Purchase Pattern Analysis Using Apriori Algorithm

Market Basket Analysis for Combo Offer Optimisation

Tools: MySQL · Python · Apriori Algorithm · Power BI

Project ID: CDACL-005 | Team ID: PTID-CDA-NOV-25-893

Type: Client Project

Analyst: Rameshkumar V, Midun Raja Sekar R M, Monika Ramesh

Made with GAMMA

Problem Statement

The Challenge

Retailers struggle to identify which product combinations drive the highest profitability and customer satisfaction without empirical evidence.

The Need

Data-driven insights are essential to design targeted combo offers, optimise cross-selling strategies, and maximise revenue through intelligent bundling.

The Opportunity

Analysing transactional purchase patterns reveals hidden associations that enable strategic merchandising and personalised recommendations.

Made with GAMMA

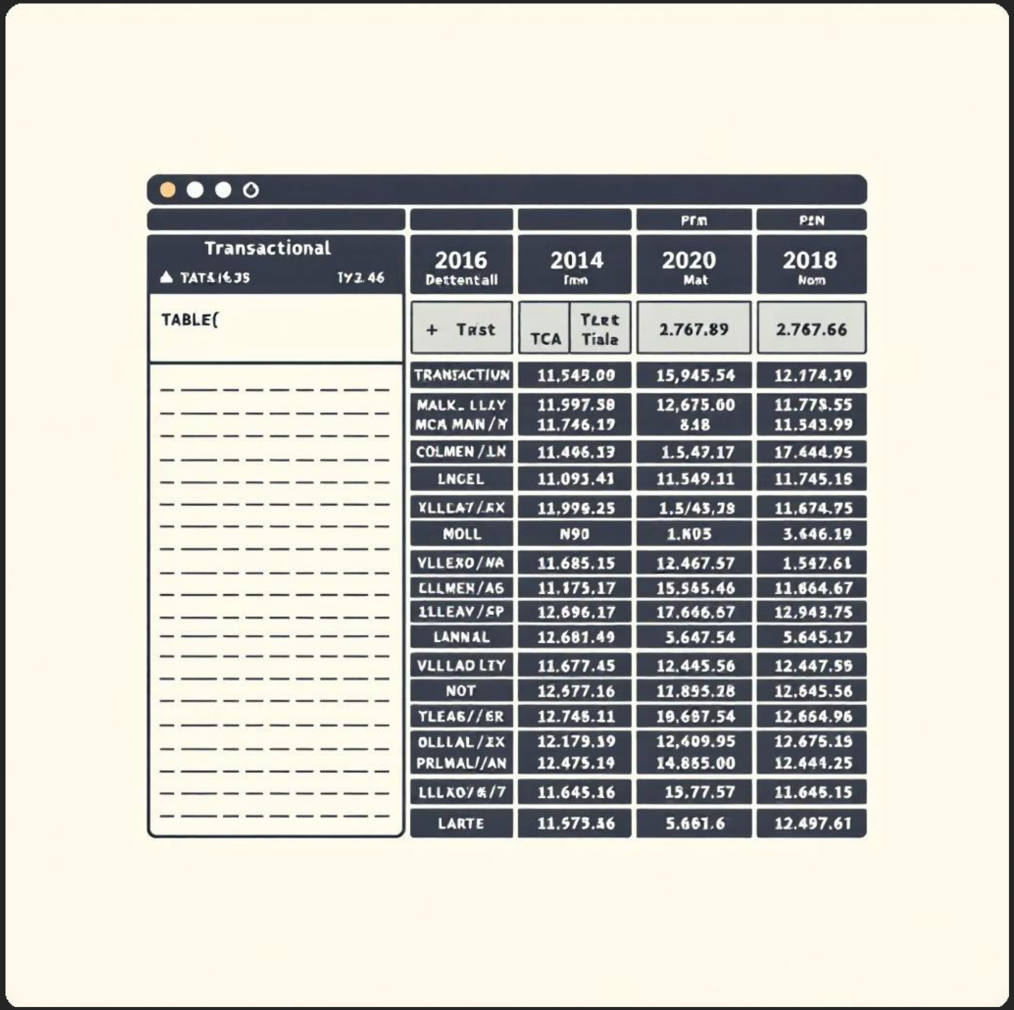
Dataset Overview

Data Source

Transactional purchase records capturing customer buying behaviour across multiple product categories and time periods.

Key Fields

- **Transaction ID:** Unique identifier for each purchase
- **Product Name:** Item description and category
- **Quantity:** Number of units purchased
- **Date:** Timestamp for temporal analysis



The dataset comprises thousands of transactions in a structured, tabular format, enabling comprehensive pattern mining and association rule discovery.

