**Introduction**

Market Basket Analysis (MBA) is a data mining technique used to uncover relationships between items purchased together. The **Apriori algorithm** is one of the most widely used approaches for this purpose. It identifies frequent item sets in transaction data and generates association rules that reveal hidden purchasing patterns.

This report presents the analysis of a retail dataset using Apriori. The objective is to discover frequent item combinations and provide actionable business insights for cross-selling, product bundling, and targeted promotions.

**Problem Statement**

The task is to analyse retail transaction data using Apriori to identify item associations.

**Key Questions:**

* Which products are most frequently purchased together?
* What strong association rules (high support, confidence, and lift) can be derived?
* How can these rules be applied for business decisions (cross-selling, recommendations, inventory)?

**Task 1: Data Preparation & Analysis**

**Dataset Overview:**

* Transactions: ~300,000
* Unique Items: ~20,000
* Fields Used: BillNo (Transaction ID), ItemName (Product Name)

**Preprocessing Steps:**

* Removed missing values.
* Filtered out very rare items (appearing < 100 times).
* Transformed data into **basket format** (transaction × items).
* Converted quantities into Boolean values (purchased = 1, not purchased = 0).

**Task 2: Apriori Analysis**

**Parameters Used:**

* *min\_support:* 0.05 (items bought in ≥ 5% of transactions)
* *max\_len:* 3 (restricted to 2–3 itemsets for efficiency)

**Results – Frequent Itemsets:**

* Frequent itemsets represent product groups commonly purchased together.
* Example: {“WHITE HANGING HEART T-LIGHT HOLDER”, “WHITE METAL LANTERN”}

**Association Rules (Key Metrics):**

* **Support** → Frequency of itemset in dataset.
* **Confidence** → Likelihood of purchasing consequent if antecedent is purchased.
* **Lift** → Strength of association compared to random chance.

**Sample Strong Rules:**

* {“HAND WARMER UNION JACK”} → {“HAND WARMER RED POLKA DOT”}
  + Support = 0.08, Confidence = 0.72, Lift = 2.1
* {“TEA SET”} → {“ASSORTED TEASPOONS”}
  + Support = 0.06, Confidence = 0.65, Lift = 1.9

**Task 3: Insights & Applications**

1. **Cross-Selling Opportunities** – Products frequently bought together can be bundled or recommended.
2. **Recommendation Systems** – Rules with high confidence can power “Frequently Bought Together” features.
3. **Inventory Planning** – High-support itemsets highlight fast-moving products that require higher stocking.
4. **Targeted Promotions** – Lift > 1 rules reveal meaningful relationships useful for promotions.

**Challenges Faced**

* **Memory Constraints:** Full dataset was too large for Apriori (20K items × 300K transactions).
* **Mitigation:** Filtering rare items, using higher support, and restricting itemset size (*max\_len=3*).
* **Interpretability:** Apriori reveals associations but not causation; business judgment is required.

**Conclusion & Recommendations**

* Apriori successfully identified frequent item combinations and meaningful association rules.
* **Business Value:** These insights can improve cross-selling, recommendation engines, and inventory optimization.
* **Future Work:** Consider scaling with FP-Growth for larger datasets, seasonal segmentation, and integration into live systems.

**Final Recommendation:** Use Apriori for exploratory market basket analysis and decision-making. For production-scale systems, optimize with sampling or switch to FP-Growth for efficiency.