 15:32|    0|      null|United Kingdom|
|537001|    null|      -6|03.12.2010 15:33|    0|      null|United Kingdom|
|537003|    null|      -2|03.12.2010 15:33|    0|      null|United Kingdom|
|537004|    null|     -30|03.12.2010 15:34|    0|      null|United Kingdom|
|537005|    null|     -70|03.12.2010 15:35|    0|      null|United Kingdom|
|537006|    null|    -130|03.12.2010 15:36|    0|      null|United Kingdom|
|537007|    null|     -80|03.12.2010 15:36|    0|      null|United Kingdom|
|537008|    null|    -120|03.12.2010 15:37|    0|      null|United Kingdom|
|537009|    null|     -80|03.12.2010 15:38|    0|      null|United Kingdom|
|537010|    null|     -40|03.12.2010 15:38|    0|      null|United Kingdom|
|537011|    null|      -5|03.12.2010 15:38|    0|      null|United Kingdom|
|537012|    null|     -12|03.12.2010 15:39|    0|      null|United Kingdom|
|537013|    null|     -25|03.12.2010 15:40|    0|      null|United Kingdom|
|537014|    null|     -20|03.12.2010 15:40|    0|      null|United Kingdom|
|537015|    null|     -14|03.12.2010 15:41|    0|      null|United Kingdom|
+------+--------+--------+---------------+-----+----------+--------------+
only showing top 20 rows
```

Filtering out these rows as we can see it is noise in the data.

In [10]:
```
df = df.filter(df.Quantity>0)
df.show()
```

```
+------+--------------------+--------+---------------+-----+----------+----------
----+
|BillNo|            Itemname|Quantity|           Date|Price|CustomerID|       Cou
ntry|
+------+--------------------+--------+---------------+-----+----------+----------
----+
|536365|WHITE HANGING HEA...|       6|01.12.2010 08:26| 2,55|     17850|United Kin
gdom|
|536365| WHITE METAL LANTERN|       6|01.12.2010 08:26| 3,39|     17850|United Kin
gdom|
|536365|CREAM CUPID HEART...|       8|01.12.2010 08:26| 2,75|     17850|United Kin
gdom|
|536365|KNITTED UNION FLA...|       6|01.12.2010 08:26| 3,39|     17850|United Kin
gdom|
|536365|RED WOOLLY HOTTIE...|       6|01.12.2010 08:26| 3,39|     17850|United Kin
gdom|
|536365|SET 7 BABUSHKA NE...|       2|01.12.2010 08:26| 7,65|     17850|United Kin
gdom|
|536365|GLASS STAR FROSTE...|       6|01.12.2010 08:26| 4,25|     17850|United Kin
gdom|
|536366|HAND WARMER UNION...|       6|01.12.2010 08:28| 1,85|     17850|United Kin
gdom|
|536366|HAND WARMER RED P...|       6|01.12.2010 08:28| 1,85|     17850|United Kin
gdom|
|536367|ASSORTED COLOUR B...|      32|01.12.2010 08:34| 1,69|     13047|United Kin
gdom|
|536367|POPPY'S PLAYHOUSE...|       6|01.12.2010 08:34|  2,1|     13047|United Kin
gdom|
|536367|POPPY'S PLAYHOUSE...|       6|01.12.2010 08:34|  2,1|     13047|United Kin
gdom|
|536367|FELTCRAFT PRINCES...|       8|01.12.2010 08:34| 3,75|     13047|United Kin
gdom|
|536367|IVORY KNITTED MUG...|       6|01.12.2010 08:34| 1,65|     13047|United Kin
gdom|
|536367|BOX OF 6 ASSORTED...|       6|01.12.2010 08:34| 4,25|     13047|United Kin
gdom|
|536367|BOX OF VINTAGE JI...|       3|01.12.2010 08:34| 4,95|     13047|United Kin
gdom|
|536367|BOX OF VINTAGE AL...|       2|01.12.2010 08:34| 9,95|     13047|United Kin
gdom|
|536367|HOME BUILDING BLO...|       3|01.12.2010 08:34| 5,95|     13047|United Kin
gdom|
|536367|LOVE BUILDING BLO...|       3|01.12.2010 08:34| 5,95|     13047|United Kin
gdom|
|536367|RECIPE BOX WITH M...|       4|01.12.2010 08:34| 7,95|     13047|United Kin
gdom|
+------+--------------------+--------+---------------+-----+----------+----------
----+
only showing top 20 rows
```

Checking Null Values in Itemnames

In [11]:
```
missing_counts = df.select([count(when(isnull('Itemname') | isnan('Itemname'), 'It
missing_counts.show()
```

```
+--------+
|Itemname|
+--------+
|     592|
+--------+
```

Removing rows where Itemnames is missing

In [12]:
```
df = df.dropna()
```

In [13]:
```
# NULl values
df.select([count(when(isnull(c) | isnan(c), c)).alias(c) for c in df.columns]).sho
```

```
+------+--------+--------+----+-----+----------+-------+
|BillNo|Itemname|Quantity|Date|Price|CustomerID|Country|
+------+--------+--------+----+-----+----------+-------+
|     0|       0|       0|   0|    0|         0|      0|
+------+--------+--------+----+-----+----------+-------+
```

In [14]:
```
df = df.withColumn('Price', regexp_replace('Price', ',', '.'))
df = df.withColumn('Total_Price', round(df.Price * df.Quantity, 2))
df = df.withColumn('Hour', split('Date', ' ')[1])
df = df.withColumn('Date', split('Date', ' ')[0])
df = df.withColumn( 'Date', concat_ws('-', split('Date', '\.')[2], split('Date', '
df = df.withColumn('Date', to_date('Date'))
df.show()
```

