# Intelligent Bundle Generation for Personalized Promotion Optimization in Retail


**VINOD KUMAR**
**MA23M026**

**Guide:** Prof. Satyajit Roy
Department of Mathematics , IIT Madras

**Co-Guide:** Dr. Sharadha Ramanan
TCS Research and Innovation


MA5990 - M.Tech Project
19 May 2025


Indian Institute of Technology Madras

# Introduction & Modern Retail

- Traditional bundles (e.g., shampoo + conditioner) offer the same set to all customers.
- Generic promotions lead to reduced engagement and lower conversions.
- Today's consumers prefer bundles tailored to their needs and preferences.
- e.g., A health-conscious customer prefers granola + almond milk, while a parent might seek kid-friendly snacks.
- Personalization is key to improving customer satisfaction and driving sales.

# Literature & Traditional Strategies

- **Traditional Strategies:**
  - Category-level offers (e.g., BOGO).
  - Loyalty or segment-based discounts.
  - Fixed discounting — lacks personalization.
  - Rule-based bundling — same combos for all.
  - Generic campaigns — mass coupons, flat discounts.
  - Consequences: low engagement, poor ROI.
- **Literature Insights:**
  - Elasticity estimation using linear/log-log models.
  - Shift towards ML, DL, and optimization models.
  - **Sun & Li:** Discriminative vs. generative bundling.
  - **Sun & Yang:** Behavioral promotion alignment.
  - **Pathak & Gupta:** Collaborative filtering for bundles.
  - **Zhu & Harrington:** Bundle Recommendation Problem.
  - **Uma G et al.:** MINLP with cross-elasticity.

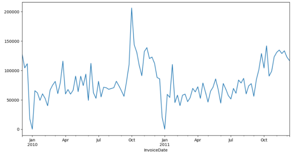
# Objectives

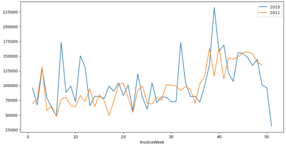
- Analyze historical sales and behavior
- Discover product affinities
- Generate optimal bundles
- Personalize recommendations
- Optimize promotional strategies

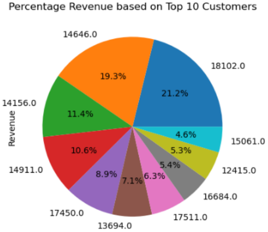

*Traditionally static bundles (shampoo + conditioner) are replaced by AI-powered, personalized bundles to enhance experience, loyalty & revenue.*

# EDA and Data Preprocessing

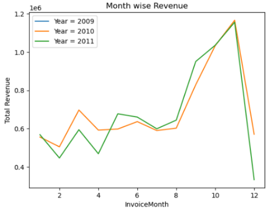
- **EDA:** Trends, correlations, outlier and missing value handling.
- **Preprocessing:** Feature engineering, encoding, normalization.

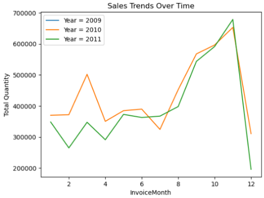


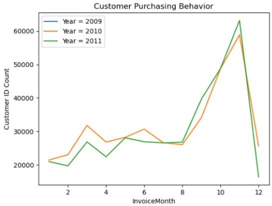



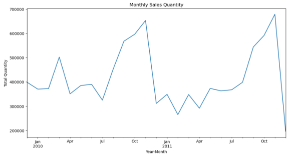



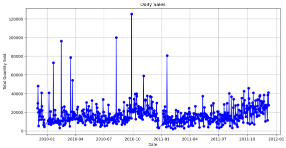
# EDA and Preprocessing

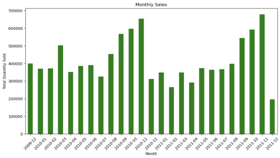


# Customer Segmentation

- **Why Segment?** To offer personalized promotions based on customer behavior and value.
- **RFM Segmentation:** Classifies customers by Recency, Frequency, and Monetary value.
  *eg, Target high-frequency, high-spend users for exclusive bundles.*
- **Behavioral Segmentation:** Groups users by transaction patterns and preferences.
  *Example: Identify health-conscious vs. convenience-driven customers.*
- **Clustering Techniques:**
  - **k-means:** Efficient for large datasets, assumes spherical clusters.
  - **DBSCAN:** Detects dense regions, ideal for arbitrary-shaped clusters.
  - **Hierarchical:** Builds a tree of clusters, useful for small datasets.
- **Outcome:** Better targeting, higher campaign ROI (Return on Investment), and improved customer satisfaction.

