

# MARKET BASKET INSIGHTS

## PHASE – 5 Documentation & Submission

### Introduction:

Market Basket Analysis is a data mining technique that allows businesses to uncover patterns and associations within their transactional data, particularly in the context of retail and e-commerce. The main objective of Market Basket Analysis is to identify relationships between products that are frequently purchased together. This information can be invaluable for businesses looking to optimize their marketing strategies, improve product placement, and enhance the overall customer shopping experience.

The term "market basket" originates from the concept of a shopping basket filled with items that a customer intends to purchase during a single shopping trip. Market Basket Analysis helps retailers and businesses understand the contents of these baskets, enabling them to make data-driven decisions and ultimately increase revenue and customer satisfaction.

Key Concepts and Objectives of Market Basket Analysis:

#### 1. **Association Rules:**

The fundamental concept in Market Basket Analysis is the discovery of association rules, which describe relationships between products. These rules typically take the form of "If a customer buys product A, they are likely to buy product B." The two key metrics associated with association rules are support and confidence.

- **Support:**

This measures the frequency of a product or product combination in all transactions. It tells us how often a particular association occurs.

- **Confidence:**

This indicates the likelihood of one product being bought when another product is already in the customer's basket.

#### 2. **Apriori Algorithm:**

The Apriori algorithm is a popular method for finding association rules in large datasets. It works by progressively identifying frequent itemsets (combinations of products) and generating association rules based on these frequent itemsets.

### 3. Use Cases:

Market Basket Analysis has several practical applications, including:

- **Recommendation Systems:**

E-commerce websites can use the insights gained from market basket analysis to suggest related or complementary products to customers.

- **Inventory Management:**

Retailers can optimize their inventory by ensuring that frequently associated products are stocked together.

- **Pricing and Promotions:**

Understanding which products are often purchased together can inform pricing strategies and promotional campaigns.

- **Store Layout and Product Placement:**

Retail stores can improve the layout and placement of products on shelves to encourage cross-selling.

- **Customer Segmentation:**

Segmenting customers based on their purchase patterns can help tailor marketing efforts.

### 4. Challenges:

Market Basket Analysis also faces challenges, such as dealing with large datasets, handling sparse data, and ensuring that the discovered associations are meaningful and not purely coincidental.

## Code:

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

```

import matplotlib.pyplot as plt
import seaborn as sns
import squarify
from statsmodels.tsa.seasonal import seasonal_decompose
import os

for dirname, _, filenames in os.walk('/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))
data = pd.read_csv("../input/market-basket-analysis/Assignment-1_Data.csv", sep = ";")
data.head(10)
data.isnull().sum()
plt.figure(figsize=(15,6))
sns.heatmap(data.isna().transpose())
data.shape
data = data.dropna(subset=["Itemname"])
data.isnull().sum()
plt.figure(figsize=(15,6))
sns.heatmap(data.isna().transpose())
data["Itemname"].value_counts()
data.shape
data.dtypes
data[data['CustomerID'].isnull()]
data = data.fillna(0)
data[data["CustomerID"].isnull()]
data.isnull().sum()
data.shape
data["Date"] = pd.to_datetime(data["Date"])

```

```

data["Price"] = data["Price"].str.replace(",",".")
data["Price"] = pd.to_numeric(data["Price"])
data["Price"]
data.dtypes
data["Sales"] = data["Quantity"] * data["Price"]
top20items = pd.DataFrame(data["Itemname"].value_counts().head(20))
top20items = top20items.reset_index()
top20items.columns = ["Itemname","Frequency"]
labels = top20items["Itemname"]
sizes = top20items["Frequency"]
top20items
fig = plt.figure(figsize=(16,6))
colors = sns.color_palette("Spectral",20)
squarify.plot(sizes, label=labels, color = colors)
plt.title("Top 20 Products")
t_light = data[data["Itemname"]== "WHITE HANGING HEART T-LIGHT HOLDER"]
t_light
fig = plt.figure(figsize=(24,8))
sns.lineplot(x = t_light["Date"], y = t_light["Quantity"] )
startdate = t_light["Date"] >= "20110101"
t_light_2011 = t_light.loc[startdate]
t_light_2011.head()
plt.figure(figsize=(16,6))
sns.lineplot(t_light_2011["Date"],t_light_2011["Sales"])
country_specific = data.groupby(["Country"]).sum().reset_index()
country_specific = country_specific.sort_values(["Sales"], ascending = False)
country_specific_10 = country_specific.head(10)
country_specific_10