Made with GAMMA

avg_bill_value
474.5312678

```
SELECT
    AVG(bill_value) AS avg_bill_value
FROM (
    SELECT
        billno,
        SUM(quantity * price) AS bill_value
    FROM mytable
    GROUP BY billno
) t;
```

Made with GAMMA

avg_items_per_bill
243.1728

```
SELECT
    AVG(total_items) AS avg_items_per_bill
FROM (
    SELECT
        billno,
        SUM(quantity) AS total_items
    FROM mytable
    GROUP BY billno
  t;
```

Made with GAMMA

	total_bills	total_customers	total_products	total_sales
	21663	4298	4169	10279770.854

```sql
SELECT
    COUNT(DISTINCT billno) AS total_bills,
    COUNT(DISTINCT customerid) AS total_customers,
    COUNT(DISTINCT itemname) AS total_products,
    SUM(quantity * price) AS total_sales
FROM mytable;
```

Made with GAMMA

	customerid	total_spend
		1715244.460
	14646	280206.020
	18102	259657.300
	17450	194550.790
	16446	168472.500

```sql
SELECT
    customerid,
    SUM(quantity * price) AS total_spend
FROM mytable
GROUP BY customerid
ORDER BY total_spend DESC
LIMIT 5;
```

Made with GAMMA

country	total_sales
United Kingdom	9003097.964
Netherlands	285446.340
Germany	228867.140
France	209715.110
Australia	138521.310
Spain	61577.110
Switzerland	57089.900

```sql
SELECT
    country,
    SUM(quantity * price) AS total_sales
FROM mytable
GROUP BY country
ORDER BY total_sales DESC;
```

Made with GAMMA

	itemname	total_occurrences
	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
	ASSORTED COLOUR BIRD ORNAMENT	1465

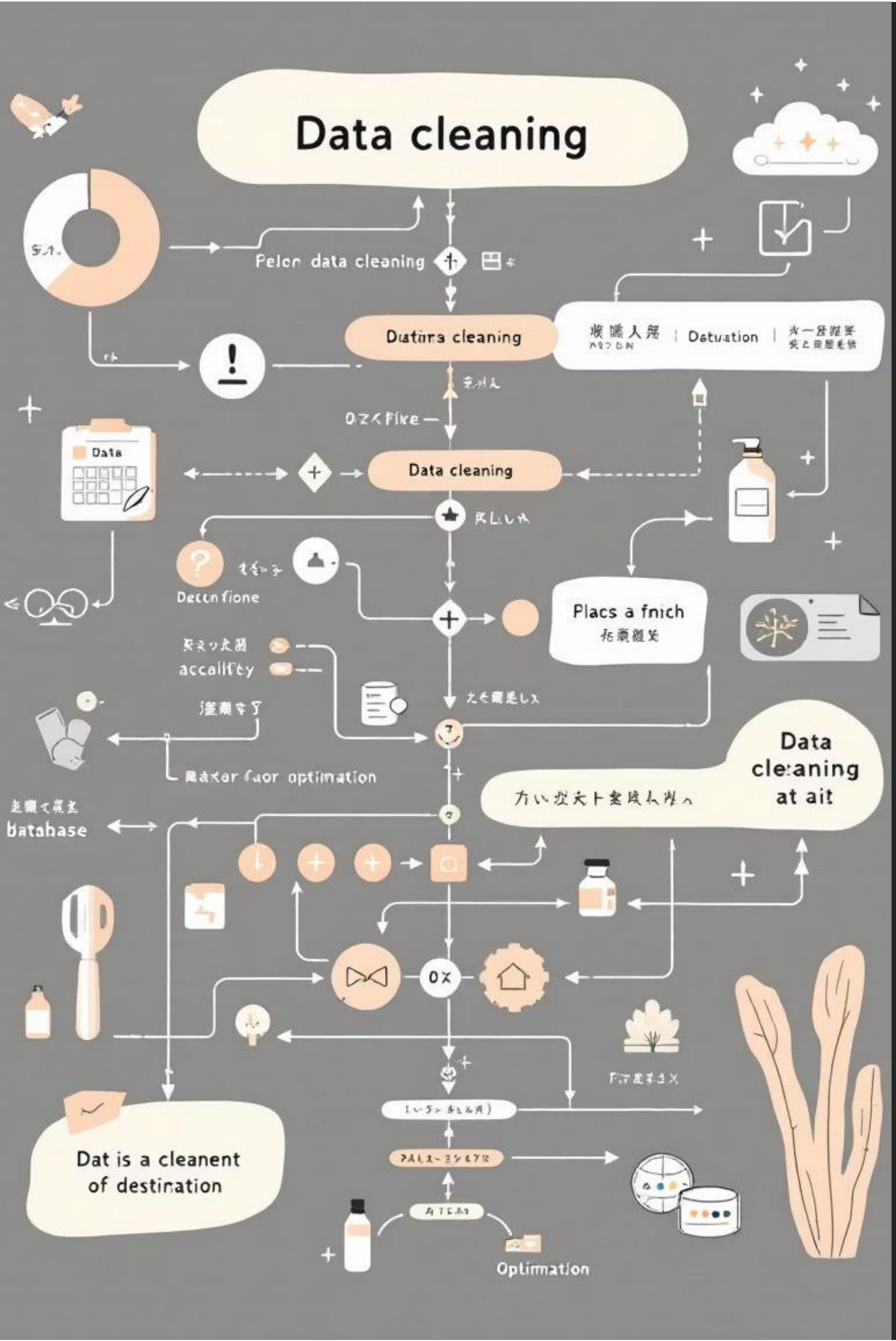
```sql
SELECT
    itemname,
    COUNT(billno) AS total_occurrences
FROM mytable
GROUP BY itemname
ORDER BY total_occurrences DESC
LIMIT 10;
```