# Interrelated Items

- Understand relationships between products — are they substitutes or complements?
- **Substitutes:** Items that can replace each other.
  Eg: Coke (Stock out) then Pepsi (Demand up).
- **Complements:** Items bought together.
  Eg: if coffee sales rise, sugar sales may also increase.
- **Substitution Matrix:**
  - Constructed from co-purchase and switching behavior.
  - Value at (i, j) = ratio of customers who switched from product $i$ to $j$.
- **Simple Example:**
  - 100 customers bought Coke in month 1; 20 switched to Pepsi in month 2 $\rightarrow$ Substitution ratio = 20
- **Visualization:**
  - Use heatmaps or graphs to cluster related products.
  - Darker cells in heatmap = stronger relationships (substitution or complementarity).

# Price Elasticity Modeling

- $PriceElasticity(E\_ii) = \frac{Percentage\_Change\_in\_Demand\_of\_item\_i}{Percentage\_Change\_in\_Price\_of\_item\_i}$
- We use a **log-log regression model** to estimate the own price elasticity of demand.
- The model takes the form:

$$\log(Q) = \beta_0 + \beta_1 \log(P) + \varepsilon$$

  where:
  - $Q$ = Quantity sold
  - $P$ = Price of the product
  - $\beta_1$ = Own price elasticity (interpreted directly from the model)
- **Interpretation:** If $\beta_1 = -1.5$, a 1% increase in price leads to a 1.5% decrease in quantity sold.
- Eg: For a soft drink brand:
  - Regression output: $\log(Q) = 5.2 - 1.8 \log(P)$
  - Interpretation: The demand is price elastic. A 10% increase in price is expected to reduce sales by 18%.
- Such elasticity estimates help optimize pricing strategy to maximize revenue or profit.

# Cross Elasticity Modeling

- **Cross-price elasticity** measures how the demand for one product responds to a price change in another:

$$E_{ij} = \frac{\%\ \Delta Q_j}{\%\ \Delta P_i}$$

  where:
  - $E_{ij}$: Elasticity of demand for product $j$ with respect to price of product $i$
  - $Q_j$: Quantity sold of product $j$
  - $P_i$: Price of product $i$
- **Interpretation:**
  - $E_{ij} > 0$: Products $i$ and $j$ are substitutes (e.g., Pepsi and Coke)
  - $E_{ij} < 0$: Products $i$ and $j$ are complements (e.g., printers and ink)
- **Eg:** A 5% increase in the price of Coke leads to a 3% increase in sales of Pepsi $\rightarrow E_{\text{Coke,Pepsi}} = \frac{+3\%}{+5\%} = +0.6$
- **Construct a cross-elasticity matrix** across products or categories to detect:
  - Cannibalization within your brand
  - Complementary relationships that boost total sales

# Demand Forecasting Model & Pipeline

- Predict future sales for a product/store over a time horizon.
- **Input Features:**Price, promotions, holidays, day of week, and past sales (lags, rolling averages).
- **Models Used:**
  - **Statistical:** SARIMA, Holt-Winters — good for capturing seasonality and trends.
  - **Machine Learning:** Random Forest, XGBoost — model non-linear effects and interactions.
  - **Deep Learning:** RNN, LSTM — handle temporal dependencies and multiple covariates.
- **Output:**
  - Multi-step sales forecast $\hat{y}_{t,i}$, where $t$ = time step, $i$ = item/store.
  - Example: Predict next 14 days of sales for SKU123 at Store A.
- **Evaluation Metrics:** RMSE (Root Mean Square Error), MAPE (Mean Absolute Percentage Error).

Helped in inventory planning, pricing, and promotion decisions.

