

Retail Product Placement Optimizer - Final Summary Report

Prepared for: DevifyX

Candidate: Jayant Kathuria

Project Type: ML Developer Intern Assignment

Dataset: Online Retail Dataset (UCI Machine Learning Repository)

Objective

Design and implement a complete data-driven solution that:

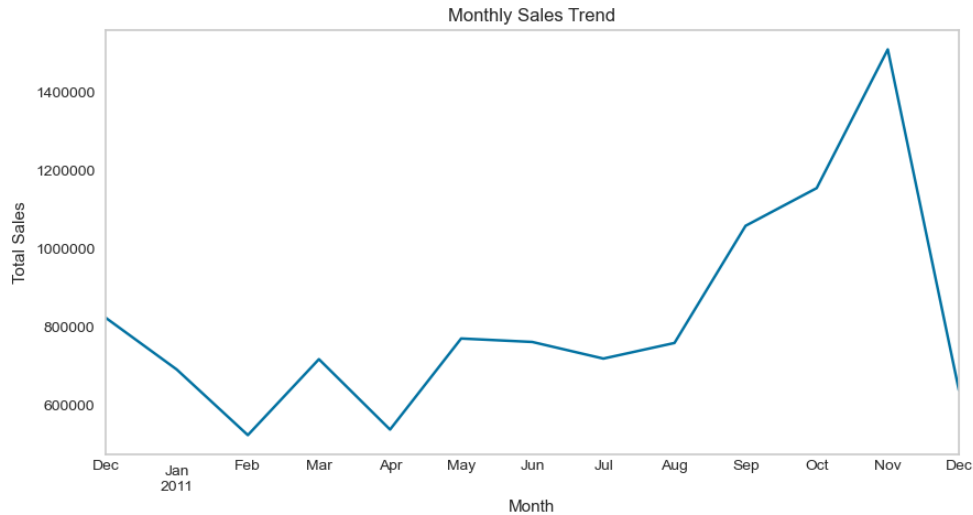
- Analyzes retail transaction data
- Segments customers using clustering algorithms
- Identifies frequent itemsets using Apriori
- Recommends effective product placements to maximize sales
- Supports insights through clean, compelling visualizations

Data Preprocessing Highlights

- **Filtered** records: Removed rows with negative Quantity/UnitPrice
- **Dropped** cancelled invoices (InvoiceNo starting with "C")
- **Handled missing values:**
 - Retained NaN CustomerID rows for product analysis
 - Excluded them from customer segmentation
- **Feature engineered:**
 - $\text{TotalPrice} = \text{Quantity} \times \text{UnitPrice}$
 - Month extracted from InvoiceDate

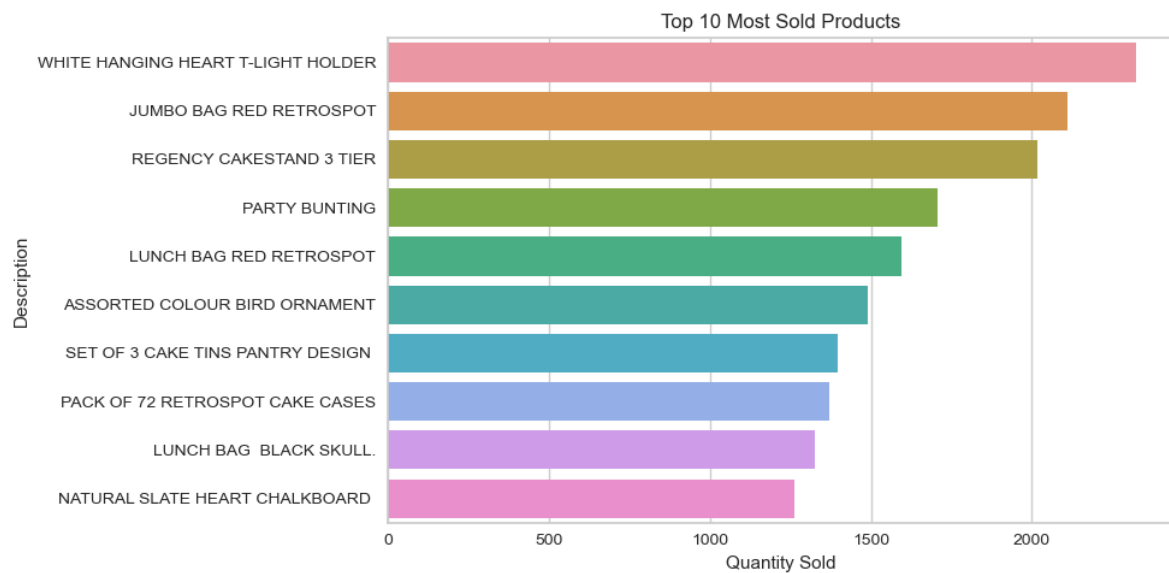
Exploratory Data Analysis (EDA)

