[26]:
```
# Purchase Pattern Analytics - Market Basket Analysis (Apriori)
```

[28]:
```
import os
os.listdir()
```

[28]:
```
['.ipynb_checkpoints', 'clientproject1.csv', 'Purchase_Pattern_Apriori.ipynb']
```

[8]:
```
df = pd.read_csv("clientproject1.csv", encoding="latin1")
df.head()
```

```
C:\Users\priya\AppData\Local\Temp\ipykernel_33308\2040283746.py:1: DtypeWarning: Columns (0,4) have mixed types. Specify dtype option on import or set low_memory=False.
  df = pd.read_csv("clientproject1.csv", encoding="latin1")
```

[8]:

	BillNo	Itemname	Quantity	Present_Date	Price	CustomerID	Country
0	536365	WHITE HANGING HEART T-LIGHT HOLDER	6.0	01-12-2010 08:26	2.55	17850.0	United Kingdom
1	536365	WHITE METAL LANTERN	6.0	01-12-2010 08:26	3.39	17850.0	United Kingdom
2	536365	CREAM CUPID HEARTS COAT HANGER	8.0	01-12-2010 08:26	2.75	17850.0	United Kingdom
3	536365	KNITTED UNION FLAG HOT WATER BOTTLE	6.0	01-12-2010 08:26	3.39	17850.0	United Kingdom
4	536365	RED WOOLLY HOTTIE WHITE HEART.	6.0	01-12-2010 08:26	3.39	17850.0	United Kingdom

[7]:
```
df.columns
```

[7]:
```
Index(['BillNo', 'Itemname', 'Quantity', 'Present_Date', 'Price', 'CustomerID',
       'Country'],
      dtype='object')
```

[37]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 522064 entries, 0 to 522063
Data columns (total 7 columns):
 #   Column        Non-Null Count   Dtype  
---  ------        --------------   -----  
 0   BillNo        522064 non-null  object 
 1   Itemname      520609 non-null  object 
 2   Quantity      522064 non-null  float64
 3   Present_Date  522063 non-null  object 
 4   Price         522064 non-null  object 
 5   CustomerID    388023 non-null  float64
 6   Country       522063 non-null  object 
dtypes: float64(2), object(5)
memory usage: 27.9+ MB
```

[39]:
```
df = df.rename(columns={
    'BillNo': 'InvoiceNo',
    'Itemname': 'Description',
    'Quantity': 'Quantity'
})

df.head()
```

[39]:

	InvoiceNo	Description	Quantity	Present_Date	Price	CustomerID	Country
0	536365	WHITE HANGING HEART T-LIGHT HOLDER	6.0	01-12-2010 08:26	2.55	17850.0	United Kingdom
1	536365	WHITE METAL LANTERN	6.0	01-12-2010 08:26	3.39	17850.0	United Kingdom
2	536365	CREAM CUPID HEARTS COAT HANGER	8.0	01-12-2010 08:26	2.75	17850.0	United Kingdom
3	536365	KNITTED UNION FLAG HOT WATER BOTTLE	6.0	01-12-2010 08:26	3.39	17850.0	United Kingdom
4	536365	RED WOOLLY HOTTIE WHITE HEART.	6.0	01-12-2010 08:26	3.39	17850.0	United Kingdom

[40]:
```
# Remove missing Invoice or Product names
df = df.dropna(subset=['InvoiceNo', 'Description'])

# Convert InvoiceNo to string
df['InvoiceNo'] = df['InvoiceNo'].astype(str)

# Remove cancelled invoices (if any)
df = df[~df['InvoiceNo'].str.startswith('C')]

# Keep only positive quantity
df['Quantity'] = pd.to_numeric(df['Quantity'], errors='coerce')
df = df[df['Quantity'] > 0]

df.head()
```

[40]:

	InvoiceNo	Description	Quantity	Present_Date	Price	CustomerID	Country
0	536365	WHITE HANGING HEART T-LIGHT HOLDER	6.0	01-12-2010 08:26	2.55	17850.0	United Kingdom
1	536365	WHITE METAL LANTERN	6.0	01-12-2010 08:26	3.39	17850.0	United Kingdom
2	536365	CREAM CUPID HEARTS COAT HANGER	8.0	01-12-2010 08:26	2.75	17850.0	United Kingdom
3	536365	KNITTED UNION FLAG HOT WATER BOTTLE	6.0	01-12-2010 08:26	3.39	17850.0	United Kingdom
4	536365	RED WOOLLY HOTTIE WHITE HEART.	6.0	01-12-2010 08:26	3.39	17850.0	United Kingdom

[44]:
```
transactions = (
    df.groupby('InvoiceNo')['Description']
      .apply(list)
      .tolist()
)