```
+------+--------------------+--------+----------+-----+----------+--------------+-
----------+-----+
|BillNo|            Itemname|Quantity|      Date|Price|CustomerID|       Country|T
otal_Price| Hour|
+------+--------------------+--------+----------+-----+----------+--------------+-
----------+-----+
|536365|WHITE HANGING HEA...|       6|2010-12-01| 2.55|     17850|United Kingdom|
15.3|08:26|
|536365| WHITE METAL LANTERN|       6|2010-12-01| 3.39|     17850|United Kingdom|
20.34|08:26|
|536365|CREAM CUPID HEART...|       8|2010-12-01| 2.75|     17850|United Kingdom|
22.0|08:26|
|536365|KNITTED UNION FLA...|       6|2010-12-01| 3.39|     17850|United Kingdom|
20.34|08:26|
|536365|RED WOOLLY HOTTIE...|       6|2010-12-01| 3.39|     17850|United Kingdom|
20.34|08:26|
|536365|SET 7 BABUSHKA NE...|       2|2010-12-01| 7.65|     17850|United Kingdom|
15.3|08:26|
|536365|GLASS STAR FROSTE...|       6|2010-12-01| 4.25|     17850|United Kingdom|
25.5|08:26|
|536366|HAND WARMER UNION...|       6|2010-12-01| 1.85|     17850|United Kingdom|
11.1|08:28|
|536366|HAND WARMER RED P...|       6|2010-12-01| 1.85|     17850|United Kingdom|
11.1|08:28|
|536367|ASSORTED COLOUR B...|      32|2010-12-01| 1.69|     13047|United Kingdom|
54.08|08:34|
|536367|POPPY'S PLAYHOUSE...|       6|2010-12-01|  2.1|     13047|United Kingdom|
12.6|08:34|
|536367|POPPY'S PLAYHOUSE...|       6|2010-12-01|  2.1|     13047|United Kingdom|
12.6|08:34|
|536367|FELTCRAFT PRINCES...|       8|2010-12-01| 3.75|     13047|United Kingdom|
30.0|08:34|
|536367|IVORY KNITTED MUG...|       6|2010-12-01| 1.65|     13047|United Kingdom|
9.9|08:34|
|536367|BOX OF 6 ASSORTED...|       6|2010-12-01| 4.25|     13047|United Kingdom|
25.5|08:34|
|536367|BOX OF VINTAGE JI...|       3|2010-12-01| 4.95|     13047|United Kingdom|
14.85|08:34|
|536367|BOX OF VINTAGE AL...|       2|2010-12-01| 9.95|     13047|United Kingdom|
19.9|08:34|
|536367|HOME BUILDING BLO...|       3|2010-12-01| 5.95|     13047|United Kingdom|
17.85|08:34|
|536367|LOVE BUILDING BLO...|       3|2010-12-01| 5.95|     13047|United Kingdom|
17.85|08:34|
|536367|RECIPE BOX WITH M...|       4|2010-12-01| 7.95|     13047|United Kingdom|
31.8|08:34|
+------+--------------------+--------+----------+-----+----------+--------------+-
----------+-----+
only showing top 20 rows
```

Apriori Method

In [15]:
```
from pyspark.ml.feature import OneHotEncoder
from pyspark.ml.feature import StringIndexer, CountVectorizer
from pyspark.sql.types import IntegerType
```

In [16]:
```
indexer = StringIndexer(inputCol = 'Itemname', outputCol = 'Item_Index')
data_indexed = indexer.fit(df).transform(df)
data_indexed.show()
```

```
+------+--------------------+--------+----------+-----+----------+--------------+-
----------+-----+----------+
|BillNo|            Itemname|Quantity|      Date|Price|CustomerID|       Country|T
otal_Price| Hour|Item_Index|
+------+--------------------+--------+----------+-----+----------+--------------+-
----------+-----+----------+
|536365|WHITE HANGING HEA...|       6|2010-12-01| 2.55|     17850|United Kingdom|
15.3|08:26|       0.0|
|536365| WHITE METAL LANTERN|       6|2010-12-01| 3.39|     17850|United Kingdom|
20.34|08:26|     437.0|
|536365|CREAM CUPID HEART...|       8|2010-12-01| 2.75|     17850|United Kingdom|
22.0|08:26|     448.0|
|536365|KNITTED UNION FLA...|       6|2010-12-01| 3.39|     17850|United Kingdom|
20.34|08:26|     267.0|
|536365|RED WOOLLY HOTTIE...|       6|2010-12-01| 3.39|     17850|United Kingdom|
20.34|08:26|     272.0|
|536365|SET 7 BABUSHKA NE...|       2|2010-12-01| 7.65|     17850|United Kingdom|
15.3|08:26|     324.0|
|536365|GLASS STAR FROSTE...|       6|2010-12-01| 4.25|     17850|United Kingdom|
25.5|08:26|    1015.0|
|536366|HAND WARMER UNION...|       6|2010-12-01| 1.85|     17850|United Kingdom|
11.1|08:28|     134.0|
|536366|HAND WARMER RED P...|       6|2010-12-01| 1.85|     17850|United Kingdom|
11.1|08:28|    2587.0|
|536367|ASSORTED COLOUR B...|      32|2010-12-01| 1.69|     13047|United Kingdom|
54.08|08:34|       3.0|
|536367|POPPY'S PLAYHOUSE...|       6|2010-12-01|  2.1|     13047|United Kingdom|
12.6|08:34|     281.0|
|536367|POPPY'S PLAYHOUSE...|       6|2010-12-01|  2.1|     13047|United Kingdom|
12.6|08:34|     233.0|
|536367|FELTCRAFT PRINCES...|       8|2010-12-01| 3.75|     13047|United Kingdom|
30.0|08:34|     172.0|
|536367|IVORY KNITTED MUG...|       6|2010-12-01| 1.65|     13047|United Kingdom|
9.9|08:34|    1797.0|
|536367|BOX OF 6 ASSORTED...|       6|2010-12-01| 4.25|     13047|United Kingdom|
25.5|08:34|    1527.0|
|536367|BOX OF VINTAGE JI...|       3|2010-12-01| 4.95|     13047|United Kingdom|
14.85|08:34|     550.0|
|536367|BOX OF VINTAGE AL...|       2|2010-12-01| 9.95|     13047|United Kingdom|
19.9|08:34|     504.0|
|536367|HOME BUILDING BLO...|       3|2010-12-01| 5.95|     13047|United Kingdom|
17.85|08:34|      49.0|
|536367|LOVE BUILDING BLO...|       3|2010-12-01| 5.95|     13047|United Kingdom|
17.85|08:34|      86.0|
|536367|RECIPE BOX WITH M...|       4|2010-12-01| 7.95|     13047|United Kingdom|
31.8|08:34|    3446.0|
+------+--------------------+--------+----------+-----+----------+--------------+-
----------+-----+----------+
only showing top 20 rows
```

In [17]:
```
df_group = data_indexed[['BillNo','Country','Itemname']].distinct()
df_group = df_group.groupBy('BillNo', 'Country').agg(collect_list('Itemname').alia
df_group.show()
```