```

```

plt.figure(figsize=(15,6))

p = sns.barplot(country_specific_10["Country"], country_specific_10["Sales"])

p.set_ylabel("Sales (millions)")

basket =
data.groupby(["BillNo", "Itemname"])["Quantity"].sum().unstack().reset_index().fillna(0).s
et_index("BillNo")

basket

def one_hot_encoding(x):
    if x <= 0:
        return 0
    if x >= 1:
        return 1

basket = basket.applymap(one_hot_encoding)

basket

frequent_itemsets = apriori(basket, min_support=0.02, use_colnames=True)

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

rules = rules.sort_values("lift", ascending = False)

rules

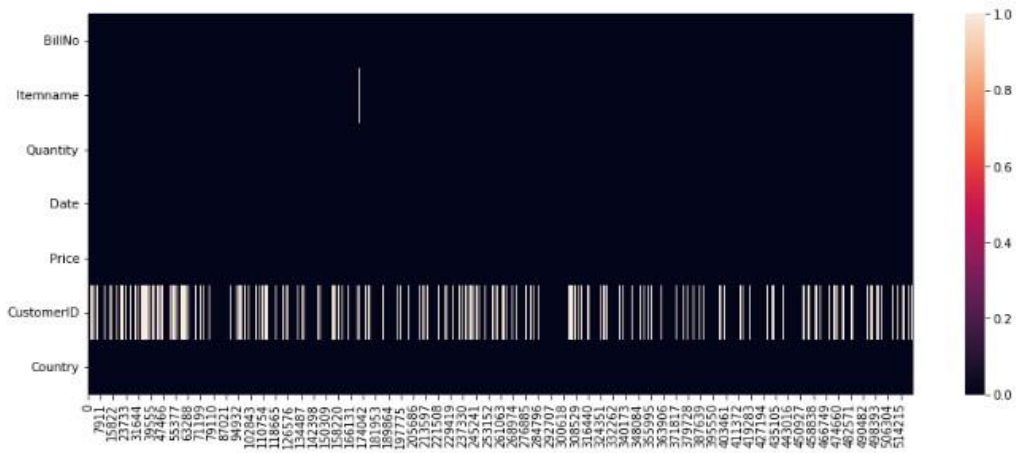
```

## Output:

	BillNo	Itemname	Quantity	Date	Price	CustomerID	Country
0	536365	WHITE HANGING HEART T-LIGHT HOLDER	6	01.12.2010 08:26	2,55	17850.0	United Kingdom
1	536365	WHITE METAL LANTERN	6	01.12.2010 08:26	3,39	17850.0	United Kingdom
2	536365	CREAM CUPID HEARTS COAT HANGER	8	01.12.2010 08:26	2,75	17850.0	United Kingdom
3	536365	KNITTED UNION FLAG HOT WATER BOTTLE	6	01.12.2010 08:26	3,39	17850.0	United Kingdom
4	536365	RED WOOLLY HOTTIE WHITE HEART.	6	01.12.2010 08:26	3,39	17850.0	United Kingdom
5	536365	SET 7 BABUSHKA NESTING BOXES	2	01.12.2010 08:26	7,65	17850.0	United Kingdom
6	536365	GLASS STAR FROSTED T-LIGHT HOLDER	6	01.12.2010 08:26	4,25	17850.0	United Kingdom
7	536366	HAND WARMER UNION JACK	6	01.12.2010 08:28	1,85	17850.0	United Kingdom
8	536366	HAND WARMER RED POLKA DOT	6	01.12.2010 08:28	1,85	17850.0	United Kingdom
9	536367	ASSORTED COLOUR BIRD ORNAMENT	32	01.12.2010 08:34	1,69	13047.0	United Kingdom

```
BillNo      0
Itemname    1455
Quantity    0
Date        0
Price       0
CustomerID  134041
Country     0
dtype: int64
```

