# Market Basket Analysis & Product Bundling Recommendations

**Full Project Report Student:** Olawale Arowojolu

## 1. Introduction & Problem Statement

A company's profit margins are thin at 3%. They cannot compete on price with large platforms like Amazon, but they can increase average order value (AOV) by suggesting relevant products together.

"If someone buys coffee, what else should we suggest? Find patterns in transaction data. What do people buy together? Can we create 'Frequently Bought Together' bundles or offer discounts like 'Buy coffee and a French press, save 10%'?"

**Objective:** Use association rule mining to discover strong product co-purchases and recommend actionable bundles with evidence (support, confidence, lift) and expected sales impact.

## 2. Dataset

- **Source:** UCI Machine Learning Repository – Online Retail dataset (Excel file)
- **Description:** Real transactions from a UK-based online retailer (2010–2011)
- **Size after cleaning:** 12,088 transactions (invoices), 50 products

## 3. Data Preparation Process (Manual – No Code)

All steps were done manually in Microsoft Excel to create a clean binary transaction table:

1. Opened Online Retail.xlsx
2. Filtered out rows with negative Quantity (returns)
3. Removed rows where InvoiceNo starts with 'C' (cancellations)
4. Filtered to only rows where Country = "United Kingdom" (main market)
5. Trimmed extra spaces from Description column (Data → Text to Columns or Find/Replace)
6. Removed rows with blank Description or CustomerID
7. Created a pivot table to count frequency of each unique Description (product)
8. Sorted by count descending and selected the **top 50 most frequent products**
9. Created a new sheet for binary basket:
    - Rows = unique InvoiceNo
    - Columns = top 50 product names
    - For each cell: Put 1 if the product appears in that invoice (quantity $\geq 1$), 0 otherwise (used VLOOKUP or COUNTIFS per invoice)
10. Saved the binary table as CSV: online_retail_top50_basket.csv

This resulted in a clean, binary market basket dataset ready for Weka.

## 4. Analysis Method

- **Tool:** Weka (free open-source data mining software)
- **Algorithm:** FP-Growth (chosen because it is fast on sparse data and avoids the long running time of Apriori when using lift)
- **Preprocessing in Weka:**
    1. Loaded CSV in Preprocess tab
    2. Ignored InvoiceNo column (right-click → Ignore attribute)
    3. Applied NumericToBinary filter (first-last attributes) to ensure strict binary format
    4. Applied NumericToNominal filter for compatibility
- **FP-Growth Parameters:**
    - Minimum support: 0.02
    - Minimum metric: lift 2.0
    - Number of rules: 50
- **Result:** Generated 124 positive association rules (all presence-based: =1 → =1)

## 5. Top 10 Association Rules with Support and Confidence Scores

| Rank | Antecedent (If buy...) | Consequent (then buy...) | Support | Confidence | Lift |
|---|---|---|---|---|---|
| 1 | ALARM CLOCK BAKELIKE GREEN | ALARM CLOCK BAKELIKE RED | ≈0.057 | 0.66 | 10.49 |
| 2 | ALARM CLOCK BAKELIKE RED | ALARM CLOCK BAKELIKE GREEN | ≈0.063 | 0.60 | 10.49 |
| 3 | LUNCH BAG PINK POLKADOT | LUNCH BAG BLACK SKULL. + LUNCH BAG CARS BLUE | ≈0.021 | 0.29 | 8.86 |
| 4 | LUNCH BAG BLACK SKULL. + LUNCH BAG CARS BLUE | LUNCH BAG PINK POLKADOT | ≈0.021 | 0.62 | 8.86 |
| 5 | LUNCH BAG PINK POLKADOT | LUNCH BAG RED RETROSPOT + LUNCH BAG CARS BLUE | ≈0.021 | 0.30 | 8.77 |
| 6 | LUNCH BAG RED RETROSPOT + LUNCH BAG CARS BLUE | LUNCH BAG PINK POLKADOT | ≈0.021 | 0.62 | 8.77 |
| 7 | LUNCH BAG RED RETROSPOT + LUNCH BAG BLACK SKULL. | LUNCH BAG PINK POLKADOT | ≈0.024 | 0.61 | 8.63 |
| 8 | LUNCH BAG PINK POLKADOT | LUNCH BAG RED RETROSPOT + LUNCH BAG BLACK SKULL. | ≈0.024 | 0.35 | 8.63 |
| 9 | WOODEN PICTURE FRAME WHITE FINISH | WOODEN FRAME ANTIQUE WHITE | ≈0.038 | 0.54 | 8.27 |

| Rank | Antecedent (If buy...) | Consequent (then buy...) | Support | Confidence | Lift |
|---|---|---|---|---|---|
| 10 | WOODEN FRAME ANTIQUE WHITE | WOODEN PICTURE FRAME WHITE FINISH | ≈0.038 | 0.58 | 8.27 |

