MARKET BASKET ANALYSIS FOR A MID-SIZED FASHION E-COMMERCE STORE

1. Background and Problem Statement:

A growing fashion e-commerce brand offering apparel, footwear, and accessories noticed low average order values (AOV) and inconsistent cross-category purchases despite a broad catalog. Customers often bought single items, missing out on complementary products. The marketing team lacked actionable insights on product pairings to inform bundle offers or upselling campaigns. To address this, the brand initiated a **market basket analysis project** to identify high-affinity product combinations and design data-driven product bundles to increase revenue per transaction.

2. Objectives:

- To identify which products are most frequently bought together using transaction data

- To calculate support, confidence, and lift values for product pairings

- To recommend cross-sell and upsell bundles based on customer behavior

- To develop strategic suggestions for combo offers and onsite product placement

- To raise AOV through targeted recommendations and dynamic bundling

3. Methodology:

Data Collection and Preparation:

- Transaction logs of 48,000 completed orders over the past 6 months

- Data included product ID, category, subcategory, brand, and quantity per basket

- Cleaned and transformed into a binary transaction matrix (1 = purchased, 0 = not purchased)

Analytical Techniques:

- **Market Basket Analysis (MBA)** using the Apriori algorithm

- Metrics calculated:

  - **Support** = proportion of baskets containing the item pair

  - **Confidence** = probability of buying B given A

- o **Lift** = strength of association (Lift > 1 implies positive correlation)
- Tool used: **Python** with mlxtend for association rule mining
- Additional visualization in **Excel and Tableau**

4. Results:

- Identified over 120 strong product pair rules (Lift > 1.2, Confidence > 0.5)
- **Top Pairings:**
    - o **"Ankle Boots" + "Long Coats"** → Lift: 1.8, Confidence: 0.62
    - o **"Dresses" + "Belts"** → Lift: 2.1, Confidence: 0.59
    - o **"Sneakers" + "No-Show Socks"** → Lift: 2.3, Confidence: 0.71
- **Triplet Pattern:** "T-shirts + Joggers + Baseball Cap" occurred in 7.4% of male orders
- Segment-wise insight:
    - o Women's accessories had the highest cross-category lift
    - o Return customers showed higher adoption of suggested combinations vs. first-time buyers

5. Interpretation and Insights:

- Customers are already exhibiting logical bundling behavior, but not all patterns are promoted
- Many complementary products (e.g., belts) were not displayed on PDPs (Product Detail Pages)
- Seasonal patterns influenced pairing (e.g., boots + coats in winter)
- Product combinations with high lift and low conversion represent **untapped bundling opportunities**

6. Recommendations:

- Implement **"Frequently Bought Together" widgets** for top 15 pairings
- Launch combo discounts (e.g., "Buy Dress + Belt and get 10% off")
- Update site merchandising logic to auto-show relevant accessories on product pages

- Use dynamic bundling in email marketing (based on last order SKU)

- Run A/B tests to compare manual vs. algorithmic bundles in conversion performance

7. Future Work:

- Extend analysis with time-based association rules (weekday vs. weekend buying behavior)

- Integrate real-time bundling suggestions into checkout and cart pages

- Add demographic segmentation to personalize pairing suggestions

- Connect this model with ad targeting logic for cross-sell ad campaigns

8. Stakeholder Relevance:

Academic:

- Demonstrates practical application of association rule mining in retail

- Suitable for courses in retail analytics, consumer behavior, and business intelligence

Corporate:

- Enables online retailers to optimize cross-selling strategies and increase average order value

- Provides a replicable framework to analyze product affinity and guide merchandising decisions