



Groceries Dataset for Market Basket Analysis (MBA)



Author: Badal Singh

Created with Pi

CONTENTS



1. Overview

3. Use Cases

5. How to Use

2. Dataset Description

4. Example Applications

6. Example (Python - Apriori)

A faded, light gray background image of a mountain range, likely the Alps, with snow-capped peaks and forested slopes.

01

Overview



Overview

This dataset contains grocery store transaction records that can be used for Market Basket Analysis (MBA), a common technique in data mining to uncover associations between items frequently purchased together. MBA is widely applied in recommendation systems, cross-selling strategies, and retail analytics.

A faded, light gray background image of a mountain range, likely the Himalayas, with snow-capped peaks and rugged terrain.

02

Dataset Description



Dataset Description

Each record represents a customer's shopping basket.

Items purchased together appear in the same transaction.

The dataset is suitable for applying association rule mining algorithms such as Apriori, FP-Growth, or Eclat.

Key Features

- 1 Transaction ID → Unique ID for each purchase.
- 2 Items → List of grocery items bought in that transaction.





03

Use Cases



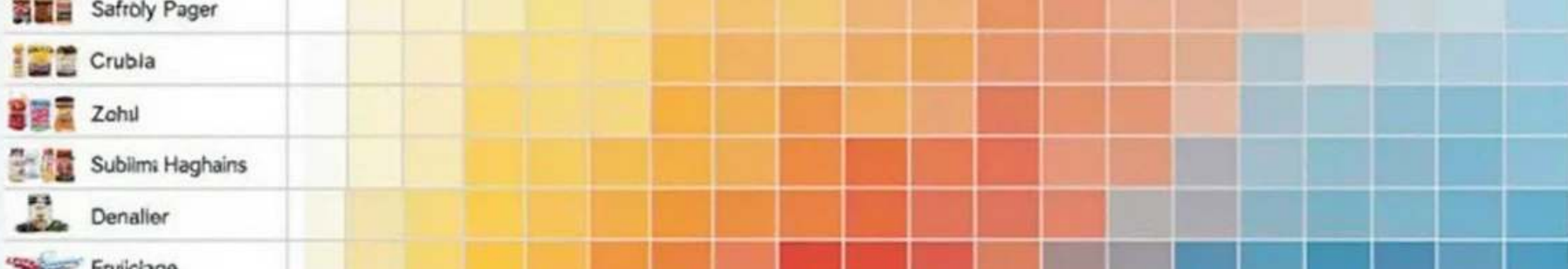
Use Cases

- 1 Identify items frequently bought together.
- 2 Create product bundling and cross-sell opportunities.
- 3 Develop recommendation systems for retail/online stores.
- 4 Analyze shopping patterns and seasonal demand.



04

Example Applications



Example Applications

Generate rules like:

1

{Bread, Butter} → {Jam}

2

{Milk} → {Cookies}

Visualize frequent itemsets with heatmaps, bar charts, or network graphs.

Support decision-making in retail marketing.



05

How to Use



How to Use

Load the dataset using Python (pandas) or R.

Apply association rule mining with libraries such as:

1

mlxtend (Python)

2

arules (R)

Interpret the results using support, confidence, and lift metrics.

06

Example (Python - Apriori)



Example (Python - Apriori)

```
from mlxtend.frequent_patterns import apriori, association_rules
import pandas as pd
/# Load dataset
data = pd.read_csv("groceries.csv")
/# Apply Apriori
frequent_itemsets = apriori(data, min_support=0.02, use_colnames=True)
rules = association_rules(frequent_itemsets, metric="lift", min_threshold=1)
print(rules.head())
```



Thank You