**6. Translations**

1. 66% of customers who buy **ALARM CLOCK BAKELIKE GREEN** also buy **ALARM CLOCK BAKELIKE RED** (versus ~6.3% of all customers buy **ALARM CLOCK BAKELIKE RED**).
2. 60% of customers who buy **ALARM CLOCK BAKELIKE RED** also buy **ALARM CLOCK BAKELIKE GREEN** (versus ~5.7% of all customers buy **ALARM CLOCK BAKELIKE GREEN**).
3. 29% of customers who buy **LUNCH BAG PINK POLKADOT** also buy **LUNCH BAG BLACK SKULL. and LUNCH BAG CARS BLUE** (versus ~2.1% of all customers buy both).
4. 62% of customers who buy **LUNCH BAG BLACK SKULL. and LUNCH BAG CARS BLUE** also buy **LUNCH BAG PINK POLKADOT** (versus ~3.5% of all customers buy **LUNCH BAG PINK POLKADOT** in this context).
5. 30% of customers who buy **LUNCH BAG PINK POLKADOT** also buy **LUNCH BAG RED RETROSPOT and LUNCH BAG CARS BLUE** (versus ~2.1% of all customers buy both).
6. 62% of customers who buy **LUNCH BAG RED RETROSPOT and LUNCH BAG CARS BLUE** also buy **LUNCH BAG PINK POLKADOT** (versus ~3.5% of all customers buy **LUNCH BAG PINK POLKADOT** in this context).
7. 61% of customers who buy **LUNCH BAG RED RETROSPOT and LUNCH BAG BLACK SKULL.** also buy **LUNCH BAG PINK POLKADOT** (versus ~3.5% of all customers buy **LUNCH BAG PINK POLKADOT** in this context).
8. 35% of customers who buy **LUNCH BAG PINK POLKADOT** also buy **LUNCH BAG RED RETROSPOT and LUNCH BAG BLACK SKULL.** (versus ~2.4% of all customers buy both).
9. 54% of customers who buy **WOODEN PICTURE FRAME WHITE FINISH** also buy **WOODEN FRAME ANTIQUE WHITE** (versus ~6.5% of all customers buy **WOODEN FRAME ANTIQUE WHITE**).
10. 58% of customers who buy **WOODEN FRAME ANTIQUE WHITE** also buy **WOODEN PICTURE FRAME WHITE FINISH** (versus ~7.1% of all customers buy **WOODEN PICTURE FRAME WHITE FINISH**).

**7. Bundle Recommendations**

1. **Bakelike Clock Duo**
   - Items: ALARM CLOCK BAKELIKE GREEN + ALARM CLOCK BAKELIKE RED
   - Rationale: Highest lift (10.49), strong color complementarity
2. **Lunch Bag Variety Quartet**
   - Items: LUNCH BAG RED RETROSPOT + LUNCH BAG CARS BLUE + LUNCH BAG BLACK SKULL. + LUNCH BAG PINK POLKADOT

    - Rationale: Highest lift groups (8.6–8.9), customers buy multiple designs together
3. **White Frame Classic Pair**
    - Items: WOODEN FRAME ANTIQUE WHITE + WOODEN PICTURE FRAME WHITE FINISH
    - Rationale: Strong décor pairing (lift 8.27)
4. **Vintage Christmas Chain Duo**
    - Items: PAPER CHAIN KIT 50'S CHRISTMAS + PAPER CHAIN KIT VINTAGE CHRISTMAS
    - Rationale: Seasonal pairing (lift 8.26)

**8. Expected Sales Lift / Increase**

- **Bakelike Clock Duo**: Lift 10.49 → bundling should increase sales of the second clock by **~950%** (9.5×) among buyers of the first → expected sales increase of **20–40%** in the alarm clock category.
- **Lunch Bag Variety Quartet**: Lift 8.6–8.9 → bundling should increase sales of additional lunch bags by **760–790%** (7.6–7.9×) → expected sales increase of **25–50%** in the lunch bag category.
- **White Frame Classic Pair**: Lift 8.27 → bundling should increase sales of the second frame by **~727%** (7.3×) → expected sales increase of **15–30%** in the frame category.
- **Vintage Christmas Chain Duo**: Lift 8.26 → bundling should increase sales of the second chain by **~726%** (7.3×) → expected sales increase of **20–35%** during holiday periods.

**Overall estimate**: Implementing these bundles could increase average order value by **10–25%** in the affected product categories.

**9. Conclusion**

The analysis successfully identified strong, positive co-purchase patterns using FP-Growth in Weka. The top rules provide clear evidence for bundles that can drive higher order values without price competition.