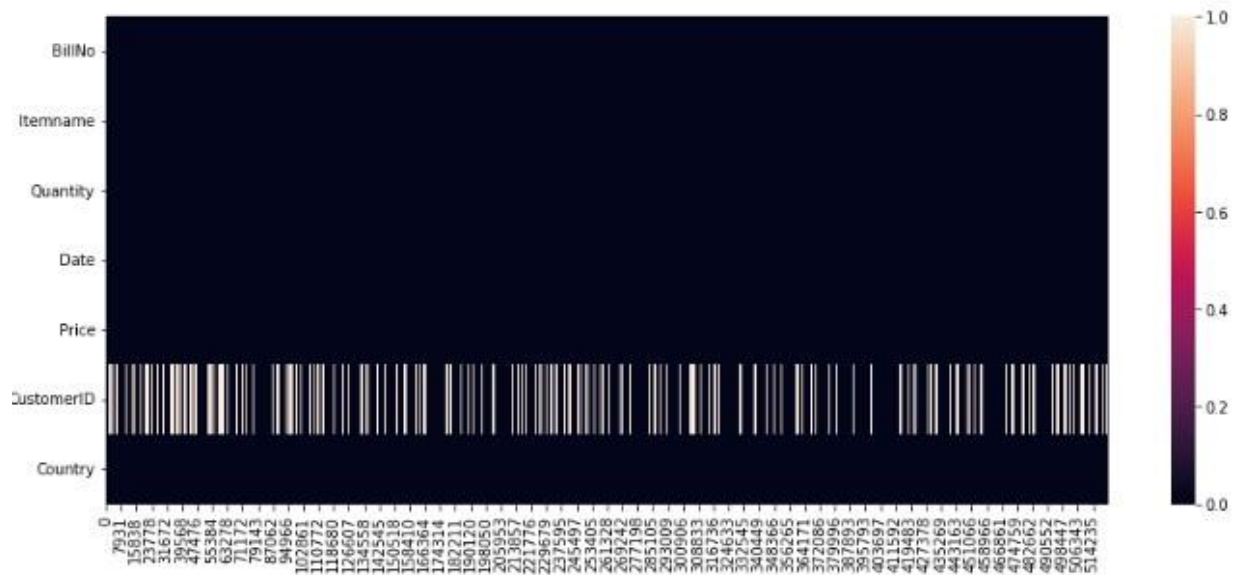
<AxesSubplot:>



(522064, 7)

```
BillNo      0
Itemname    0
Quantity    0
Date        0
Price       0
CustomerID  132586
Country     0
dtype: int64
```

<AxesSubplot:>



```
WHITE HANGING HEART T-LIGHT HOLDER      2269
JUMBO BAG RED RETROSPOT                  2087
REGENCY CAKESTAND 3 TIER                  1930
PARTY BUNTING                            1677
LUNCH BAG RED RETROSPOT                    1570
...
FOOD COVER WITH BEADS , SET 2 SIZES        1
POLYESTER FILLER PAD 60x40cm                1
damages/credits from ASOS.                  1
samples                                     1
PAPER CRAFT , LITTLE BIRDIE                 1
Name: Itemname, Length: 4185, dtype: int64
```

(520609, 7)

```
BillNo      object
Itemname     object
Quantity     int64
Date         object
Price        object
CustomerID   float64
Country      object
dtype: object
```



	BillNo	Itemname	Quantity	Date	Price	CustomerID	Country
1410	536544	DECORATIVE ROSE BATHROOM BOTTLE	1	01.12.2010 14:32	2,51	NaN	United Kingdom
1411	536544	DECORATIVE CATS BATHROOM BOTTLE	2	01.12.2010 14:32	2,51	NaN	United Kingdom
1412	536544	POLKADOT RAIN HAT	4	01.12.2010 14:32	0,85	NaN	United Kingdom
1413	536544	RAIN PONCHO RETROSPOT	2	01.12.2010 14:32	1,66	NaN	United Kingdom
1414	536544	VINTAGE SNAP CARDS	9	01.12.2010 14:32	1,66	NaN	United Kingdom
...	...	...	...	...	...	...	...
521695	581498	JUMBO BAG RED RETROSPOT	5	09.12.2011 10:26	4,13	NaN	United Kingdom
521696	581498	JUMBO BAG BAROQUE BLACK WHITE	4	09.12.2011 10:26	4,13	NaN	United Kingdom
521697	581498	LADIES & GENTLEMEN METAL SIGN	1	09.12.2011 10:26	4,96	NaN	United Kingdom
521698	581498	S/4 CACTI CANDLES	1	09.12.2011 10:26	10,79	NaN	United Kingdom

BillNo	Itemname	Quantity	Date	Price	CustomerID	Country
--------	----------	----------	------	-------	------------	---------

```

BillNo      0
Itemname    0
Quantity    0
Date        0
Price       0
CustomerID  0
Country     0
dtype: int64
(520609, 7)

```

```

0      2.55
1      3.39
2      2.75
3      3.39
4      3.39
...
522059  0.85
522060  2.10
522061  4.15
522062  4.15
522063  4.95

