CLUSTER-BASED PRODUCT CATEGORIZATION USING EXCEL FOR INVENTORY OPTIMIZATION

1. Background and Problem Statement:

A mid-sized fashion e-commerce company with over 2,000 SKUs faced difficulty in identifying which products performed similarly in terms of sales, returns, and stock movements. Their manual product grouping (e.g., "summer tops," "winter coats") failed to reflect real performance behavior. They sought a way to use unsupervised data mining in Excel to cluster products into logical segments, helping them improve reorder decisions, clearance strategies, and promotion targeting.

2. Objectives:

- To perform unsupervised clustering on product-level data using Excel

- To define data-driven product categories based on sales, returns, and stock turnover

- To build an interactive dashboard for category-wise comparison and optimization

- To support procurement and merchandising teams in targeting fast- and slow-moving SKUs

3. Methodology:

3.1 Data Preparation:

- Dataset: 24-month sales data for 2,100 SKUs

- Variables used:

    o Monthly units sold

    o Average selling price

    o Return rate (%)

    o Stock turnover ratio

    o Shelf time (average number of days in inventory)

- Cleaning: Used Excel functions (e.g., IFERROR, TRIM, ISBLANK) to handle missing and invalid data

3.2 Normalization:

- Applied min-max normalization using Excel formula: =(Value - Min(Range)) / (Max(Range) - Min(Range))

3.3 Clustering:

- Used 4 key variables (sales, return rate, turnover, shelf time) for grouping

- Implemented K-means-like clustering in Excel:

  - Assigned initial centroids manually for 4 clusters

  - Computed Euclidean distances between products and centroids

  - Iterated manually using Excel's cell-based logic for 5 rounds

  - Assigned final clusters using lowest distance logic

3.4 Dashboard Development:

- Created PivotTables linked to the clustered data

- Used slicers to filter by cluster, season, product category

- Visualized metrics with charts:

  - Cluster-wise average sales

  - Return rate by cluster

  - Inventory holding trends

4. Key Features of the Excel Model:

- Custom cluster assignment and label for each product

- Cluster-level summary of performance metrics

- Slicers for real-time cluster filtering

- Conditional formatting to flag low-turnover or high-return SKUs

- Built-in dropdown to simulate new SKU performance and see cluster allocation

5. Results and Benefits:

- Products grouped into 4 performance-based clusters:

  - Cluster A: High sales, low returns (Keep in stock)

  - Cluster B: High returns, long shelf time (Reduce promotions)

- o Cluster C: Low sales, low turnover (Target for clearance)

  o Cluster D: New/inconclusive patterns (Monitor)

- Helped reduce inventory holding cost by 12%

- Improved clearance planning by focusing discounts on Cluster C

- Enabled sales team to push high-performing products in Cluster A

6. Deliverables:

- Excel workbook with cluster model, raw and processed data sheets

- Dashboard sheet with slicers, charts, and filters

- Documentation explaining clustering logic and update instructions

- Slide deck summarizing insights for internal teams

7. Stakeholder Relevance:

Academic:

- Useful for teaching unsupervised learning, clustering logic without needing advanced software

- Applied use case of data mining with Excel formulas

Corporate:

- Applicable for retail businesses with large SKU portfolios

- No-code solution for product analytics and inventory decision-making