transactions[:5]
```

[44]:
```
[['WHITE HANGING HEART T-LIGHT HOLDER',
  'WHITE METAL LANTERN',
  'CREAM CUPID HEARTS COAT HANGER',
  'KNITTED UNION FLAG HOT WATER BOTTLE',
  'RED WOOLLY HOTTIE WHITE HEART.',
  'SET 7 BABUSHKA NESTING BOXES',
  'GLASS STAR FROSTED T-LIGHT HOLDER'],
 ['HAND WARMER UNION JACK', 'HAND WARMER RED POLKA DOT'],
 ['ASSORTED COLOUR BIRD ORNAMENT',
  "POPPY'S PLAYHOUSE BEDROOM",
  "POPPY'S PLAYHOUSE KITCHEN",
  'FELTCRAFT PRINCESS CHARLOTTE DOLL',
  'IVORY KNITTED MUG COSY',
  'BOX OF 6 ASSORTED COLOUR TEASPOONS',
  'BOX OF VINTAGE JIGSAW BLOCKS',
  'BOX OF VINTAGE ALPHABET BLOCKS',
  'HOME BUILDING BLOCK WORD',
  'LOVE BUILDING BLOCK WORD',
  'RECIPE BOX WITH METAL HEART',
  'DOORMAT NEW ENGLAND'],
 ['JAM MAKING SET WITH JARS',
  'RED COAT RACK PARIS FASHION',
  'YELLOW COAT RACK PARIS FASHION',
  'BLUE COAT RACK PARIS FASHION'],
 ['BATH BUILDING BLOCK WORD']]
```

[15]:
```
# one-hot encoding
te = TransactionEncoder()
te_array = te.fit(transactions).transform(transactions)
basket = pd.DataFrame(te_array, columns=te.columns_)

basket.head()
```

[15]:

	*Boombox Ipod Classic	*USB Office Mirror Ball	10 COLOUR SPACEBOY PEN	12 COLOURED PARTY BALLOONS	12 DAISY PEGS IN WOOD BOX	12 EGG HOUSE PAINTED WOOD	12 HANGING EGGS HAND PAINTED	12 IVORY ROSE PEG PLACE SETTINGS	12 MESSAGE CARDS WITH ENVELOPES	12 PENCIL SMALL TUBE WOODLAND	...	returned	taig adjust	test	to push order througha s stock was	website fixed	wror co 20
0	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	F
1	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	F
2	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	F
3	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	F
4	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	F

5 rows × 4056 columns

[20]:
```
# Count how many times each product appears
item_freq = basket.sum()

# Keep only products bought at least 100 times
basket_reduced = basket[item_freq[item_freq >= 100].index]

basket_reduced.shape
```

[20]:
```
(19735, 1531)
```

[21]:
```
from mlxtend.frequent_patterns import apriori

frequent_itemsets = apriori(
    basket_reduced,
    min_support=0.02,   # increase support
    use_colnames=True
)

frequent_itemsets.head()
```

[21]:

	support	itemsets
0	0.023157	(3 STRIPEY MICE FELTCRAFT)
1	0.023512	(4 TRADITIONAL SPINNING TOPS)
2	0.047732	(6 RIBBONS RUSTIC CHARM)
3	0.021130	(60 CAKE CASES DOLLY GIRL DESIGN)
4	0.030403	(60 CAKE CASES VINTAGE CHRISTMAS)

[22]:
```
frequent_itemsets.sort_values(by='support', ascending=False).head(10)
```

[22]:

	support	itemsets
276	0.111578	(WHITE HANGING HEART T-LIGHT HOLDER)
104	0.104586	(JUMBO BAG RED RETROSPOT)
199	0.096478	(REGENCY CAKESTAND 3 TIER)
159	0.083912	(PARTY BUNTING)
129	0.078085	(LUNCH BAG RED RETROSPOT)
15	0.072511	(ASSORTED COLOUR BIRD ORNAMENT)
220	0.068204	(SET OF 3 CAKE TINS PANTRY DESIGN)
151	0.064809	(PACK OF 72 RETROSPOT CAKE CASES)
122	0.063846	(LUNCH BAG BLACK SKULL.)
141	0.062427	(NATURAL SLATE HEART CHALKBOARD)

[23]:
```
from mlxtend.frequent_patterns import association_rules