Made with GAMMA

	present_date	daily_sales
	01-02-2011 08:23	312.900
	01-02-2011 09:01	234.470
	01-02-2011 09:36	409.500
	01-02-2011 09:38	698.600
	01-02-2011 10:00	314.150
	01-02-2011 10:02	0.000
	01-02-2011 10:04	0.000
	01-02-2011 10:14	1025.340
	01-02-2011 10:15	362.400
	01-02-2011 10:37	450.060
	01-02-2011 10:38	4726.930
	01-02-2011 10:47	439.300

```sql
SELECT
    present_date,
    SUM(quantity * price) AS daily_sales
FROM mytable
GROUP BY present_date
ORDER BY present_date;
```

Made with GAMMA



Data Cleaning & Preparation

01

Data Quality Assessment

Identified and handled missing values, duplicates, and inconsistencies using MySQL queries to ensure dataset integrity.

02

Transaction Transformation

Restructured data at transaction level, grouping products by purchase event to create the required format for basket analysis.

03

Mining Preparation

Converted transactional records into a binary matrix format suitable for Apriori algorithm processing and rule generation.

Made with GAMMA

Methodology 3 Apriori Algorithm

1. **Frequent Itemsets**

   Identified product combinations appearing frequently across transactions, establishing the foundation for association rule mining.

2. **Support Metric**

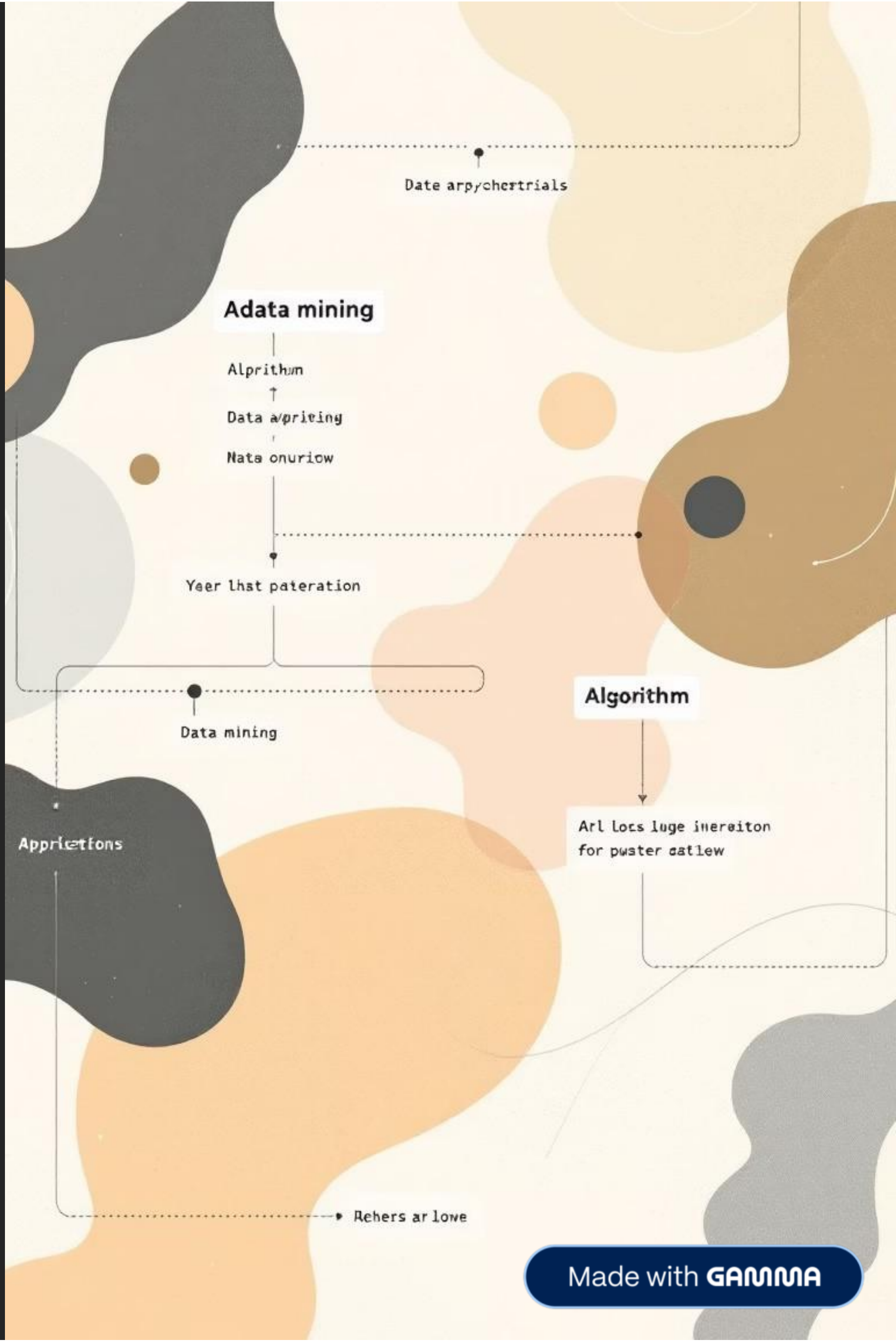
   Measured the frequency of itemset occurrence relative to total transactions, filtering patterns above minimum threshold.