```

Name: Price, Length: 520609, dtype: float64

```

BillNo      object
Itemname     object
Quantity     int64
Date         datetime64[ns]
Price        float64
CustomerID   float64
Country      object
dtype: object

```

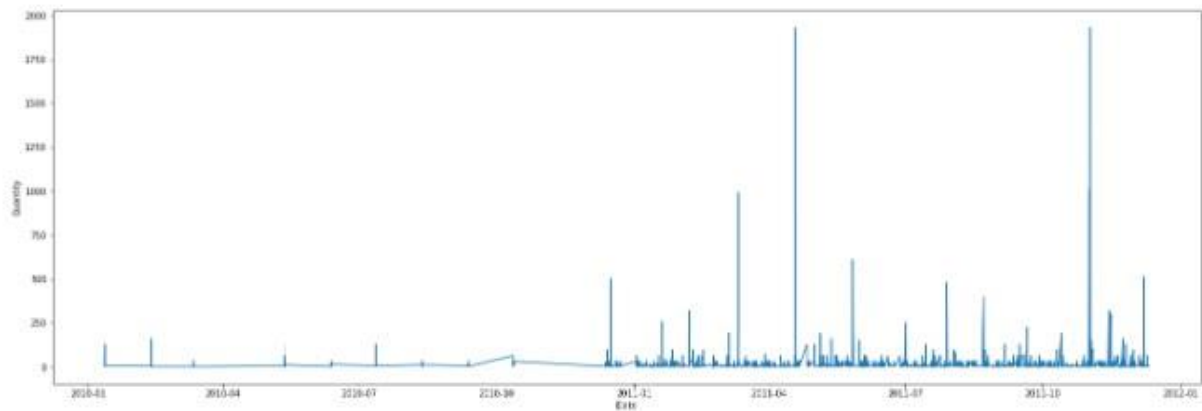
	BillNo	Itemname	Quantity	Date	Price	CustomerID	Country	Sales
0	536365	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-01-12 08:26:00	2.55	17850.0	United Kingdom	15.30
1	536365	WHITE METAL LANTERN	6	2010-01-12 08:26:00	3.39	17850.0	United Kingdom	20.34
2	536365	CREAM CUPID HEARTS COAT HANGER	8	2010-01-12 08:26:00	2.75	17850.0	United Kingdom	22.00
3	536365	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-01-12 08:26:00	3.39	17850.0	United Kingdom	20.34
4	536365	RED WOOLLY HOTTIE WHITE HEART.	6	2010-01-12 08:26:00	3.39	17850.0	United Kingdom	20.34

	Itemname	Frequency
0	WHITE HANGING HEART T-LIGHT HOLDER	2269
1	JUMBO BAG RED RETROSPOT	2087
2	REGENCY CAKESTAND 3 TIER	1930
3	PARTY BUNTING	1677
4	LUNCH BAG RED RETROSPOT	1570
5	ASSORTED COLOUR BIRD ORNAMENT	1465
6	SET OF 3 CAKE TINS PANTRY DESIGN	1360
7	PACK OF 72 RETROSPOT CAKE CASES	1328
8	LUNCH BAG BLACK SKULL.	1315
9	NATURAL SLATE HEART CHALKBOARD	1246
10	JUMBO BAG PINK POLKADOT	1231
11	HEART OF WICKER SMALL	1206
12	JUMBO STORAGE BAG SUKI	1191
13	PAPER CHAIN KIT 50'S CHRISTMAS	1183
14	JUMBO SHOPPER VINTAGE RED PAISLEY	1181
15	LUNCH BAG SPACEBOY DESIGN	1169
16	LUNCH BAG CARS BLUE	1156
17	SPOTTY BUNTING	1133
18	JAM MAKING SET PRINTED	1123
19	LUNCH BAG SUKI DESIGN	1112

Text(0.5, 1.0, 'Top 20 Products')