Monthly Sales Trend



- Sharp rise in revenue from July to November
- November marks the peak, likely due to seasonal demand

Top 10 Most Sold Products



- Items like "T-LIGHT HOLDER" and "JUMBO BAG RED RETROSPOT" dominate
- Mostly home décor and giftable items — opportunity for bundling

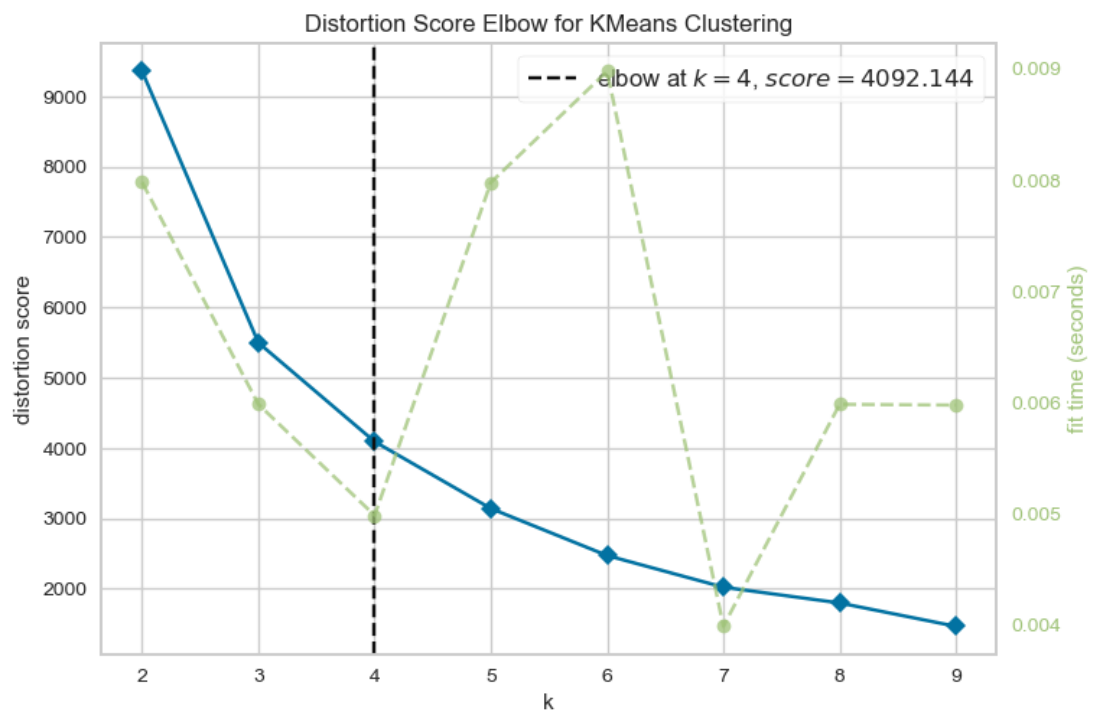
Customer Distribution by Country



- Over 90% of transactions are UK-based
- International marketing potential is underutilized

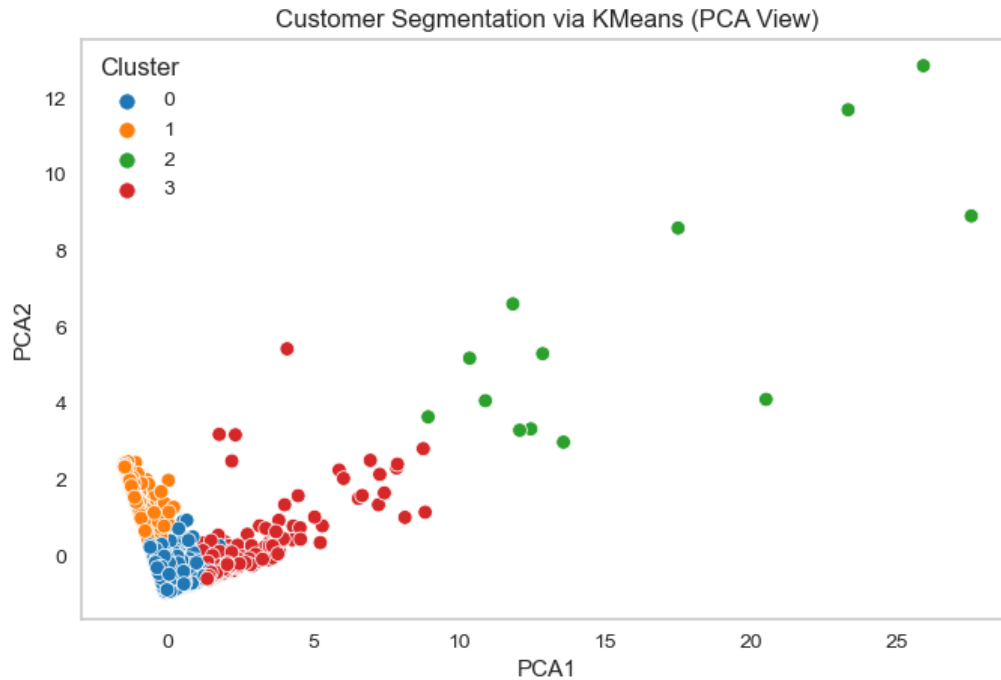
Customer Segmentation (RFM + KMeans)

