**Association Model for Market Basket Analysis**

**1. Introduction**

Market Basket Analysis (MBA) is a data mining technique used to uncover relationships between products frequently purchased together. It is widely applied in retail for **product placement, cross-selling, and promotional strategies**. This project employs the **Apriori algorithm**, a popular association rule mining method, to analyze transaction data and identify meaningful product associations.

**2. Objective**

The primary objectives of this analysis were:

* To identify **frequently co-purchased products** in a transactional dataset.
* To generate **association rules** (e.g., "If Product A is bought, then Product B is likely bought").
* To measure the strength of these associations using **support, confidence, and lift**.
* To derive actionable insights for **retail strategy optimization**.

**3. Dataset Link**

<https://drive.google.com/file/d/1vWBogbxSCUGk01ZPLTCaYZqkJu1-2taq/view?usp=sharing>

**4. Understanding the Data**

* **Source**: Transaction records (MBA.csv) with:
  + **Rows**: **72 entries** (individual product purchases).
  + **Columns**: **2** (ID for transaction identifier, Products for item names).
* **Key Details**:
  + **15 unique transactions** (IDs: 1–15).
  + **15 unique products** (labeled A–O).

**5. Code Link**

<https://github.com/Ishita2003M/Association-Model-for-Market-Basket-Analysis/blob/main/Association%20Model%20for%20Market%20Basket%20Analysis.docx>

**6. Procedure for Coding (Steps Followed)**

* Loaded Libraries and Data
* Converted Data to Transaction Format
* Generated a summary of transactions (e.g., item frequency, transaction lengths).
* Plotted top 5 most frequent items
* Extracted association rules with thresholds:
  + Support = 0.005 (minimum frequency of item sets).
  + Confidence = 0.8 (minimum likelihood of rule being true).
* Evaluated Top Rules by Lift

**7. Interpretation and Conclusion**

**Key Findings**

1. **Top Association Rules**:
   * All top rules involved **Product D** as the consequent (RHS), showing it was frequently bought with:
     + **Product N** (lift = 5, confidence = 100%).
     + Combinations like {Product G, N}, {Product E, N}, etc.
   * **Interpretation**:
     + Product N strongly predicts Product D (e.g., a complementary or bundled product).
2. **Item Frequency**:
   * **Product I** had the highest occurrence (73%), making it a staple item.
   * Products A, C, F, H were also popular (~47% each).
3. **Statistical Metrics**:
   * **High confidence (1.0)**: Rules were always true in the observed data.
   * **High lift (5)**: Co-occurrence of these items was 5x more likely than random chance.

**Business Implications**

* **Cross-Selling**: Promote Product D alongside Product N (e.g., discounts or placements).
* **Inventory Management**: Ensure high stock for frequent items (I, A, C, F, H).
* **Bundling**: Create bundles for high-lift pairs (e.g., "Buy N, get D at 10% off").

**Conclusion**

This analysis revealed **strong associations between Product N and Product D**, alongside identifying high-frequency items. Retailers can leverage these insights to optimize product placement, promotions, and inventory strategies. Future work should expand the dataset for more robust rule discovery.