```
+------+--------------+--------------------+
|BillNo|       Country|              Basket|
+------+--------------+--------------------+
|536394|United Kingdom|[FANCY FONT BIRTH...|
|536402|United Kingdom|[PAPER CHAIN KIT ...|
|536534|United Kingdom|[HAND WARMER SCOT...|
|536535|United Kingdom|[DOORMAT FANCY FO...|
|536575|United Kingdom|[LADS ONLY TISSUE...|
|536611|United Kingdom|[RED  HARMONICA I...|
|536628|United Kingdom|[CREAM CUPID HEAR...|
|536629|United Kingdom|[HAND WARMER UNIO...|
|536741|United Kingdom|[ASSORTED COLOUR ...|
|536771|United Kingdom|[JAM MAKING SET W...|
|536786|United Kingdom|[BAKING SET SPACE...|
|536806|United Kingdom|[CHRISTMAS TREE S...|
|536832|United Kingdom|[KEY CABINET MA C...|
|536838|United Kingdom|[FENG SHUI PILLAR...|
|536841|United Kingdom|[FIRST AID TIN, P...|
|536944|         Spain|[LUNCH BAG  BLACK...|
|536945|United Kingdom|[PLASTERS IN TIN ...|
|536972|United Kingdom|[RED KITCHEN SCAL...|
|536988|United Kingdom|[WHITE SKULL HOT ...|
|536992|United Kingdom|[CHARLIE + LOLA R...|
+------+--------------+--------------------+
only showing top 20 rows
```

In [18]:
```
countries = []
for country in df_group[['Country']].distinct().collect():
    countries.append(country['Country'])
```

In [19]:
```
from pyspark.ml.fpm import FPGrowth
```

In [20]:
```
minSupport=0.1
minConfidence=0.8
results = {}
```

In [21]:
```
for country in countries:
    print(country)
    fpGrowth = FPGrowth(itemsCol="Basket", minSupport=minSupport, minConfidence=mi
    model = fpGrowth.fit(df_group.filter(df_group['Country']==country))
#     model.associationRules.show()
    results[country] = model.associationRules
```

```
Sweden
C:\Users\dhruv\anaconda3\lib\site-packages\pyspark\sql\context.py:125: FutureWarni
ng: Deprecated in 3.0.0. Use SparkSession.builder.getOrCreate() instead.
  warnings.warn(
```

```
Germany
France
Belgium
Italy
Lithuania
Norway
Spain
Iceland
Switzerland
Japan
Poland
Portugal
Australia
Austria
United Kingdom
Netherlands
Singapore
Greece
Israel
Saudi Arabia
United Arab Emirates
Lebanon
Unspecified
USA
Brazil
Malta
Bahrain
RSA
```

In [22]:
```
from pyspark.sql.functions import lit
country = countries[0]
apriori = results[country].withColumn('country', lit(country))

for country in countries[1:5]:
    df_temp = results[country].withColumn('country', lit(country))
    apriori =  apriori.union(df_temp)
```

In [24]:
```
sqlCtx = SQLContext(spark)
apriori.createOrReplaceTempView("apriori")
rules = sqlCtx.sql("""SELECT antecedent, consequent, COUNT(DISTINCT country) as n_
rules.sort(col('n_country').desc(), col('mean_lift').desc()).show(30, truncate=Fal
```

```
C:\Users\dhruv\anaconda3\lib\site-packages\pyspark\sql\context.py:77: FutureWarnin
g: Deprecated in 3.0.0. Use SparkSession.builder.getOrCreate() instead.
  warnings.warn(
```