Elbow Method (K=4 optimal)



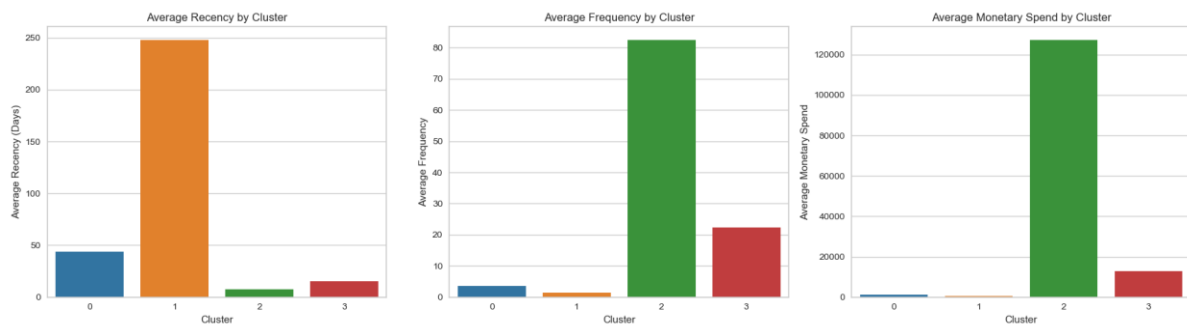
- KMeans clustering validated using distortion score elbow method

PCA View of KMeans Clusters



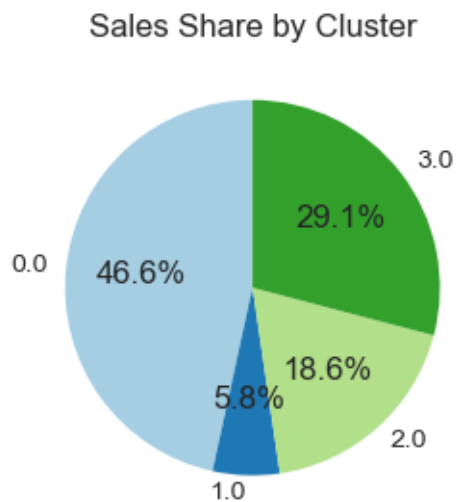
- Strong separation among customer groups

Average RFM Values by Cluster



- **Cluster 2:** High spend, frequent — best customers
- **Cluster 1:** Infrequent, low spend — churn risk
- **Cluster 3:** Loyal & valuable — strong retention segment