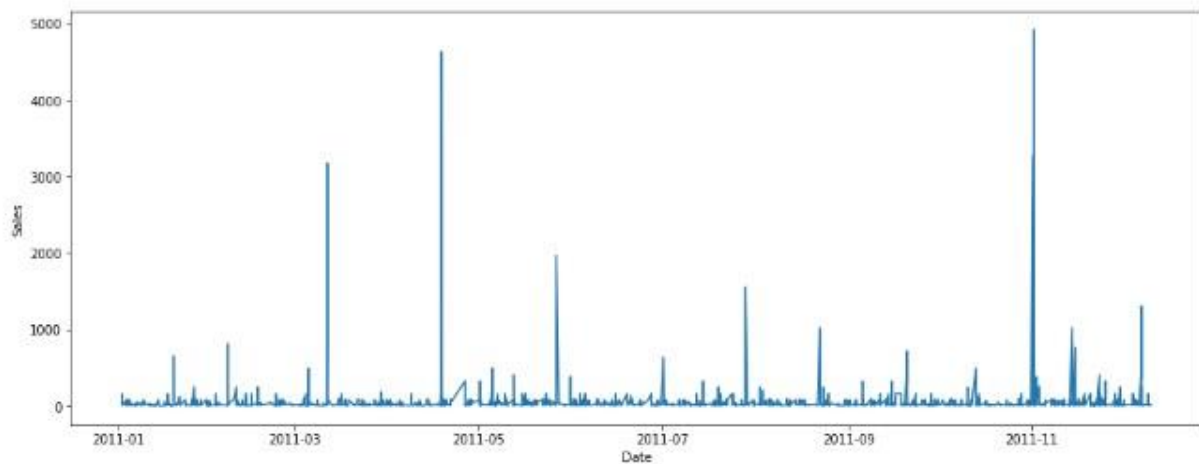


	BillNo	Itemname	Quantity	Date	Price	CustomerID	Country	Sales
0	536365	WHITE HANGING HEART T- LIGHT HOLDER	6	2010-01- 12 08:26:00	2.55	17850.0	United Kingdom	15.30
49	536373	WHITE HANGING HEART T- LIGHT HOLDER	6	2010-01- 12 09:02:00	2.55	17850.0	United Kingdom	15.30
66	536375	WHITE HANGING HEART T- LIGHT HOLDER	6	2010-01- 12 09:32:00	2.55	17850.0	United Kingdom	15.30
218	536390	WHITE HANGING HEART T- LIGHT HOLDER	64	2010-01- 12 10:19:00	2.55	17511.0	United Kingdom	163.20
253	536394	WHITE HANGING HEART T- LIGHT HOLDER	32	2010-01- 12 10:39:00	2.55	13408.0	United Kingdom	81.60
...	...	...	...	...	...	...	...	...
517571	581246	WHITE HANGING HEART T-	1	2011-08- 12	2.95	15453.0	United Kingdom	2.95



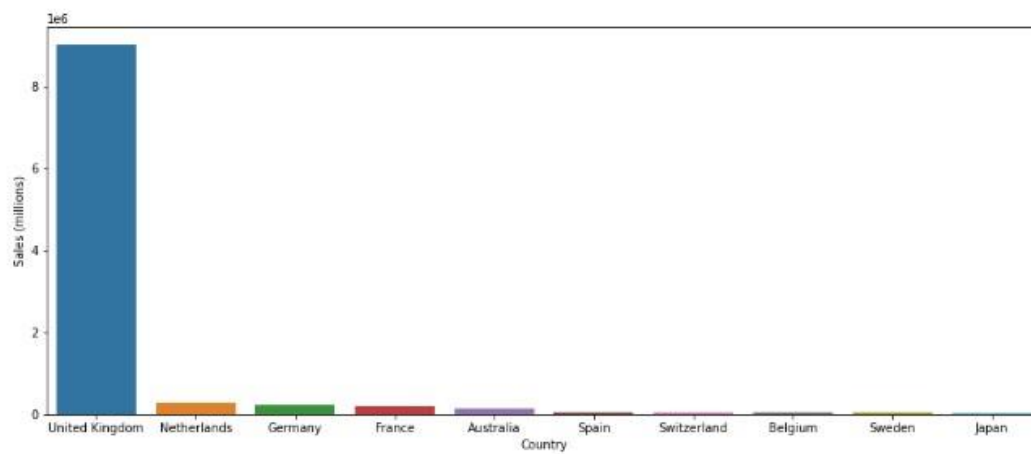
	BillNo	Itemname	Quantity	Date	Price	CustomerID	Country	Sales
41255	539993	WHITE HANGING HEART T-LIGHT HOLDER	12	2011-04-01 10:00:00	2.95	13313.0	United Kingdom	35.40
41400	540014	WHITE HANGING HEART T-LIGHT HOLDER	6	2011-04-01 11:34:00	2.95	14606.0	United Kingdom	17.70
41448	540016	WHITE HANGING HEART T-LIGHT HOLDER	10	2011-04-01 11:42:00	2.95	16282.0	United Kingdom	29.50
41787	540026	WHITE HANGING HEART T-LIGHT HOLDER	1	2011-04-01 13:25:00	5.91	0.0	United Kingdom	5.91
42168	540033	WHITE HANGING HEART T-LIGHT HOLDER	32	2011-04-01 14:12:00	2.55	13680.0	United Kingdom	81.60