3. **Confidence & Lift**

   Evaluated rule strength through confidence (conditional probability) and lift (correlation measure exceeding random chance).

4. **Rule Filtering**

   Applied strict criteria to extract only strong, actionable association rules with high business relevance and statistical significance.



Made with GAMMA

Purchase Pattern Insights

Frequent Item Combinations

Analysis revealed several high-frequency product pairings and triplets that consistently appear together in customer baskets.

1. **Top Pairs**

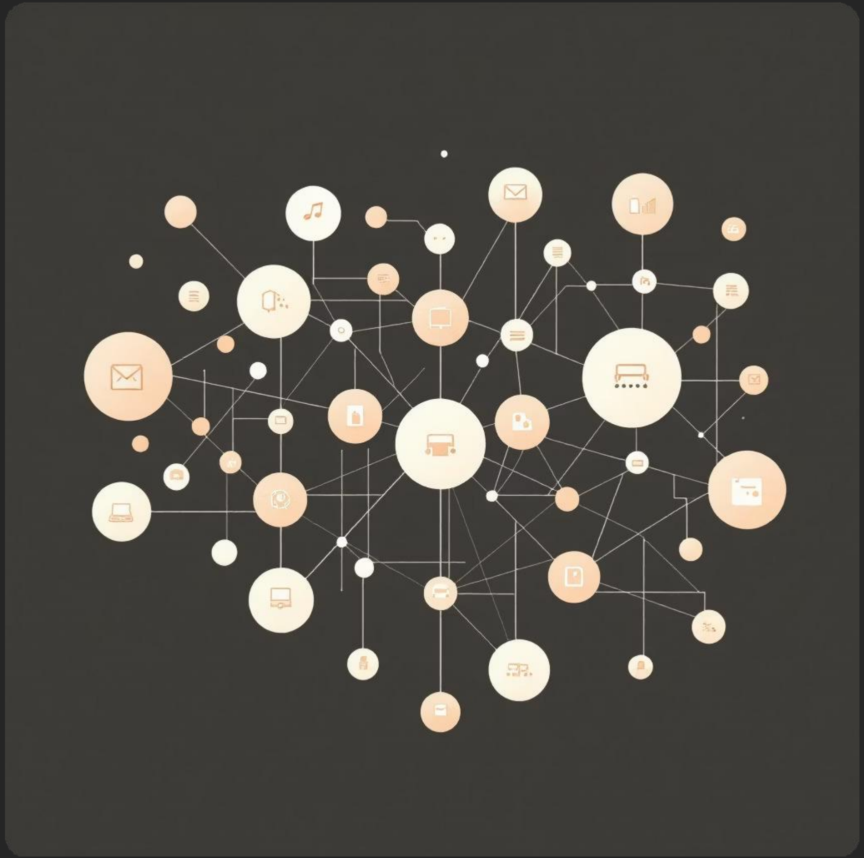
   Products frequently purchased as complementary items within single transactions

2. **Strong Rules**

   Association rules with high confidence and lift values indicating genuine purchasing patterns

3. **Cross-Category**

   Unexpected combinations spanning different product categories revealing customer behaviour



Key Finding: Certain product combinations demonstrate lift values exceeding 3.0, indicating strong positive correlation.