```
+-------------------------------------------------------------------------+---------
--------------------------+---------+---------+--------+
|antecedent                                                               |consequen
t                         |n_country|mean_lift|min_lift|
+-------------------------------------------------------------------------+---------
--------------------------+---------+---------+--------+
|[REGENCY CAKESTAND 3 TIER]                                               |[POSTAGE]
|4        |1.24     |1.021   |
|[RABBIT NIGHT LIGHT]                                                     |[POSTAGE]
|3        |1.271    |1.021   |
|[RED TOADSTOOL LED NIGHT LIGHT]                                          |[POSTAGE]
|3        |1.269    |1.021   |
|[ROUND SNACK BOXES SET OF4 WOODLAND]                                     |[POSTAGE]
|3        |1.119    |1.021   |
|[PLASTERS IN TIN SPACEBOY]                                               |[POSTAGE]
|3        |1.09     |1.021   |
|[PLASTERS IN TIN WOODLAND ANIMALS]                                       |[POSTAGE]
|3        |1.038    |1.021   |
|[ROUND SNACK BOXES SET OF 4 FRUITS, POSTAGE]                             |[ROUND SN
ACK BOXES SET OF4 WOODLAND] |2        |2.791    |2.211   |
|[ROUND SNACK BOXES SET OF 4 FRUITS]                                      |[ROUND SN
ACK BOXES SET OF4 WOODLAND] |2        |2.755    |2.11    |
|[PLASTERS IN TIN CIRCUS PARADE]                                          |[POSTAGE]
|2        |1.1      |1.061   |
|[ROUND SNACK BOXES SET OF 4 FRUITS, ROUND SNACK BOXES SET OF4 WOODLAND]|[POSTAGE]
|2        |1.091    |1.021   |
|[LUNCH BAG WOODLAND]                                                     |[POSTAGE]
|2        |1.074    |1.021   |
|[ROUND SNACK BOXES SET OF 4 FRUITS]                                      |[POSTAGE]
|2        |1.073    |0.974   |
|[WOODLAND CHARLOTTE BAG]                                                 |[POSTAGE]
|2        |1.069    |1.021   |
|[LUNCH BOX WITH CUTLERY RETROSPOT]                                       |[POSTAGE]
|2        |1.031    |1.021   |
|[LUNCH BAG RED RETROSPOT]                                                |[POSTAGE]
|2        |1.029    |1.021   |
|[LUNCH BAG CARS BLUE]                                                    |[LUNCH BA
G WOODLAND]                |1        |9.5      |9.5     |
|[LUNCH BAG WOODLAND]                                                     |[LUNCH BA
G CARS BLUE]                |1        |9.5      |9.5     |
|[SET OF 20 KIDS COOKIE CUTTERS, RETROSPOT TEA SET CERAMIC 11 PC]         |[GINGERBR
EAD MAN COOKIE CUTTER]      |1        |9.5      |9.5     |
|[PACK OF 72 RETROSPOT CAKE CASES]                                        |[PACK OF
60 SPACEBOY CAKE CASES]     |1        |9.0      |9.0     |
|[PACK OF 60 SPACEBOY CAKE CASES]                                         |[PACK OF
72 RETROSPOT CAKE CASES]    |1        |9.0      |9.0     |
|[DOORMAT UNION FLAG, DOORMAT WELCOME TO OUR HOME]                        |[DOORMAT
AIRMAIL]                    |1        |7.6      |7.6     |
|[SET OF TEA COFFEE SUGAR TINS PANTRY]                                    |[VINTAGE
CREAM DOG FOOD CONTAINER]   |1        |7.6      |7.6     |
|[TOY TIDY SPACEBOY, RECYCLING BAG RETROSPOT]                             |[TOY TIDY
PINK POLKADOT]              |1        |7.6      |7.6     |
|[TOY TIDY SPACEBOY, CHILDRENS APRON APPLES DESIGN]                       |[TOY TIDY
PINK POLKADOT]              |1        |7.6      |7.6     |
|[SET OF 20 KIDS COOKIE CUTTERS]                                          |[GINGERBR
EAD MAN COOKIE CUTTER]      |1        |7.6      |7.6     |
|[TOY TIDY PINK POLKADOT, RECYCLING BAG RETROSPOT]                        |[TOY TIDY
SPACEBOY]                   |1        |7.6      |7.6     |
|[VINTAGE CREAM DOG FOOD CONTAINER]                                       |[SET OF T
EA COFFEE SUGAR TINS PANTRY]|1        |7.6      |7.6     |
|[TOY TIDY PINK POLKADOT, CHILDRENS APRON APPLES DESIGN]                  |[TOY TIDY
SPACEBOY]                   |1        |7.6      |7.6     |
|[GINGERBREAD MAN COOKIE CUTTER, RETROSPOT TEA SET CERAMIC 11 PC]         |[SET OF 2
0 KIDS COOKIE CUTTERS]      |1        |7.6      |7.6     |
```

```
|[TOY TIDY PINK POLKADOT]                                                    |[TOY TIDY
SPACEBOY]                  |1        |7.6      |7.6     |
+--------------------------------------------------------------------------+---------
---------------------------+---------+---------+--------+
only showing top 30 rows
```

In [42]:

```
from pyspark.ml.fpm import FPGrowth
# from time import time
minSupport=0.01
minConfidence=0.4

fpGrowth = FPGrowth(itemsCol="Basket", minSupport=minSupport, minConfidence=minCon
model = fpGrowth.fit(df_group)
print('fitted')
results = model.associationRules
results.show()
```

```
fitted
+--------------------+--------------------+--------------------+------------------+
--------------------+
|          antecedent|          consequent|          confidence|              lift|
support|
+--------------------+--------------------+--------------------+------------------+
--------------------+
|[LUNCH BAG WOODLA...|[LUNCH BAG  BLACK...| 0.5149253731343422| 9.001529886659087|
0.011396795683532576|
|[LUNCH BAG WOODLA...|[LUNCH BAG RED RE...| 0.6343283582089657| 9.100557638348693|
0.01403953091449655|
|[LUNCH BAG WOODLA...|[LUNCH BAG CARS B...| 0.5174129353233968| 9.913260700716547|
0.011451852667510993|
|[LUNCH BAG WOODLA...|[LUNCH BAG PINK P...| 0.4925373134328361|   9.7876971814888|
0.010901282827726832|
|[LUNCH BAG WOODLA...|[LUNCH BAG DOLLY ...| 0.4601990049751255|14.239513675235443|
0.010185542036007423|
|[PINK REGENCY TEA...|[ROSES REGENCY TE...| 0.8402777777777823|20.792868225854804|
0.019985685184165932|
|[PINK REGENCY TEA...|[REGENCY CAKESTAN...| 0.5787037037037157| 6.460353638826239|
0.013764245994604468|
|[JUMBO STORAGE BA...|[JUMBO BAG RED RE...| 0.5611979166666792| 6.479998576234516|
0.02372956009469823|
|[LUNCH BAG CARS B...|[LUNCH BAG  BLACK...|  0.439873417721519|7.6895292455014035|
0.022958762319000403|
|[LUNCH BAG CARS B...|[LUNCH BAG SPACEB...|0.40717299578059074| 7.561843683397845|
0.021251995815669503|
|[LUNCH BAG CARS B...|[LUNCH BAG RED RE...| 0.4736286919831224| 6.795037861366304|
0.024720585806309717|
|[LUNCH BAG CARS B...|[LUNCH BAG PINK P...| 0.4440928270042194| 8.825008771200666|
0.023178990254914066|
|[LUNCH BAG CARS B...|[LUNCH BAG SUKI D...|0.40717299578059074| 8.290900361393808|
0.021251995815669503|
|   [JUMBO BAG PEARS]|[JUMBO BAG RED RE...|  0.437229437229439| 5.048568511378449|
0.011121510763640497|
|   [JUMBO BAG PEARS]|  [JUMBO BAG APPLES]| 0.6666666666666762| 17.17541371158417|
0.016957551065353047|
|   [JUMBO BAG PEARS]|[JUMBO BAG VINTAG...| 0.5129870129870269|12.406635308765784|
0.01304850520288506|
|[CANDLEHOLDER PIN...|[WHITE HANGING HE...| 0.7347560975609833| 6.954338196977467|
0.013268733138798724|
|[JUMBO  BAG BAROQ...|[JUMBO BAG RED RE...| 0.5569422776911077| 6.430859879023488|
0.01965534328029544|
|[REGENCY TEA PLAT...|[REGENCY TEA PLAT...| 0.6877076411960222| 50.77574750830487|
0.011396795683532576|
|[60 CAKE CASES DO...|[PACK OF 72 RETRO...| 0.5430267062314671| 9.942534340002608|
0.01007542806805059|
+--------------------+--------------------+--------------------+------------------+
--------------------+
only showing top 20 rows
```

In [43]:

```
for column in ['confidence', 'lift', 'support']:
    results = results.withColumn(column, round(results[column], 3))

results.sort(col('lift').desc()).show(50, truncate=False)
```

```
+-----------------------------------------------------------------------------
-----------+-------------------------------------+-----------------+---------
----+--------------------+
|antecedent
|consequent                           |confidence        |lift             |suppor
t             |
+-----------------------------------------------------------------------------
-----------+-------------------------------------+-----------------+---------
----+--------------------+
|[REGENCY SUGAR BOWL GREEN]
|[REGENCY MILK JUG PINK]              |0.763             |56.601           |0.01
|
|[REGENCY MILK JUG PINK]
|[REGENCY SUGAR BOWL GREEN]           |0.751             |56.601           |0.01
|
|[REGENCY TEA PLATE ROSES]
|[REGENCY TEA PLATE GREEN]            |0.688             |50.776           |0.011
|
|[REGENCY TEA PLATE GREEN]
|[REGENCY TEA PLATE ROSES]            |0.841             |50.776           |0.011
|
|[POPPY'S PLAYHOUSE BEDROOM]
|[POPPY'S PLAYHOUSE LIVINGROOM]       |0.645             |48.009           |0.011
|
|[POPPY'S PLAYHOUSE LIVINGROOM]
|[POPPY'S PLAYHOUSE BEDROOM]          |0.811             |48.009           |0.011
|
|[SET/6 RED SPOTTY PAPER CUPS]
|[SET/6 RED SPOTTY PAPER PLATES]      |0.824             |47.347           |0.013
|
|[SET/6 RED SPOTTY PAPER PLATES]
|[SET/6 RED SPOTTY PAPER CUPS]        |0.725             |47.347           |0.013
|
|[POPPY'S PLAYHOUSE LIVINGROOM]
|[POPPY'S PLAYHOUSE KITCHEN]          |0.852             |46.081           |0.011
|
|[POPPY'S PLAYHOUSE KITCHEN]
|[POPPY'S PLAYHOUSE LIVINGROOM]       |0.619             |46.081           |0.011
|
|[CHILDRENS CUTLERY DOLLY GIRL]
|[CHILDRENS CUTLERY SPACEBOY]         |0.776             |45.3             |0.011
|
|[CHILDRENS CUTLERY SPACEBOY]
|[CHILDRENS CUTLERY DOLLY GIRL]       |0.656             |45.3             |0.011
|
|[POPPY'S PLAYHOUSE BEDROOM]
|[POPPY'S PLAYHOUSE KITCHEN]          |0.801             |43.316           |0.014
|
|[POPPY'S PLAYHOUSE KITCHEN]
|[POPPY'S PLAYHOUSE BEDROOM]          |0.732             |43.316           |0.014
|
|[SCANDINAVIAN PAISLEY PICNIC BAG]
|[PINK VINTAGE PAISLEY PICNIC BAG]    |0.628             |41.019           |0.011
|
|[PINK VINTAGE PAISLEY PICNIC BAG]
|[SCANDINAVIAN PAISLEY PICNIC BAG]    |0.698             |41.019           |0.011
|
|[SMALL MARSHMALLOWS PINK BOWL]
|[SMALL DOLLY MIX DESIGN ORANGE BOWL] |0.779             |39.742           |0.012
|
|[SMALL DOLLY MIX DESIGN ORANGE BOWL]
|[SMALL MARSHMALLOWS PINK BOWL]       |0.624             |39.742           |0.012
|
|[FELTCRAFT CUSHION BUTTERFLY]
```