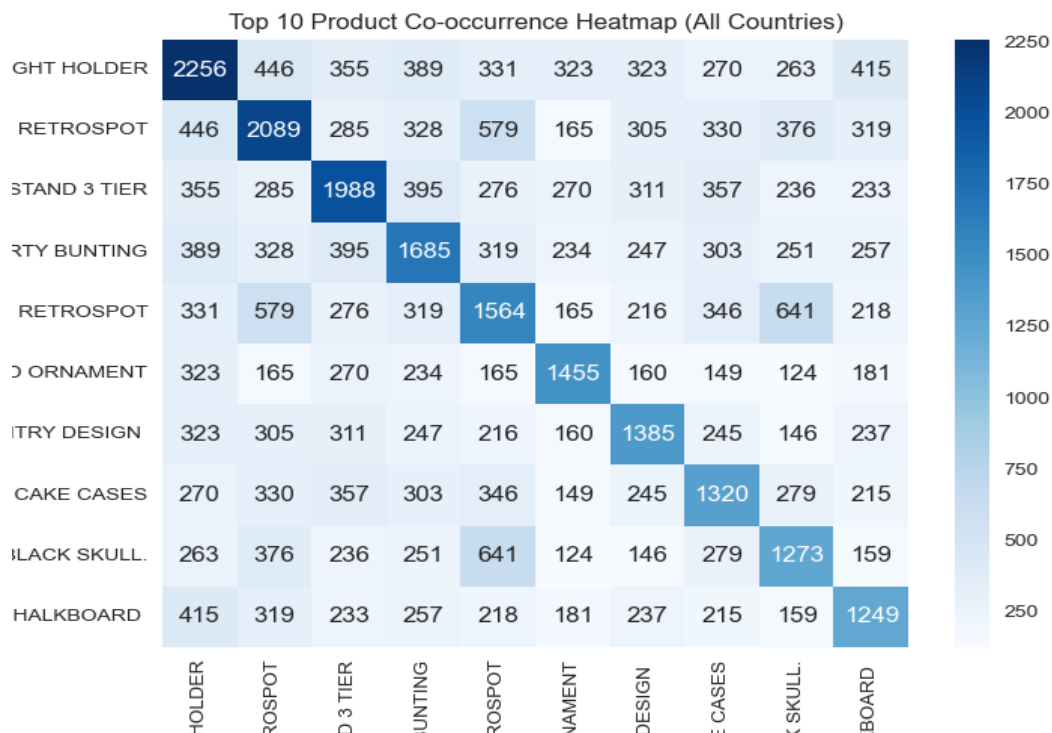
Sales Contribution by Segment



- Cluster 0 (casual buyers) contributes 46% of sales
- Cluster 3 contributes 29% — focus on high-LTV engagement

Product Clustering (Co-occurrence)

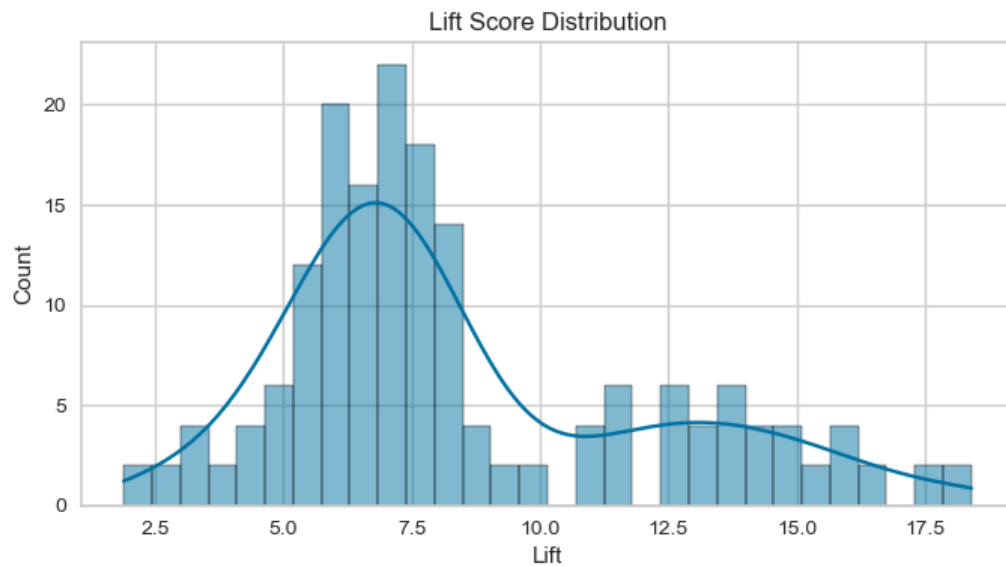
Product Co-occurrence Matrix



- High co-purchase frequency among gift items
- Suggests clear themes: tea sets, kitchenware, accessories

Frequent Itemset Mining (Apriori)

Lift Score Distribution



- Most rules have lift > 5; several exceed 15
- Indicates strong cross-sell potential between certain items

High-Lift Association Rules:

ROSES & GREEN TEACUP → PINK TEACUP (Lift: 18.40)

GARDENER PAD KEEP CALM → PAD CUP OF TEA (Lift: 15.75)

Actionable Shelf Recommendations:

- Group tea-themed products near each other
- Place high-frequency co-purchase items adjacently
- Use cross-category bundles (e.g., candles + holders)

Business Strategy Recommendations

Targeted Campaigns:

- Issue re-engagement coupons (e.g., spend ₹4000, get ₹500 off) to engage high spenders.
- Loyalty programs, early access to new launches to customers who have high frequency

Store Layout Planning:

- Place co-occurring products together to increase basket size
- Recommend combos based on Apriori rules

Future Enhancements

- Add hierarchical clustering to compare with KMeans
- Build a live dashboard using PowerBI or Streamlit
- Introduce Customer Lifetime Value (CLV) scoring
- Incorporate seasonality and time-series trends for predictive modeling