# Methodology and Mathematical Formulation

- Generate personalized product bundles and optimize promotions to maximize revenue.
- **Formulation:** Mixed-Integer Nonlinear Programming (MINLP).
- **Key Constraints:** Budget, forecasted demand, minimum bundle size.
- **Decision Variables:**
  - $x_i$: Include product $i$ in bundle (binary).
  - $\delta[i]$, $P[i]$, $P_d[i]$: Discount rate, regular price, and discounted price.
  - $y_i$: Quantity of product $i$.
- **Demand Modeling with Cross-Elasticity:**

$$D_i = x_i \cdot \hat{D}_i \cdot \left(1 + \sum_{j \neq i} \epsilon_{ij} \cdot \frac{P_j - P_i}{P_i}\right)$$

  - Captures influence of other product prices on $D_i$.
  - $\epsilon_{ij}$: Cross-price elasticity between product $i$ and $j$.
- **Goal:** Satisfy minimum revenue $R_{\text{min}}$, bundle size $M_{\text{min}}$, and cost constraints $B$.

# Objective Function – Maximizing Revenue

- Maximize the total revenue from the selected product bundle.
- **Revenue Formula:**

$$R_B = \sum_{i \in L} P[i] \cdot \left( x_i \cdot \hat{D}_i \cdot \left( 1 + \sum_{j \neq i} \epsilon_{ij} \cdot \frac{P[j] - P_d[i]}{P_d[i]} \right) \cdot y_i \right)$$

- Where:
  - $x_i$: whether product $i$ is selected.
  - $\hat{D}_i$: base demand for product $i$.
  - $\epsilon_{ij}$: how demand for $i$ changes with price of $j$ (cross elasticity).
  - $P_d[i] = P[i] \cdot (1 - \delta[i])$: discounted price of product $i$.
  - $y_i$: number of units of product $i$ in the bundle.
- Why Nonlinear?

  Demand depends on other products' prices (interaction between products). Cross-elasticity introduces complexity in how price changes affect demand and revenue.
- Suppose Product A is discounted, and Product B is a substitute. Lowering price of A may reduce demand for B. The model adjusts both demand and revenue based on these interactions.

# Constraints in Bundle Optimization

- **1. Budget Constraint:** $\sum_{i \in L} C_i \cdot x_i \leq B$
  Total cost of selected products must stay within the available budget.
- **2. Demand Limitation:** $x_i \cdot y_i \leq \hat{D}_i \quad \forall i$
  Can't include more units than forecasted demand.
- **3. Revenue Requirement:** $\sum_{i \in L} P_d[i] \cdot y_i \geq R_{\text{min}}$
  Ensure total revenue from the bundle meets a minimum threshold.
- **4. Decision Variable Constraints:** $x_i \in \{0, 1\}, \quad 0 \leq y_i \leq k \quad \forall i$
  $x_i$ indicates selection (1 or 0), and $y_i$ must be a valid non-negative quantity.
- **5. Bundle Size Constraint:** $\sum_{i \in L} x_i \geq M_{\text{min}}$
  At least $M_{\text{min}}$ products must be selected in the bundle.
- **6. Discount Limits:** $0 \leq \delta[i] \leq \delta_{\text{max}} \quad \forall i$
  Discount on any product must be within allowable bounds.

# Full MINLP Model

**Algorithm 1:** Intelligent Bundle Generation with Optimal Discounting using MINLP Framework

**Input** : ProductList, Total Budget (B), Minimum Revenue Threshold (R_min), Minimum Bundle Size (M_min), Max Discount $\delta_{max}$

**Variables:** Binary selection variable ($x[i]$),
Quantity variable ($y[i]$),
Discount rate ($\delta[i]$),
Discounted price ($P_d[i] = P[i] \cdot (1 - \delta[i])$),
Demand ($D[i]$)

**Output** : BestBundle, Quantities, Discounts, ExpectedRevenue

**for** *each product i in ProductList* **do**
- $x[i] \in \{0, 1\}$
- $y[i] \geq 0$
- $0 \leq \delta[i] \leq \delta_{max}$
- $P_d[i] = P[i] \cdot (1 - \delta[i])$
- **if** $x[i] == 1$ **then**
  - $D[i] = \hat{D[i]} \cdot \left(1 + \sum_{j \neq i} \epsilon[i][j] \cdot \frac{P_d[j] - P_d[i]}{P_d[i]}\right)$
- **else**
  - $D[i] = 0$

