

# Association Rule Mining on Online Retail Dataset

---

Aimé-232

---

## Introduction

The objective of this project is to explore **association rule mining**, a key data mining technique that uncovers relationships between products in transactional datasets. By analyzing patterns of products bought together, businesses can gain insights into **customer buying behavior**, which can guide strategies for **product bundling, cross-selling, and recommendation systems**.

For this analysis, we are using the **Online Retail Dataset (UK)**, which contains over **500,000 transactions** from a UK-based online retail store. The dataset includes the following fields:

- **InvoiceNo**: Invoice number for each transaction
- **StockCode**: Product code
- **Description**: Product description
- **Quantity**: Number of units purchased
- **InvoiceDate**: Date and time of the transaction
- **UnitPrice**: Price per unit
- **CustomerID**: Unique identifier for each customer
- **Country**: Customer country

Since the dataset does not contain **product categories**, we enhanced it by adding a **ProductCategory** field. This was done using **natural language processing (NLP)** and **clustering techniques**, grouping similar products based on their descriptions. Adding the category field allows us to perform **more generalized and meaningful association analysis**, capturing patterns that might be too sparse at the item level.

## Project Workflow

### 1. Data Cleaning & Preparation

- Remove duplicates and missing values
- Filter transactions to include only positive quantities
- Convert transactional data into a **basket format** suitable for association analysis
- Add the **ProductCategory** column

### 2. Frequent Itemset Mining

- Apply **Apriori** algorithm for item-level and category-level frequent itemsets
- Experiment with different minimum support thresholds
- Visualize the top frequent itemsets

### 3. Association Rules Generation

- Generate association rules using frequent itemsets
- Analyze rules using **support, confidence, and lift**
- Interpret key insights and suggest actionable business applications

# 1. Dataset Description

- **Source:** [Online Retail Dataset](#)
- **Description:** Transactional data for an online retail store, including invoices, stock codes, product descriptions, quantities, prices, invoice dates, and customer IDs.
- **Preparation Steps:**
  1. Removed duplicate transactions and missing values
  2. Filtered transactions to include only positive quantities
  3. Converted transactional data into a basket format
  4. Added the **ProductCategory** field

# 2. Frequent Itemset Mining

## 2.1 Methodology

- Algorithm used: **Apriori** for item-level and category-level mining
- Minimum support thresholds:
  - Item-level: 0.02, 0.05
  - Category-level: 0.05, 0.1

## 2.2 Results

### Top 10 Frequent Itemsets (Item-Level, min\_support=0.02)

Support	Itemsets
0.106	WHITE HANGING HEART T-LIGHT HOLDER
0.092	REGENCY CAKESTAND 3 TIER
0.086	JUMBO BAG RED RETROSPOT
0.074	PARTY BUNTING
0.074	ASSORTED COLOUR BIRD ORNAMENT
0.070	LUNCH BAG RED RETROSPOT
0.062	SET OF 3 CAKE TINS PANTRY DESIGN
0.059	POSTAGE
0.057	LUNCH BAG BLACK SKULL.
0.056	PACK OF 72 RETROSPOT CAKE CASES

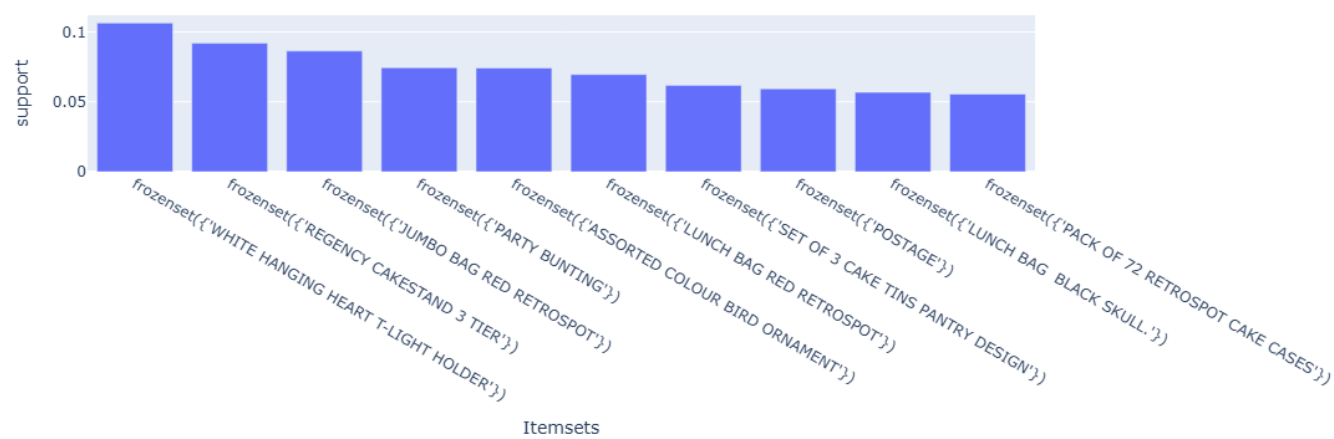
### Top 10 Frequent Itemsets (Category-Level, min\_support=0.05)