<AxesSubplot:xlabel='Date', ylabel='Sales'>



	Country	Quantity	Price	CustomerID	Sales
28	United Kingdom	4540871	1845443.914	5.510928e+09	9003097.964
16	Netherlands	200937	6247.730	3.407895e+07	285446.340
6	Germany	119263	33532.140	1.143723e+08	228867.140
5	France	112104	36992.790	1.057630e+08	209715.110
0	Australia	84209	3612.500	1.475853e+07	138521.310
23	Spain	27951	9504.340	3.208028e+07	61577.110
25	Switzerland	30630	6634.360	2.332474e+07	57089.900
3	Belgium	23237	7372.850	2.524670e+07	41196.340
24	Sweden	36083	1695.790	6.637579e+06	38378.330
12	Japan	26016	657.210	4.095425e+06	37416.370

Text(0, 0.5, 'Sales (millions)')





Itemname	*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
BillNo								
536365	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
536366	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
536367	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
536368	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
536369	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
...	...	...	...	...	...	...	...	...
572057	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
572058	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
A563185	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
A563186	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
A563187	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Itemname	*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
BillNo								
536365	0	0	0	0	0	0	0	0
536366	0	0	0	0	0	0	0	0
536367	0	0	0	0	0	0	0	0
536368	0	0	0	0	0	0	0	0
536369	0	0	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...
572057	0	0	0	0	0	0	0	0
572058	0	0	0	0	0	0	0	0
A563185	0	0	0	0	0	0	0	0
A563186	0	0	0	0	0	0	0	0
A563187	0	0	0	0	0	0	0	0

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
153	(ROSES REGENCY TEACUP AND SAUCER, GREEN REGENC...	(PINK REGENCY TEACUP AND SAUCER)	0.036170	0.036418	0.025482	0.704514	19.345428	0.024165	3.261013
156	(PINK REGENCY TEACUP AND SAUCER)	(ROSES REGENCY TEACUP AND SAUCER, GREEN REGENC...	0.036418	0.036170	0.025482	0.699728	19.345428	0.024165	3.209858
154	(PINK REGENCY TEACUP AND SAUCER, ROSES REGENCY...	(GREEN REGENCY TEACUP AND SAUCER)	0.028204	0.048243	0.025482	0.903509	18.728115	0.024122	9.863659
155	(GREEN REGENCY TEACUP AND SAUCER)	(PINK REGENCY TEACUP AND SAUCER, ROSES REGENCY...	0.048243	0.028204	0.025482	0.528205	18.728115	0.024122	2.059785
25	(GREEN REGENCY TEACUP AND SAUCER)	(PINK REGENCY TEACUP AND SAUCER)	0.048243	0.036418	0.029936	0.620513	17.038810	0.028179	2.539170
...	...	...	...	...	...	...	...	...	...
149	(WOODEN PICTURE FRAME WHITE FINISH)	(WHITE HANGING HEART T-LIGHT HOLDER)	0.054033	0.108956	0.020040	0.370879	3.403936	0.014152	1.416332
131	(WHITE HANGING HEART T-LIGHT HOLDER)	(NATURAL SLATE HEART CHALKBOARD)	0.108956	0.060960	0.020336	0.186649	3.061823	0.013695	1.154532
130	(NATURAL SLATE HEART CHALKBOARD)	(WHITE HANGING HEART T-LIGHT HOLDER)	0.060960	0.108956	0.020336	0.333604	3.061823	0.013695	1.337109
80	(JUMBO BAG RED RETROSPOT)	(WHITE HANGING HEART T-LIGHT HOLDER)	0.102128	0.108956	0.021870	0.214147	1.965448	0.010743	1.133856

## Conclusion:

Market Basket Analysis is a data mining technique that uncovers valuable insights about product associations in customer transactions. It is a vital tool for businesses in retail and e-commerce, enabling them to make informed decisions regarding inventory management, pricing strategies, recommendation systems, store layouts, and customer segmentation. By understanding which products are frequently purchased together, businesses can enhance their operations, improve the customer experience, and boost revenue.