Define objective:
Maximize total revenue: $R\_B = \sum_i P_d[i] \cdot D[i] \cdot y[i]$

**Subject to:**
1. Budget constraint: $\sum_i C[i] \cdot x[i] \leq B$
2. Demand limitation: $x[i] \cdot y[i] \leq D[i] \quad , \forall i$
3. Revenue requirement: $\sum_i P_d[i] \cdot y[i] \geq R\_min$
4. Bundle size constraint: $\sum_i x[i] \geq M\_min$
5. Discount bounds: $0 \leq \delta[i] \leq \delta_{max}$
6. Binary and non-negativity constraints: $x[i] \in \{0, 1\}, \quad y[i] \geq 0$

**return** *BestBundle* $= \{i \mid x[i] = 1\}$,
*Quantities* $= y[i]$ *for each product in BestBundle,*
*Discounts* $= \delta[i]$ *for each product in BestBundle,*
*ExpectedRevenue* $= R_B$

# Solution Approach

- **Challenges:** Mixed-Integer Nonlinear Programming (MINLP)
  - Binary decisions (include/exclude products)
  - Nonlinear demand (due to cross-price elasticities)
- **Solution Techniques:**
  - **Branch and Bound / Branch and Cut:**
    For handling binary variables $x_i$
  - **Nonlinear Solvers (APOPT, IPOPT, KNITRO):**
    For solving the nonlinear demand equations
  - **Heuristics**
    Genetic Algorithms, Simulated Annealing for large or complex problems where exact solutions are slow
- **How It Works:**
  - Iterates over possible bundles
  - Considers product selection, price effects, and demand interactions
  - Finds the bundle that maximizes revenue while meeting all constraints

# Implementation and Data Source

**Implementation Details:**

- Implemented in **Python** using the **GEKKO optimization suite**. Uses the **APOPT solver** to solve the MINLP problem. Suitable for nonlinear constraints and integer decision variables

**Data Source:**

- Due to lack of real retail data, a **synthetic dataset** was created based on realistic retail patterns
- **Components Simulated: Prices and Costs:** Based on average market prices and retailer markups. **Demand Forecasts:** Synthetic time-series with seasonal and promotional effects. **Cross Elasticities:** Modeled using assumed substitution and complementarity behavior.

**Assumptions Made:**

- Fixed product categories and price ranges
- Seasonal demand variations (e.g., holidays, weekends)
- Elasticity matrix to capture realistic product interactions
- Budget and promotion limits at category/store level

# Experiments and Results

**Data Exploration:**

- Conducted extensive **EDA** on the simulated retail dataset
- Generated 10+ graphs to understand: Customer behavior patterns, Product reorder frequency, Purchase trends over time.

**Customer Segmentation:**

- Applied **clustering techniques** to divide customers into 4 segments
- Segmentation based on:
  Ordering frequency, Product preferences, Reordering behavior.

**Segment-wise Optimization:**

- Used the optimization method **individually on each customer segment and each store level**
- Improved accuracy and effectiveness compared to a combined approach

*"Segment-specific modeling revealed personalization in retail price promo optimization"*

