ISYE Homework 13

Problem: Optimizing Shelf Space Allocation to Maximize Profit and Validate Effectiveness
To tackle this problem, we can combine data analysis, predictive modeling, optimization, and real-world testing. This way, we can make sure our recommendations are practical, effective, and backed by solid data.

Data Requirements
A robust dataset is essential for effective analysis and modeling. Required data:
1. Product-Level Data
    - Product ID: Unique identifier for each product
    - Category: Classification of products into groups
    - Price and Cost: Retail price and supplier cost for calculating profit margins
    - Package Size: Dimensions to calculate shelf space requirements
    - Historical Shelf Space: Allocation in previous periods
    - Cross-Selling Potential: Groupings of frequently purchased products
2. Sales Data
    - Transaction Data: Includes basket-level purchases, quantity sold, and timestamps
    - Seasonality: Historical sales trends to identify seasonal patterns and demand spikes
    - Store ID: Links sales to specific store layouts and characteristics
3. Store-Level Data
    - Shelf Space Constraints: Minimum and maximum space requirements per product type
    - Total Space: Store-wide capacity for shelf space
    - Store Layout: Aisle arrangements, shelf configurations, and product placement impacts
    - Customer Demographics: Buyer behavior linked to age, income, and other characteristics

---

Analytical Models
This phase involves using advanced analytics to understand relationships, predict outcomes, and optimize shelf space.
A. Time-Series Analysis
    - Given: Historical sales data, including seasonality and trends
    - Use:
        - Apply exponential smoothing or ARIMA models to adjust for seasonality and isolate long-term trends
        - Identify high-demand periods for specific products
    - To:
        - Ensure sales data used in further models accurately reflects product performance without seasonal bias

B. Regression Analysis
    - Given: Seasonally adjusted sales data, shelf space allocation, and price data.
    - Use:
        - Fit linear or polynomial regression models to quantify the relationship between shelf space and sales or profit
        - Use advanced regression methods (lasso, ridge, elastic net) to identify key drivers of sales
    - To:
        - Test the hypothesis: Does increasing shelf space lead to higher sales or profit?
        - Determine elasticity: How much sales increase for each additional unit of shelf space

C. k-Means Clustering

- Given: Transaction data showing products frequently purchased together and product attributes
- Use:
  - Cluster products based on co-occurrence and shared attributes
  - Identify cross-selling opportunities by grouping complementary items
- To:
  - Inform shelf layouts by placing related products close together to encourage cross-selling

D. Optimization

- Given:
  - Regression outputs, clustering results, and store constraints
- Use:
  - Develop an optimization model with the objective function to maximize profit:
  - Constraints:
    - Total space available in the store.
    - Minimum and maximum shelf space limits for each product.
    - Placement of clustered products for cross-selling opportunities.
- To:
  - Allocate shelf space optimally to maximize profit while adhering to physical and business constraints.

Validating Results

To ensure the effectiveness of the recommendations, employ experimental validation techniques:

A. A/B Testing

- Given:
  - Current sales data as a baseline
  - Optimized shelf layouts versus legacy layouts
- Use:
  - Test the optimized layout in selected stores (treatment group) and retain the legacy layout in others (control group)
  - Measure metrics like total sales, profit, and cross-selling increases
- To:
  - Validate whether the optimized layout leads to statistically significant improvements.
  - Ensure the model's predictions align with real-world outcomes before scaling company-wide

B. Statistical Testing

- Given: A/B testing results and regression outputs.
- Use:
  - Perform hypothesis testing (e.g., t-tests) to determine if shelf space changes significantly impact sales or profit
- To:
  - Verify that observed improvements are not due to chance

Practical Considerations

1. Scalability:
   - Ensure models and recommendations can be scaled across multiple stores with varying layouts and constraints
2. Data Quality:

    - Ensure all data is accurate, complete, and up-to-date for reliable insights
3. Customer Behavior:
    - Incorporate customer demographics and local preferences to fine-tune recommendations for each store
4. Diminishing Returns:
    - Account for diminishing returns in sales as shelf space for a product increases