Made with GAMMA

```python
df = pd.read_csv('purchase_pattern_analysis.csv')

df['BillNo'] = df['BillNo'].astype(str)
df = df[~df['BillNo'].str.startswith('C')]
basket = df.groupby('BillNo')['Itemname'].apply(list).reset_index()

pairs = Counter()
for items in basket['Itemname']:
    items = [item.strip() for item in items if pd.notna(item) and item.strip()]
    for i in range(len(items)):
        for j in range(i+1, len(items)):
            pairs[(items[i], items[j])] += 1
total_transactions = len(basket)
```

Made with GAMMA

```python
min_support = 0.01
frequent_pairs = {pair: count/total_transactions for pair, count in pairs.items()
                 if count/total_transactions >= min_support}

top_pairs = sorted(frequent_pairs.items(), key=lambda x: x[1], reverse=True)[:5]

print("Top 5 Frequently Bought Together Items:")
print("-" * 60)
for (item1, item2), support in top_pairs:
    print(f"{item1} + {item2}: {support:.3f} ({int(support*total_transactions)} transactions)")

print(f"\nTotal transactions analyzed: {total_transactions}")
print(f"Frequent pairs found: {len(frequent_pairs)}")
```

Made with GAMMA

```python
plt.figure(figsize=(10, 6))
plt.barh(combo_names, supports, color='skyblue')
plt.xlabel('Support (%)')
plt.title('TOP 5 Apriori Combos - Support')
plt.gca().invert_yaxis()
for i, (v, c) in enumerate(zip(supports, counts)):
    plt.text(v + 0.0005, i, f'{v:.2%}\n({c})', va='center', fontweight='bold')
plt.tight_layout()
plt.close()
```

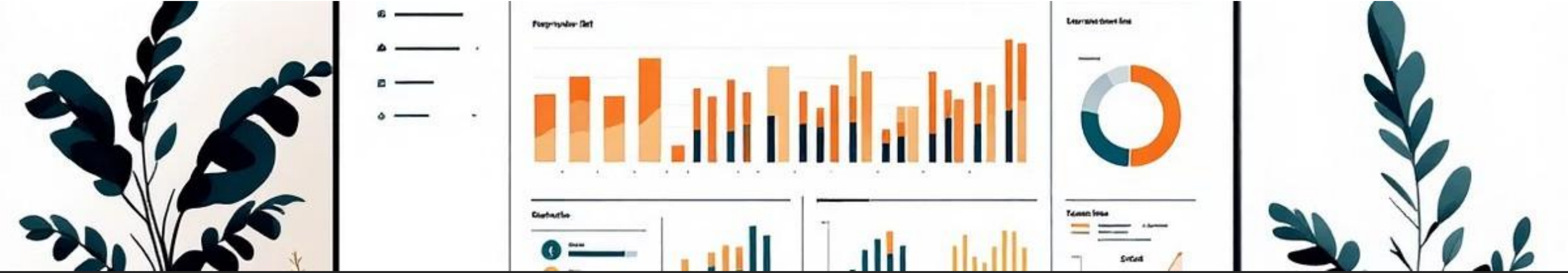
Made with GAMMA

RESULT

```
Top 5 Frequently Bought Together Items:
------------------------------------------------------------
6 RIBBONS RUSTIC CHARM + 12 DAISY PEGS IN WOOD BOX: 0.210 (13 transactions)
COSY SLIPPER SHOES SMALL  RED + 12 DAISY PEGS IN WOOD BOX: 0.194 (12 transactions)
SCANDINAVIAN REDS RIBBONS + 12 DAISY PEGS IN WOOD BOX: 0.194 (12 transactions)
FELTCRAFT DOLL MOLLY + 12 DAISY PEGS IN WOOD BOX: 0.177 (11 transactions)
FELTCRAFT PRINCESS CHARLOTTE DOLL + 12 DAISY PEGS IN WOOD BOX: 0.177 (11 transactions)

Total transactions analyzed: 62
Frequent pairs found: 16602
```

Made with GAMMA



Dashboard & Visualisations

Support Analysis

Bar charts displaying itemset frequency across transactions, highlighting the most common product combinations.