# Results & Codes (Click here for codes)

OPTIMIZATION RESULTS: ITEM-LEVEL OVERVIEW

| Product Name | Selected | Discount % | Revenue |
|---|---|---|---|
| Whole Wheat Loaf | Selected | 22.04% | $689.83 |
| Bundle of 4 Organic Eggs + 4 Oatmeal Packs | Selected | 44.04% | $753.77 |
| Bundle of 2 Cheddar Blocks + 2 Gra-nola Boxes | Not Selected | 0.0% | $118.23 |
| Bundle of 3 Salted Butter + 4 Greek Yogurts | Selected | 29.98% | $694.04 |
| Bundle of 4 Herb Butter + 1 Baguette | Not Selected | 0.0% | $71.64 |
| Bundle of 5 Orange Juices + 5 Almond Milks | Selected | 38.77% | $667.81 |
| Bundle of 3 Cornflakes + 2 Soy Milks | Selected | 20.61% | $623.26 |
| Bundle of 3 Vanilla Yogurts + 3 Ghee Tubs | Selected | 22.29% | $376.49 |
| Bundle of 5 Mixed Vegetables + 2 Fresh Berries | Selected | 56.18% | $489.86 |
| Bundle of 1 Garlic Butter + 1 Free-range Eggs | Selected | 54.39% | $779.5 |
| Bundle of 2 Seasonal Fruits + 1 Salted Butter | Selected | 59.58% | $670.26 |
| Salted Butter | Selected | 25.72% | $1007.21 |
| Multigrain Cereal | Selected | 30.37% | $945.92 |
| Cottage Cheese | Not Selected | 0.0% | $285.37 |
| Brown Eggs | Selected | 19.57% | $1563.42 |
| Seasonal Fruits | Selected | 23.01% | $1360.35 |
| Tropical Juice | Selected | 25.15% | $1087.35 |
| Skimmed Milk | Selected | 22.31% | $587.34 |
| Fresh Vegetables | Selected | 29.04% | $1728.79 |
| Fruit Yogurt Cups | Selected | 9.9% | $831.27 |

**Summary Metrics:**
Total Revenue: **$15331.68**

Figure: Results

# Challenges and Future Work

**Key Challenges:**

- **Retail Data Complexity:** Large product base and diverse customer behavior required robust data preprocessing and feature engineering.
- **Customer Segmentation:** Choosing the right clustering method and validating with metrics like silhouette score was critical.
- **Cross-Elasticity Modeling:** Capturing interdependencies among products was computationally intensive but crucial for bundle design.
- **Optimization with Constraints:** Balancing budget, revenue, and fairness across customer segments posed significant difficulties.

**Future Directions:**

- **Scalability & Generalization:** Extend to enterprise-scale data using distributed systems for real-time processing. Test and adapt the framework across diverse retail domains (e.g., electronics, apparel).
- **Usability:** Develop a UI for parameter input and bundle generation.
- **External Data Integration:** Incorporate competitor pricing, economic trends, and campaign schedules.

# References I


- Beal, L. T., & Hedengren, J. D. (2020). *GEKKO Optimization Suite*. `https://github.com/BYU-PRISM/GEKKO`
- IBM. (2023). *IBM ILOG CPLEX Optimization Studio*. `https://www.ibm.com/products/ilog-cplex-optimization-studio`
- Sun, H., & Li, X. (2022). A Survey on Bundle Recommendation: Methods, Challenges, and Applications. *IEEE Access*, 10, 46827–46847.
- Wei, Z., & Liu, F. (2021). Towards Personalized Bundle Creative Generation. In *Proc. 30th Int. Conf. on Computational Linguistics*, 1329–1337.
- Sun, H., & Yang, X. (2022). Revisiting Bundle Recommendation: Datasets, Tasks, and Challenges. *ACM Computing Surveys*, 54(8), 1–25.
- Zhu, Y., & Harrington, K. (2020). Bundle Recommendation in E-commerce: The BRP and Its Applications. In *Proc. 41st Int. Conf. on Information Systems*.

# References II


Pathak, R., & Gupta, A. (2021). Generating and Personalizing Bundle Recommendations Using Matrix Factorization. *IEEE TKDE*, 33(12), 5171–5182.

Liu, S., & Fu, Y. (2020). Modeling Buying Motives for Personalized Product Bundling. *J. of Retailing and Consumer Services*, 56, 102162.

Ardakani, M. B., & Fathian, M. (2016). A Novel Model for Product Bundling in E-commerce. *Int. J. of Data Analysis Techniques and Strategies*, 8(1), 45–62.

Ettl, M., & Harsha, S. (2020). A Data-Driven Approach to Personalized Bundle Optimization. In *Proc. IEEE ICDM*, 184–193.

Maheswari, U. G., Sethuraman, S., & Ramanan, S. (2025). Optimizing Retail Promotions with MINLP and Cross-Elasticity Effects. In *Proc. Int. Conf. on Optimization and Learning (OLA2025)*.

# Thank You!

https://github.com/viinod9/M.Tech-Project

ma23m026@smail.iitm.ac.in