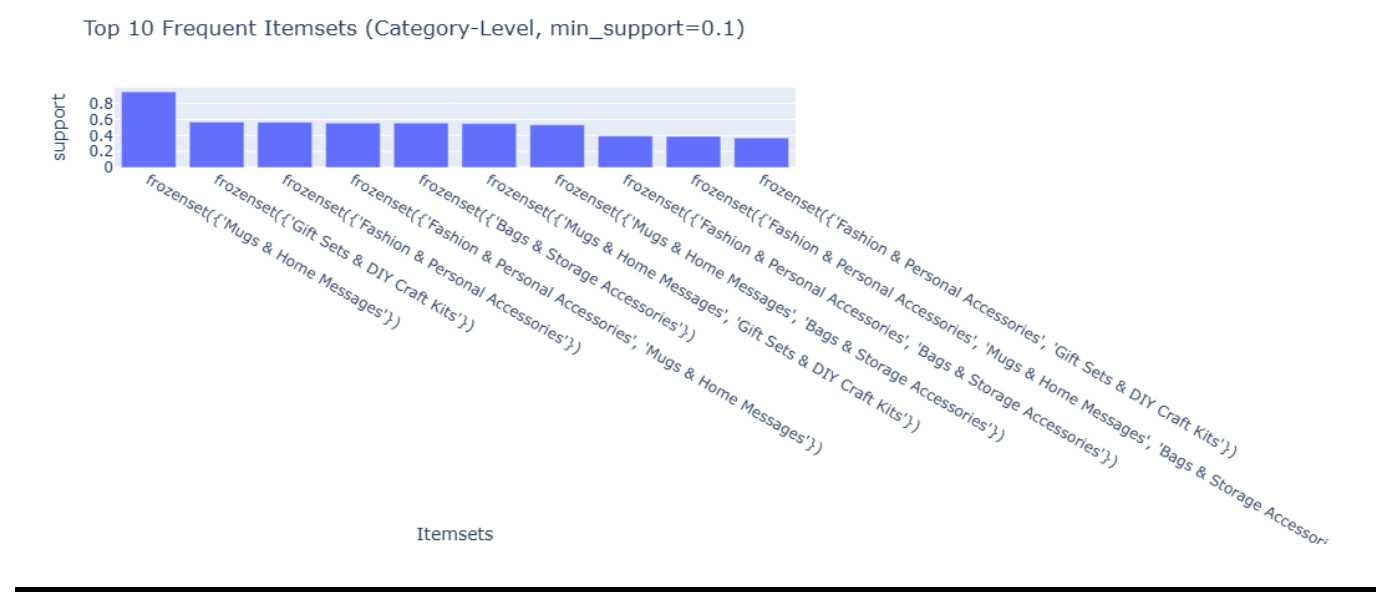
Support	Itemsets
0.946	Mugs & Home Messages
0.568	Gift Sets & DIY Craft Kits
0.566	Fashion & Personal Accessories
0.556	Fashion & Personal Accessories, Mugs & Home Messages
0.555	Bags & Storage Accessories
0.548	Mugs & Home Messages, Gift Sets & DIY Craft Kits
0.533	Mugs & Home Messages, Bags & Storage Accessories
0.394	Fashion & Personal Accessories, Bags & Storage Accessories
0.390	Fashion & Personal Accessories, Mugs & Home Messages
0.370	Fashion & Personal Accessories, Gift Sets & DIY Craft Kits

2.3 Interpretation

- **High Frequency Items:** Most commonly purchased items are small decorative or household goods, reflecting customer preference for affordable, popular products.
- **Category Insights:** Mugs, gift sets, and fashion accessories dominate purchases, which can guide **cross-selling strategies** and **promotion planning**.

Top 10 Frequent Itemsets (Item-Level, min\_support=0.02)





### 3. Association Rules

#### 3.1 Methodology

- Generated association rules using **support**, **confidence**, and **lift** metrics
- Filtered rules to identify strong relationships between product categories

#### 3.2 Top Association Rules (Item-Level)

No rules were generated at the item-level for the chosen support and confidence thresholds.