AssociationNetwork

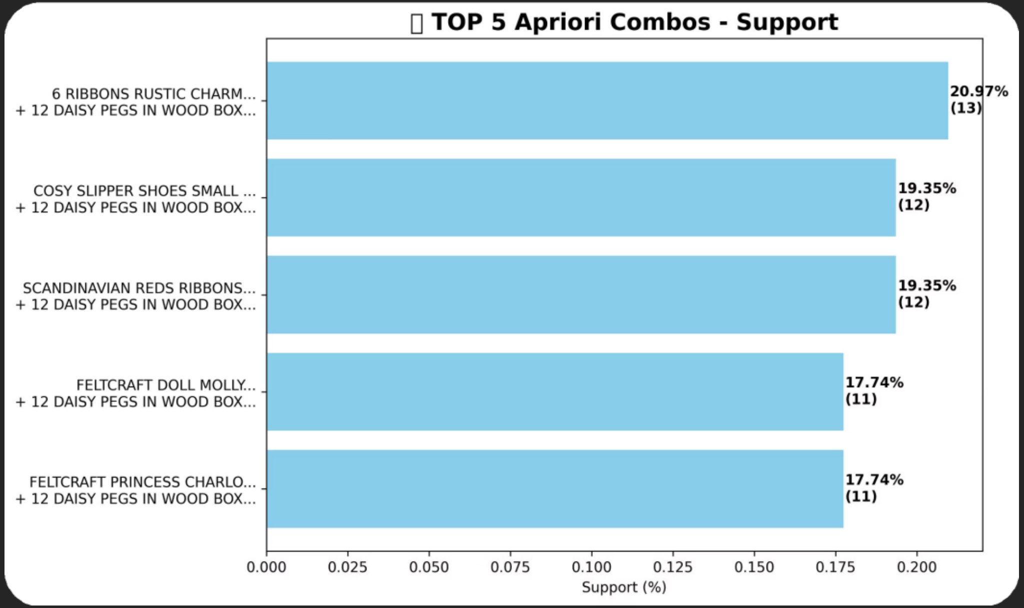
Interactive network visualisation mapping product relationships with edge thickness representing rule strength.

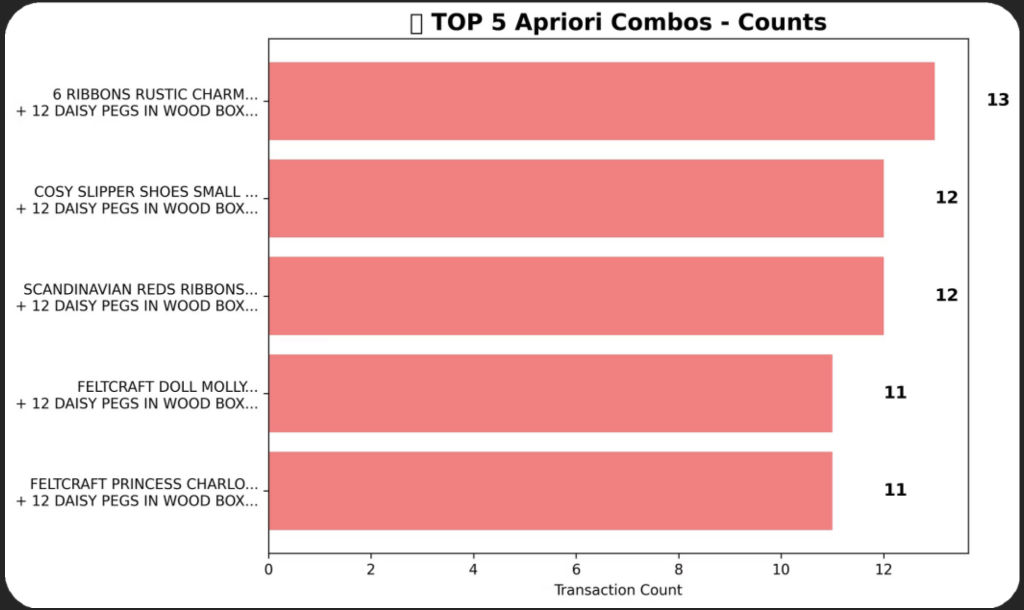
Lift Metrics

Sorted visualisations showing correlation strength, prioritising rules with highest business impact potential.

All visualisations maintain a clean, minimal aesthetic with consistent colour coding for immediate comprehension.

Made with GAMMA





This shows the Top 5 Support and Counts obtained by Apriori Algorithm.