```
|[FELTCRAFT CUSHION RABBIT]           |0.683             |39.263           |0.011
|
|[FELTCRAFT CUSHION RABBIT]
|[FELTCRAFT CUSHION BUTTERFLY]        |0.614             |39.263           |0.011
|
|[PINK HAPPY BIRTHDAY BUNTING]
|[BLUE HAPPY BIRTHDAY BUNTING]        |0.703             |37.344           |0.013
|
|[BLUE HAPPY BIRTHDAY BUNTING]
|[PINK HAPPY BIRTHDAY BUNTING]        |0.713             |37.344           |0.013
|
|[WOODEN STAR CHRISTMAS SCANDINAVIAN]
|[WOODEN TREE CHRISTMAS SCANDINAVIAN] |0.519             |35.723           |0.012
|
|[WOODEN TREE CHRISTMAS SCANDINAVIAN]
|[WOODEN STAR CHRISTMAS SCANDINAVIAN] |0.818             |35.723           |0.012
|
|[SET OF 6 SNACK LOAF BAKING CASES]
|[SET OF 12 MINI LOAF BAKING CASES]   |0.674             |34.869           |0.01
|
|[SET OF 12 MINI LOAF BAKING CASES]
|[SET OF 6 SNACK LOAF BAKING CASES]   |0.536             |34.869           |0.01
|
|[RED STRIPE CERAMIC DRAWER KNOB]
|[BLUE STRIPE CERAMIC DRAWER KNOB]    |0.645             |33.939           |0.011
|
|[BLUE STRIPE CERAMIC DRAWER KNOB]
|[RED STRIPE CERAMIC DRAWER KNOB]     |0.594             |33.939           |0.011
|
|[BATHROOM METAL SIGN]
|[TOILET METAL SIGN]                  |0.54              |32.922           |0.012
|
|[TOILET METAL SIGN]
|[BATHROOM METAL SIGN]                |0.745             |32.922           |0.012
|
|[SET OF 6 TEA TIME BAKING CASES]
|[SET OF 12 MINI LOAF BAKING CASES]   |0.602             |31.148           |0.01
|
|[SET OF 12 MINI LOAF BAKING CASES]
|[SET OF 6 TEA TIME BAKING CASES]     |0.53              |31.148           |0.01
|
|[WOODEN STAR CHRISTMAS SCANDINAVIAN]
|[WOODEN HEART CHRISTMAS SCANDINAVIAN]|0.743             |30.182           |0.017
|
|[WOODEN HEART CHRISTMAS SCANDINAVIAN]
|[WOODEN STAR CHRISTMAS SCANDINAVIAN] |0.691             |30.182           |0.017
|
|[WOODEN TREE CHRISTMAS SCANDINAVIAN]
|[WOODEN HEART CHRISTMAS SCANDINAVIAN]|0.723             |29.397           |0.011
|
|[WOODEN HEART CHRISTMAS SCANDINAVIAN]
|[WOODEN TREE CHRISTMAS SCANDINAVIAN] |0.427             |29.397           |0.011
|
|[FELTCRAFT PRINCESS CHARLOTTE DOLL]
|[FELTCRAFT PRINCESS LOLA DOLL]       |0.59              |29.21            |0.013
|
|[FELTCRAFT PRINCESS LOLA DOLL]
|[FELTCRAFT PRINCESS CHARLOTTE DOLL]  |0.624             |29.21            |0.013
|
|[SET OF 6 SNACK LOAF BAKING CASES]
|[SET OF 12 FAIRY CAKE BAKING CASES]  |0.688             |27.715           |0.011
|
|[SET OF 12 FAIRY CAKE BAKING CASES]
|[SET OF 6 SNACK LOAF BAKING CASES]   |0.426             |27.715           |0.011
```

```
|
|[CHRISTMAS CRAFT WHITE FAIRY]
|[CHRISTMAS CRAFT LITTLE FRIENDS]     |0.632             |27.664           |0.011
|
|[CHRISTMAS CRAFT LITTLE FRIENDS]
|[CHRISTMAS CRAFT WHITE FAIRY]        |0.484             |27.664           |0.011
|
|[SET OF 12 MINI LOAF BAKING CASES]
|[SET OF 12 FAIRY CAKE BAKING CASES]  |0.687             |27.652           |0.013
|
|[SET OF 12 FAIRY CAKE BAKING CASES]
|[SET OF 12 MINI LOAF BAKING CASES]   |0.534             |27.652           |0.013
|
|[FELTCRAFT CUSHION OWL]
|[FELTCRAFT CUSHION RABBIT]           |0.48              |27.612           |0.011
|
|[FELTCRAFT CUSHION RABBIT]
|[FELTCRAFT CUSHION OWL]              |0.62              |27.612           |0.011
|
|[SET OF 6 TEA TIME BAKING CASES]
|[SET OF 12 FAIRY CAKE BAKING CASES]  |0.673             |27.109           |0.011
|
|[SET OF 12 FAIRY CAKE BAKING CASES]
|[SET OF 6 TEA TIME BAKING CASES]     |0.461             |27.109           |0.011
|
|[GREEN REGENCY TEACUP AND SAUCER, ROSES REGENCY TEACUP AND SAUCER, REGENCY CAKEST
AND 3 TIER]|[PINK REGENCY TEACUP AND SAUCER]     |0.7620000000000068|26.3200000000
0016|0.012000000000000103|
|[CHRISTMAS CRAFT TREE TOP ANGEL]
|[CHRISTMAS CRAFT LITTLE FRIENDS]     |0.577             |25.27            |0.011
|
+-------------------------------------------------------------------------------
-----------+-------------------------------------+------------------+-----------
----+--------------------+
only showing top 50 rows
```

Antecedent: Represents the antecedent or the left-hand side of the association rule. It is a set or array of items that are present in the dataset and act as the condition or premise of the rule.

Consequent: Represents the consequent or the right-hand side of the association rule. It is a set or array of items that are predicted or inferred based on the presence of the antecedent.

Confidence: Indicates the strength of the association rule. It is a measure of how often the consequent appears in transactions that contain the antecedent. Confidence is calculated as the ratio of the support of the rule (support of both antecedent and consequent) to the support of the antecedent. Higher confidence values indicate stronger associations between the antecedent and consequent.

Lift: Lift is a measure of how much more likely the consequent is to appear in transactions that contain the antecedent compared to its individual occurrence. It is calculated as the ratio of the confidence of the rule to the support of the consequent. Lift values greater than 1 indicate a positive correlation between the antecedent and consequent, suggesting that the presence of the antecedent increases the likelihood of the consequent.

Support: Indicates the fraction of transactions in the dataset that contain both the antecedent and the consequent. It is calculated as the ratio of the number of transactions

containing both to the total number of transactions.

In [44]:
```
frequent_itemsets = model.freqItemsets
frequent_itemsets.show(100,truncate=False)
```