#### 3.3 Top 10 Association Rules (Category-Level)

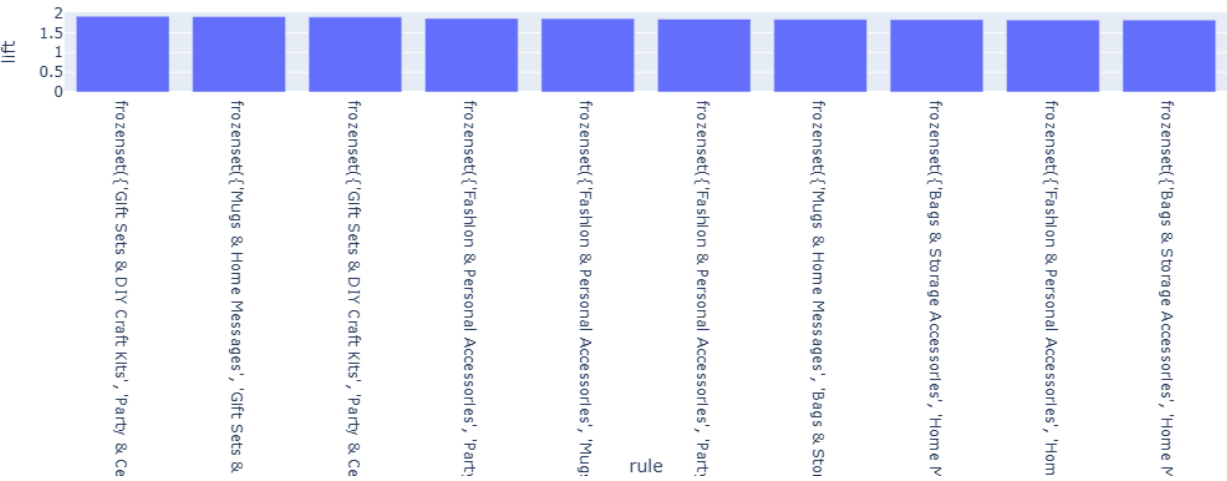
Antecedents	Consequents	Support	Confidence	Lift
(Gift Sets & DIY Craft Kits, Party & Celebration)	(Fashion & Personal Accessories, Mugs & Home Messages)	0.114882	0.746	1.913
(Mugs & Home Messages, Gift Sets & DIY Craft Kits)	(Fashion & Personal Accessories, Bags & Storage Accessories)	0.114882	0.751	1.907
(Gift Sets & DIY Craft Kits, Party & Celebration)	(Fashion & Personal Accessories, Bags & Storage Accessories)	0.114936	0.746	1.894
(Fashion & Personal Accessories, Party & Celebration)	(Bags & Storage Accessories, Mugs & Home Messages)	0.114882	0.649	1.858
(Fashion & Personal Accessories, Mugs & Home Messages)	(Bags & Storage Accessories, Gift Sets & DIY Craft Kits)	0.114882	0.654	1.853
(Fashion & Personal Accessories, Party & Celebration)	(Bags & Storage Accessories, Gift Sets & DIY Craft Kits)	0.114936	0.650	1.840
(Mugs & Home Messages, Bags & Storage Accessories)	(Fashion & Personal Accessories, Gift Sets & DIY Craft Kits)	0.105547	0.677	1.832

Antecedents	Consequents	Support	Confidence	Lift
(Bags & Storage Accessories, Home Metal Decor)	(Fashion & Personal Accessories, Mugs & Home Messages)	0.105547	0.672	1.828
(Fashion & Personal Accessories, Home Metal Decor)	(Bags & Storage Accessories, Mugs & Home Messages)	0.105547	0.636	1.819
(Bags & Storage Accessories, Home Metal Decor)	(Fashion & Personal Accessories, Gift Sets & DIY Craft Kits)	0.105655	0.672	1.819

3.4 Interpretation

- 1. **Rule 1:** Customers buying **Gift Sets & DIY Craft Kits** and **Party & Celebration** items also buy **Fashion & Personal Accessories** and **Mugs & Home Messages**. *Business Insight:* Consider cross-category promotions.
- 2. **Rule 2:** Purchases of **Mugs & Home Messages** with **Gift Sets & DIY Craft Kits** strongly co-occur with **Fashion & Personal Accessories** and **Bags & Storage Accessories**. *Insight:* Suggest these items together for upselling.
- 3. **Rule 3:** **Bags & Storage Accessories** with **Home Metal Decor** are often bought with fashion-related items, indicating complementary shopping behavior. *Insight:* Offer bundles or discounts to encourage combined purchases.

Top 10 Association Rules (Category-Level) by Lift



4. Conclusion

- Frequent itemset mining highlighted the most commonly purchased items and categories.
- Category-level association rules revealed actionable relationships for cross-selling and promotion strategies.
- Insights can guide **inventory planning**, **marketing campaigns**, and **targeted product recommendations** to increase sales and customer satisfaction.