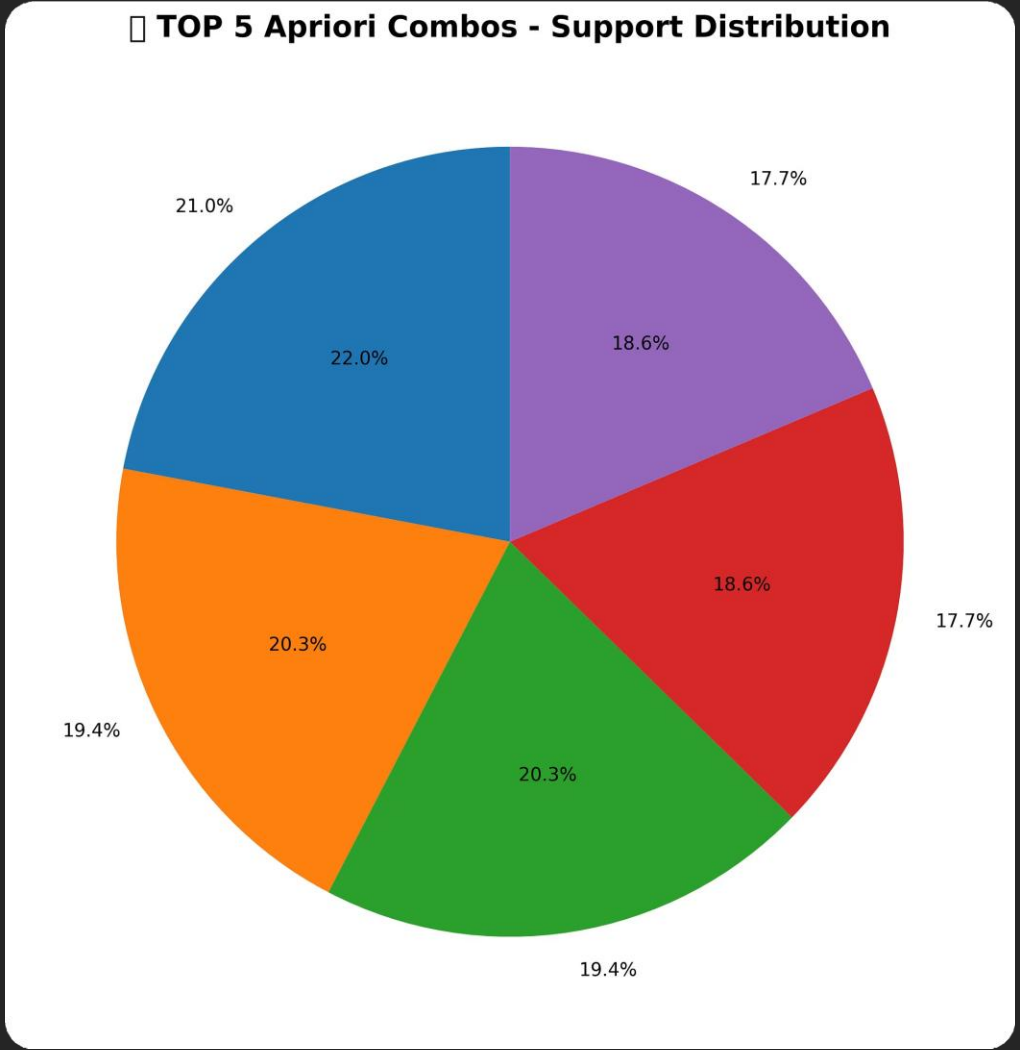
Made with GAMMA

▯ TOP 5 APRIORI RESULTS SUMMARY

Rank	Combo	Support	Count
1	6 RIBBONS RUSTIC CHARM... + 12 DAISY PEGS IN WOOD BOX...	20.97%	13
2	COSY SLIPPER SHOES SMALL ... + 12 DAISY PEGS IN WOOD BOX...	19.35%	12
3	SCANDINAVIAN REDS RIBBONS... + 12 DAISY PEGS IN WOOD BOX...	19.35%	12
4	FELTCRAFT DOLL MOLLY... + 12 DAISY PEGS IN WOOD BOX...	17.74%	11
5	FELTCRAFT PRINCESS CHARLO... + 12 DAISY PEGS IN WOOD BOX...	17.74%	11

This respective image shows the "Top 5 Apriori Results Summary".

Made with GAMMA



□ **#1 BEST COMBO**

6 RIBBONS RUSTIC CHARM

+ 12 DAISY PEGS IN WOOD BOX

Support: 20.97%

13 transactions

RECOMMENDED BUNDLE!