rules = association_rules(
    frequent_itemsets,
    metric='lift',
    min_threshold=1
)

rules[['antecedents','consequents','support','confidence','lift']].head()
```

[23]:

	antecedents	consequents	support	confidence	lift
0	(PACK OF 72 RETROSPOT CAKE CASES)	(60 TEATIME FAIRY CAKE CASES)	0.022042	0.340109	8.411103
1	(60 TEATIME FAIRY CAKE CASES)	(PACK OF 72 RETROSPOT CAKE CASES)	0.022042	0.545113	8.411103
2	(ALARM CLOCK BAKELIKE PINK)	(ALARM CLOCK BAKELIKE GREEN)	0.020826	0.541502	11.051232
3	(ALARM CLOCK BAKELIKE GREEN)	(ALARM CLOCK BAKELIKE PINK)	0.020826	0.425026	11.051232
4	(ALARM CLOCK BAKELIKE RED)	(ALARM CLOCK BAKELIKE GREEN)	0.032075	0.616358	12.578936

[24]:
```
strong_rules = rules[
    (rules['confidence'] >= 0.6) &
    (rules['lift'] > 1)
]

strong_rules
```

[24]:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	representativity	leverage	conviction	zhangs_metric	jaccard	certainty
4	(ALARM CLOCK BAKELIKE RED)	(ALARM CLOCK BAKELIKE GREEN)	0.052040	0.048999	0.032075	0.616358	12.578936	1.0	0.029525	2.478878	0.971034	0.465099	0.596592
5	(ALARM CLOCK BAKELIKE GREEN)	(ALARM CLOCK BAKELIKE RED)	0.048999	0.052040	0.032075	0.654602	12.578936	1.0	0.029525	2.744544	0.967930	0.465099	0.635641
10	(CHARLOTTE BAG PINK POLKADOT)	(RED RETROSPOT CHARLOTTE BAG)	0.037395	0.052090	0.026349	0.704607	13.526673	1.0	0.024401	3.208979	0.962048	0.417335	0.688374
18	(DOLLY GIRL LUNCH BOX)	(SPACEBOY LUNCH BOX)	0.042057	0.044794	0.026602	0.632530	14.121020	1.0	0.024719	2.599414	0.969978	0.441548	0.615298
20	(DOTCOM POSTAGE)	(JUMBO BAG RED RETROSPOT)	0.035875	0.104586	0.024576	0.685028	6.549919	1.0	0.020824	2.842840	0.878856	0.212068	0.648239
24	(GARDENERS KNEELING PAD CUP OF TEA)	(GARDENERS KNEELING PAD KEEP	0.038206	0.045858	0.027565	0.721485	15.733165	1.0	0.025813	3.425826	0.973639	0.487892	0.708100

25	(GARDENERS KNEELING PAD KEEP CALM)	(GARDENERS KNEELING PAD CUP OF TEA)	0.045858	0.038206	0.027565	0.601105	15.733165	1.0	0.025813	2.411145	0.981447	0.487892	0.585259
26	(PINK REGENCY TEACUP AND SAUCER)	(GREEN REGENCY TEACUP AND SAUCER)	0.037294	0.049405	0.030656	0.822011	16.638343	1.0	0.028814	5.340750	0.976308	0.547016	0.812760
27	(GREEN REGENCY TEACUP AND SAUCER)	(PINK REGENCY TEACUP AND SAUCER)	0.049405	0.037294	0.030656	0.620513	16.638343	1.0	0.028814	2.536860	0.988746	0.547016	0.605812
30	(ROSES REGENCY TEACUP AND SAUCER)	(GREEN REGENCY TEACUP AND SAUCER)	0.051330	0.049405	0.037041	0.721619	14.606308	1.0	0.034505	3.414727	0.981940	0.581543	0.707151
31	(GREEN REGENCY TEACUP AND SAUCER)	(ROSES REGENCY TEACUP AND SAUCER)	0.049405	0.051330	0.037041	0.749744	14.606308	1.0	0.034505	3.790792	0.979951	0.581543	0.736203