```
+--------------------------------------------------------------------------------
---+----+
|items
|freq|
+--------------------------------------------------------------------------------
---+----+
|[SET OF 36 DOILIES PANTRY DESIGN]
|185 |
|[RED RETROSPOT SUGAR JAM BOWL]
|193 |
|[PACK OF 20 SKULL PAPER NAPKINS]
|203 |
|[METAL SIGN EMPIRE TEA]
|183 |
|[DECORATIVE WICKER HEART LARGE]
|212 |
|[PINK  POLKADOT CUP]
|200 |
|[PICNIC BASKET WICKER SMALL]
|226 |
|[ENCHANTED BIRD COATHANGER 5 HOOK]
|189 |
|[BOX OF VINTAGE ALPHABET BLOCKS]
|221 |
|[PINK HEART SHAPE EGG FRYING PAN]
|197 |
|[SET OF 9 HEART SHAPED BALLOONS]
|207 |
|[FELTCRAFT PRINCESS OLIVIA DOLL]
|244 |
|[POTTING SHED TEA MUG]
|231 |
|[SET OF 3 WOODEN SLEIGH DECORATIONS]
|239 |
|[CALENDAR PAPER CUT DESIGN]
|229 |
|[RED METAL BOX TOP SECRET]
|214 |
|[CHILDREN'S APRON DOLLY GIRL]
|250 |
|[SET 6 SCHOOL MILK BOTTLES IN CRATE]
|271 |
|[PICNIC BASKET WICKER LARGE]
|246 |
|[GIANT 50'S CHRISTMAS CRACKER]
|299 |
|[MINT KITCHEN SCALES]
|281 |
|[BINGO SET]
|265 |
|[BATH BUILDING BLOCK WORD]
|235 |
|[WOODEN UNION JACK BUNTING]
|296 |
|[LONDON BUS COFFEE MUG]
|276 |
|[MEMO BOARD COTTAGE DESIGN]
|255 |
|[CLEAR DRAWER KNOB ACRYLIC EDWARDIAN]
|312 |
|[METAL 4 HOOK HANGER FRENCH CHATEAU]
|332 |
|[PINK FAIRY CAKE CHILDRENS APRON]
|386 |
```

```
|[SET OF 12 MINI LOAF BAKING CASES]
|351 |
|[SET OF 12 MINI LOAF BAKING CASES, SET OF 12 FAIRY CAKE BAKING CASES]
|241 |
|[JINGLE BELL HEART DECORATION]
|306 |
|[12 PENCILS SMALL TUBE SKULL]
|289 |
|[SOLDIERS EGG CUP]
|327 |
|[RECIPE BOX RETROSPOT]
|428 |
|[RECIPE BOX RETROSPOT, RECIPE BOX PANTRY YELLOW DESIGN]
|201 |
|[PAINTED METAL PEARS ASSORTED]
|342 |
|[PAINTED METAL PEARS ASSORTED, ASSORTED COLOUR BIRD ORNAMENT]
|248 |
|[RED STRIPE CERAMIC DRAWER KNOB]
|318 |
|[RED STRIPE CERAMIC DRAWER KNOB, BLUE STRIPE CERAMIC DRAWER KNOB]
|205 |
|[SET OF 3 BUTTERFLY COOKIE CUTTERS]
|485 |
|[SET OF 3 BUTTERFLY COOKIE CUTTERS, SET OF 3 HEART COOKIE CUTTERS]
|238 |
|[PAPER CHAIN KIT EMPIRE]
|372 |
|[JUMBO  BAG BAROQUE BLACK WHITE]
|641 |
|[JUMBO  BAG BAROQUE BLACK WHITE, JUMBO BAG RED RETROSPOT]
|357 |
|[JUMBO  BAG BAROQUE BLACK WHITE, JUMBO SHOPPER VINTAGE RED PAISLEY]
|236 |
|[JUMBO  BAG BAROQUE BLACK WHITE, JUMBO BAG PINK POLKADOT]
|248 |
|[JUMBO  BAG BAROQUE BLACK WHITE, JUMBO STORAGE BAG SUKI]
|240 |
|[JUMBO  BAG BAROQUE BLACK WHITE, JUMBO BAG STRAWBERRY]
|202 |
|[WOODLAND CHARLOTTE BAG]
|524 |
|[WOODLAND CHARLOTTE BAG, RED RETROSPOT CHARLOTTE BAG]
|286 |
|[WOODLAND CHARLOTTE BAG, CHARLOTTE BAG SUKI DESIGN]
|239 |
|[SET OF 20 KIDS COOKIE CUTTERS]
|405 |
|[DOORMAT HEARTS]
|465 |
|[POTTERING IN THE SHED METAL SIGN]
|367 |
|[HAND OVER THE CHOCOLATE   SIGN]
|589 |
|[HAND OVER THE CHOCOLATE   SIGN, GIN + TONIC DIET METAL SIGN]
|248 |
|[HAND OVER THE CHOCOLATE   SIGN, PLEASE ONE PERSON METAL SIGN]
|234 |
|[JUMBO BAG SCANDINAVIAN BLUE PAISLEY]
|445 |
|[JUMBO BAG SCANDINAVIAN BLUE PAISLEY, JUMBO BAG PINK VINTAGE PAISLEY]
|236 |
|[JUMBO BAG SCANDINAVIAN BLUE PAISLEY, JUMBO BAG RED RETROSPOT]
|263 |
```