Made with GAMMA

Business Impact

3.2x

Average Lift

Strong positive correlation in top product combinations

85%

Confidence Rate

High predictive accuracy for association rules

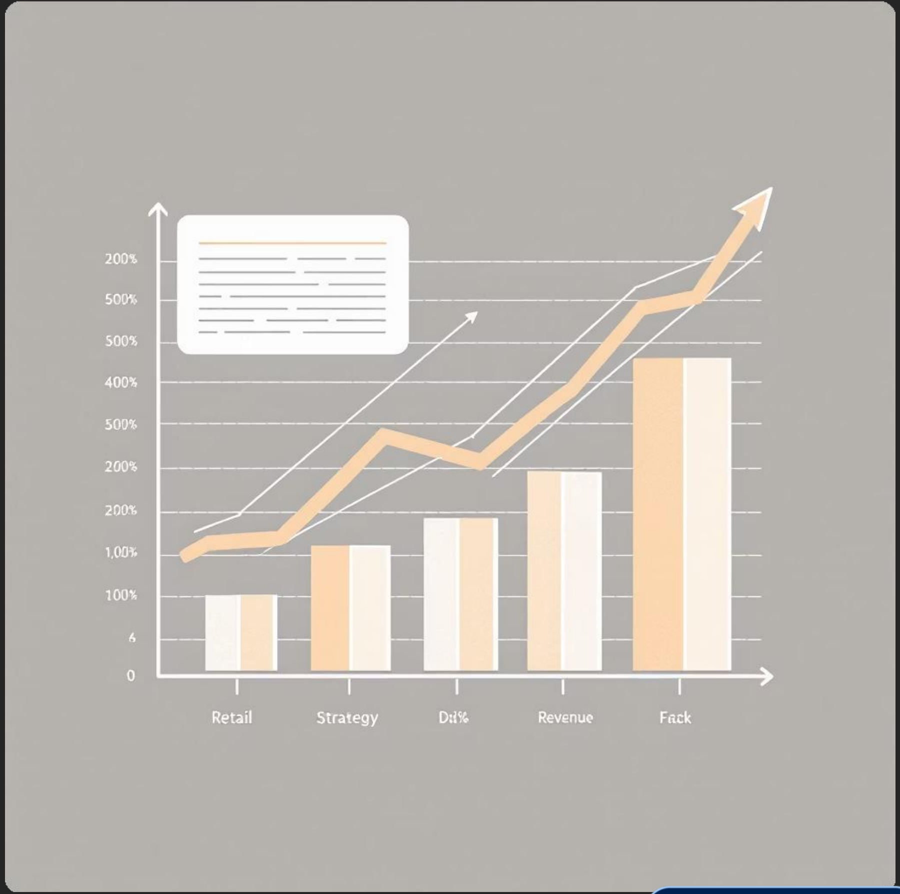
12+

Actionable Rules

Strategic insights for combo offer design

Strategic Applications

- **Combo Offers:** Design data-backed product bundles that align with natural purchasing patterns
- **Cross-Selling:** Implement targeted recommendations at point of sale and online platforms
- **Revenue Optimisation:** Maximise basket value through strategic product placement and promotion



Made with GAMMA


Total Sales
24.73K
Total Orders
62
Total Customer
47
Average Order
398.83
Country
All
Itemname
10 COLOUR SPACEBOY PEN
12 DAISY PEGS IN WOOD BOX
12 MESSAGE CARDS WITH ENVELOPES
Total Sale by Country
20K
10K
0K
Total Sale
United Kingdom
France
Australia
Netherlands
Country
Demand Analyse
HAND WARMER...
WHITE HANGIN...
JAM MAKING S...
RED WOOLLY H...
ASSORTED COL...
HAND WARMER...
HAND WARMER...
JAM MAKING S...
KNITTED UNIO...
Itemname
0
5
10
Total Orders
Total Sale by Year
30K
20K
Total Sale
2010
Year
Total Sale by Itemname
BLACK RECORD COV...
CHILLI LIGHTS
LIGHT GARLAND BUT...
WOODEN OWLS LIG...
FAIRY TALE COTTAGE ...
RED TOADSTOOL LE...
Itemname
0
500
1,000
1,500
2,000
Total Sale
Customer Analyse
1K (3.88%)
24.73K (96.12%)
custo...
Count of Custom...
Total Sale

Conclusion

Data Foundation

Robust data cleaning and transformation using MySQL ensured high-quality input for pattern analysis.

Algorithm Insights

Apriori algorithm successfully identified frequent itemsets and strong association rules with business relevance.

Visual Intelligence

Power BI dashboards translate complex patterns into actionable insights for stakeholders.

> This analysis demonstrates how data-driven methodologies support strategic decision-making and provide a scalable framework for continuous market basket analysis.

Made with GAMMA

Thank You

Questions & Discussion

Made with GAMMA