39	(JUMBO BAG BAROQUE BLACK WHITE)	(JUMBO BAG RED RETROSPOT)	0.046871	0.104586	0.029389	0.627027	5.995338	1.0	0.024487	2.400748	0.874177	0.240764	0.583463
47	(JUMBO BAG PEARS)	(JUMBO BAG APPLES)	0.029693	0.048847	0.020319	0.684300	14.008991	1.0	0.018869	3.012841	0.957035	0.348999	0.668087

65	(JUMBO BAG SPACEBOY DESIGN)	(JUMBO BAG RED RETROSPOT)	0.034558	0.104586	0.021586	0.624633	5.972452	1.0	0.017972	2.385439	0.862366	0.183621	0.580790
66	(JUMBO BAG STRAWBERRY)	(JUMBO BAG RED RETROSPOT)	0.041297	0.104586	0.026805	0.649080	6.206196	1.0	0.022486	2.551617	0.875006	0.225106	0.608092
70	(JUMBO BAG WOODLAND ANIMALS)	(JUMBO BAG RED RETROSPOT)	0.043273	0.104586	0.027008	0.624122	5.967560	1.0	0.022482	2.382192	0.870079	0.223480	0.580219
76	(JUMBO STORAGE BAG SUKI)	(JUMBO BAG RED RETROSPOT)	0.059691	0.104586	0.036534	0.612054	5.852176	1.0	0.030291	2.308092	0.881756	0.285998	0.566742
138	(PAPER CHAIN KIT VINTAGE CHRISTMAS)	(PAPER CHAIN KIT 50'S CHRISTMAS)	0.041095	0.057917	0.027616	0.672010	11.602900	1.0	0.025236	2.872289	0.952977	0.386799	0.651846
143	(PINK REGENCY TEACUP AND SAUCER)	(ROSES REGENCY TEACUP AND SAUCER)	0.037294	0.051330	0.028883	0.774457	15.087759	1.0	0.026968	4.206151	0.969892	0.483461	0.762253
144	(RED HANGING HEART T-LIGHT HOLDER)	(WHITE HANGING HEART T-LIGHT HOLDER)	0.036534	0.111578	0.024272	0.664355	5.954154	1.0	0.020195	2.646909	0.863601	0.195990	0.622201
146	(STRAWBERRY CHARLOTTE BAG)	(RED RETROSPOT CHARLOTTE BAG)	0.036281	0.052090	0.024424	0.673184	12.923437	1.0	0.022534	2.900442	0.957355	0.381933	0.655225

160	TEACUP AND SAUCER, PINK REGENCY...	TEACUP AND SAUCER)	0.028883	0.049405	0.026096	0.903509	18.287944	1.0	0.024669	9.851625	0.973435	0.500000	0.898494
161	(ROSES REGENCY TEACUP AND SAUCER, GREEN REGENC...	(PINK REGENCY TEACUP AND SAUCER)	0.037041	0.037294	0.026096	0.704514	18.890749	1.0	0.024714	3.258046	0.983493	0.540966	0.693068
162	(PINK REGENCY TEACUP AND SAUCER, GREEN REGENCY...	(ROSES REGENCY TEACUP AND SAUCER)	0.030656	0.051330	0.026096	0.851240	16.583628	1.0	0.024522	6.377170	0.969418	0.466908	0.843191
164	(PINK REGENCY TEACUP AND SAUCER)	(ROSES REGENCY TEACUP AND SAUCER, GREEN REGENC...	0.037294	0.037041	0.026096	0.699728	18.890749	1.0	0.024714	3.206959	0.983752	0.540966	0.688178
167	(JUMBO STORAGE BAG SUKI, JUMBO BAG PINK POLKADOT)	(JUMBO BAG RED RETROSPOT)	0.026045	0.104586	0.020877	0.801556	7.664107	1.0	0.018153	4.512185	0.892774	0.190212	0.778378

[27]:
```
strong_rules.to_csv("apriori_results.csv", index=False)
```