```
|[JUMBO BAG SCANDINAVIAN BLUE PAISLEY, JUMBO SHOPPER VINTAGE RED PAISLEY]
|197 |
|[JUMBO BAG SCANDINAVIAN BLUE PAISLEY, JUMBO BAG PINK POLKADOT]
|187 |
|[JUMBO STORAGE BAG SUKI]
|768 |
|[JUMBO STORAGE BAG SUKI, LUNCH BAG SUKI DESIGN]
|221 |
|[JUMBO STORAGE BAG SUKI, LUNCH BAG  BLACK SKULL.]
|183 |
|[JUMBO STORAGE BAG SUKI, JUMBO BAG RED RETROSPOT]
|431 |
|[JUMBO STORAGE BAG SUKI, LUNCH BAG RED RETROSPOT]
|237 |
|[JUMBO STORAGE BAG SUKI, JUMBO SHOPPER VINTAGE RED PAISLEY]
|249 |
|[JUMBO STORAGE BAG SUKI, JUMBO SHOPPER VINTAGE RED PAISLEY, JUMBO BAG RED RETROSP
OT]|185 |
|[JUMBO STORAGE BAG SUKI, JUMBO BAG PINK POLKADOT]
|280 |
|[JUMBO STORAGE BAG SUKI, JUMBO BAG PINK POLKADOT, JUMBO BAG RED RETROSPOT]
|219 |
|[HOT WATER BOTTLE TEA AND SYMPATHY]
|504 |
|[HOT WATER BOTTLE TEA AND SYMPATHY, CHOCOLATE HOT WATER BOTTLE]
|256 |
|[HOT WATER BOTTLE TEA AND SYMPATHY, HOT WATER BOTTLE KEEP CALM]
|228 |
|[SPACEBOY LUNCH BOX]
|688 |
|[SPACEBOY LUNCH BOX, LUNCH BAG SPACEBOY DESIGN]
|198 |
|[PAPER CHAIN KIT 50'S CHRISTMAS]
|963 |
|[HAND WARMER OWL DESIGN]
|568 |
|[JAM MAKING SET WITH JARS]
|839 |
|[JAM MAKING SET WITH JARS, SET OF 3 CAKE TINS PANTRY DESIGN]
|233 |
|[WHITE HANGING HEART T-LIGHT HOLDER]
|1919|
|[CLASSIC GLASS COOKIE JAR]
|239 |
|[FANCY FONTS BIRTHDAY WRAP]
|228 |
|[DELUXE SEWING KIT]
|214 |
|[3 TRADITIONAl BISCUIT CUTTERS  SET]
|193 |
|[GAOLERS KEYS DECORATIVE GARDEN]
|203 |
|[DOORMAT WELCOME PUPPIES]
|235 |
|[HEART FILIGREE DOVE LARGE]
|212 |
|[REGENCY TEA PLATE GREEN]
|246 |
|[REGENCY TEA PLATE GREEN, REGENCY TEA PLATE ROSES]
|207 |
|[KEY FOB , SHED]
|264 |
|[GIRAFFE WOODEN RULER]
|225 |
```

```
|[RED SPOT CERAMIC DRAWER KNOB]
|276 |
|[POPPY'S PLAYHOUSE LIVINGROOM]
|244 |
|[POPPY'S PLAYHOUSE LIVINGROOM, POPPY'S PLAYHOUSE KITCHEN]
|208 |
|[POPPY'S PLAYHOUSE LIVINGROOM, POPPY'S PLAYHOUSE BEDROOM]
|198 |
|[VINTAGE DOILY JUMBO BAG RED]
|231 |
|[LUNCH BAG PAISLEY PARK]
|295 |
|[RED RETROSPOT OVEN GLOVE]
|255 |
+----------------------------------------------------------------------------------
---+----+
only showing top 100 rows
```

In [50]:
```
fre_df = frequent_itemsets.filter(size('items') == 3)
fre_df.show(truncate=False)
```

```
+----------------------------------------------------------------------------
---+----+
|items
|freq|
+----------------------------------------------------------------------------
---+----+
|[JUMBO STORAGE BAG SUKI, JUMBO SHOPPER VINTAGE RED PAISLEY, JUMBO BAG RED RETROSP
OT]|185 |
|[JUMBO STORAGE BAG SUKI, JUMBO BAG PINK POLKADOT, JUMBO BAG RED RETROSPOT]
|219 |
|[ALARM CLOCK BAKELIKE IVORY, ALARM CLOCK BAKELIKE GREEN, ALARM CLOCK BAKELIKE RE
D]  |212 |
|[LUNCH BAG CARS BLUE, LUNCH BAG  BLACK SKULL., LUNCH BAG RED RETROSPOT]
|258 |
|[LUNCH BAG CARS BLUE, LUNCH BAG SPACEBOY DESIGN, LUNCH BAG  BLACK SKULL.]
|220 |
|[LUNCH BAG CARS BLUE, LUNCH BAG SPACEBOY DESIGN, LUNCH BAG RED RETROSPOT]
|230 |
|[LUNCH BAG WOODLAND, LUNCH BAG  BLACK SKULL., LUNCH BAG RED RETROSPOT]
|226 |
|[LUNCH BAG WOODLAND, LUNCH BAG SUKI DESIGN, LUNCH BAG  BLACK SKULL.]
|184 |
|[LUNCH BAG WOODLAND, LUNCH BAG SUKI DESIGN, LUNCH BAG RED RETROSPOT]
|220 |
|[LUNCH BAG WOODLAND, LUNCH BAG SPACEBOY DESIGN, LUNCH BAG  BLACK SKULL.]
|207 |
|[LUNCH BAG WOODLAND, LUNCH BAG SPACEBOY DESIGN, LUNCH BAG RED RETROSPOT]
|255 |
|[LUNCH BAG WOODLAND, LUNCH BAG CARS BLUE, LUNCH BAG  BLACK SKULL.]
|192 |
|[LUNCH BAG WOODLAND, LUNCH BAG CARS BLUE, LUNCH BAG SPACEBOY DESIGN]
|208 |
|[LUNCH BAG WOODLAND, LUNCH BAG CARS BLUE, LUNCH BAG RED RETROSPOT]
|225 |
|[LUNCH BAG WOODLAND, LUNCH BAG APPLE DESIGN, LUNCH BAG RED RETROSPOT]
|190 |
|[LUNCH BAG WOODLAND, LUNCH BAG PINK POLKADOT, LUNCH BAG  BLACK SKULL.]
|195 |
|[LUNCH BAG WOODLAND, LUNCH BAG PINK POLKADOT, LUNCH BAG SPACEBOY DESIGN]
|198 |
|[LUNCH BAG WOODLAND, LUNCH BAG PINK POLKADOT, LUNCH BAG RED RETROSPOT]
|234 |
|[LUNCH BAG WOODLAND, LUNCH BAG PINK POLKADOT, LUNCH BAG CARS BLUE]
|193 |
|[LUNCH BAG DOLLY GIRL DESIGN, LUNCH BAG SPACEBOY DESIGN, LUNCH BAG  BLACK SKULL.]
|182 |
+----------------------------------------------------------------------------
---+----+
only showing top 20 rows
```

In [ ]: