Retail Customer Data Analysis and Product Recommendation Report

1. Introduction

This report presents an analysis of customer purchase data for a retail company, focusing on identifying top-selling products, classifying customers based on spending behavior, and providing AI-driven product recommendations.

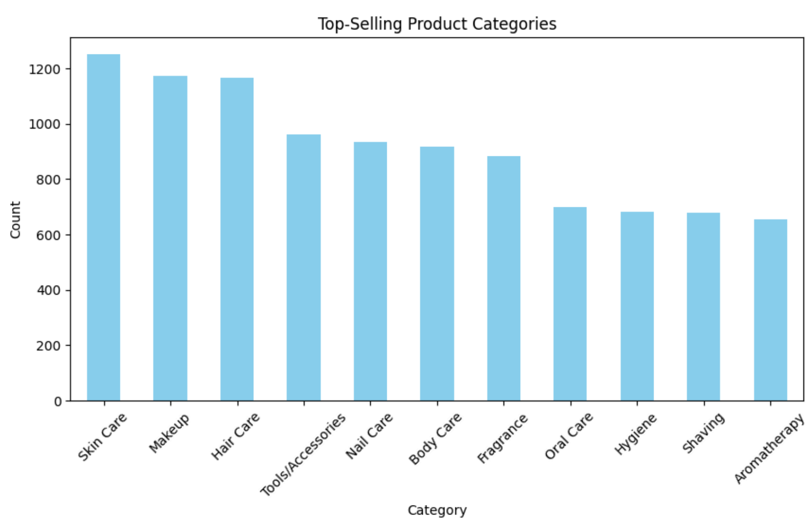
2. Data Overview

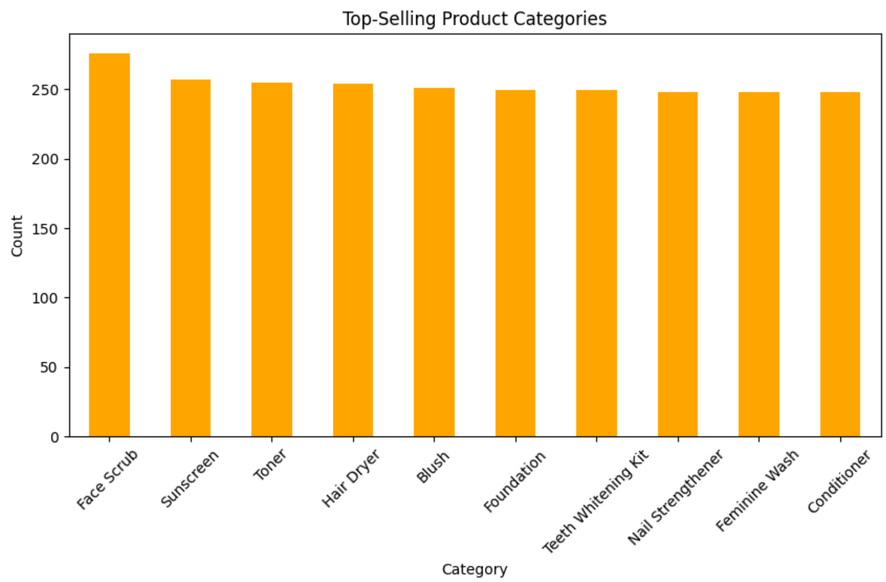
The dataset consists of anonymized customer purchase records with the following attributes:

- Customer ID
- Product ID
- Product Category
- Purchase Amount
- Purchase Date

3. Data Analysis

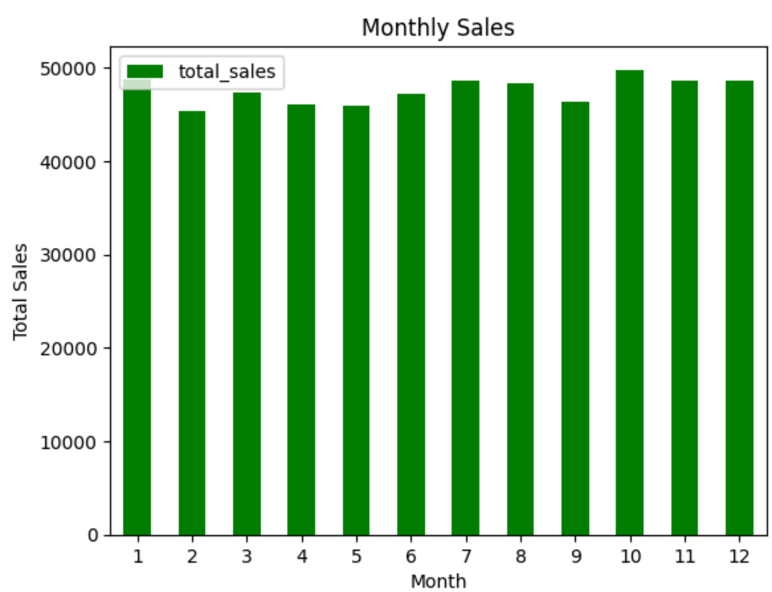
- **Top-Selling Products and Categories:**
  - The most frequently purchased products and categories were identified using value counts.
  - The top categories were visualized using a bar chart.





- **Average Spending per Customer:**
  - The total spending per customer was computed and the average was derived to understand general spending behavior.

- **Monthly Sales Trend:**
  - Analyzing the sales trend over months, we observe that the sales remain relatively stable across different months, indicating that customer purchases are not significantly affected by seasonal trends, weather changes, or sales periods.
  - This suggests that the store maintains a steady customer base with consistent purchasing habits rather than relying on seasonal spikes.



- **Purchase Frequency Distribution:**
    - Examining the distribution of purchase frequencies per customer reveals that a small percentage of customers make frequent purchases, while a larger portion consists of occasional buyers.
    - This insight helps in designing targeted marketing strategies, such as loyalty programs for frequent buyers and promotional campaigns for less active customers.

- **Spending Patterns Across Categories:**
    - Customers tend to spend more on specific product categories, showing a clear preference pattern.
    - Some categories have a high purchase volume but lower average spending, while premium product categories exhibit lower sales volume but higher spending per purchase.

4. Dataset

The Beauty Retail Shop Transactions Dataset (2024) is a synthetically generated dataset designed to simulate real-world customer purchases in a retail beauty store. It includes 10,000 transactions from 600 unique customers, covering 60 products across 11 categories[8].

Dataset Details

- Timespan: January to December 2024
- Transactions: 10,000
- Unique Customers: 600
- Products: 60
- Categories: 11

Product Categories

- Hair Care: Shampoo, Conditioner, Hair Oil, Hair Serum, Hair Mask
- Skin Care: Face Wash, Moisturizer, Sunscreen, Face Scrub, Toner
- Nail Care: Nail Polish, Nail Remover, Nail Strengthener, Cuticle Oil
- Body Care: Body Lotion, Body Wash, Body Scrub, Deodorant
- Tools/Accessories: Makeup Brushes, Hair Dryer, Straightener, Eyelash Curler
- Makeup: Foundation, Lipstick, Eyeliner, Mascara, Blush
- Aromatherapy: Essential Oils, Aroma Candles, Diffusers
- Oral Care: Toothpaste, Mouthwash, Teeth Whitening Kit
- Shaving: Razor, Shaving Cream, Aftershave
- Hygiene: Hand Sanitizer, Wet Wipes, Feminine Wash
- Fragrance: Perfume, Body Mist, Cologne, Roll-on Deodorant

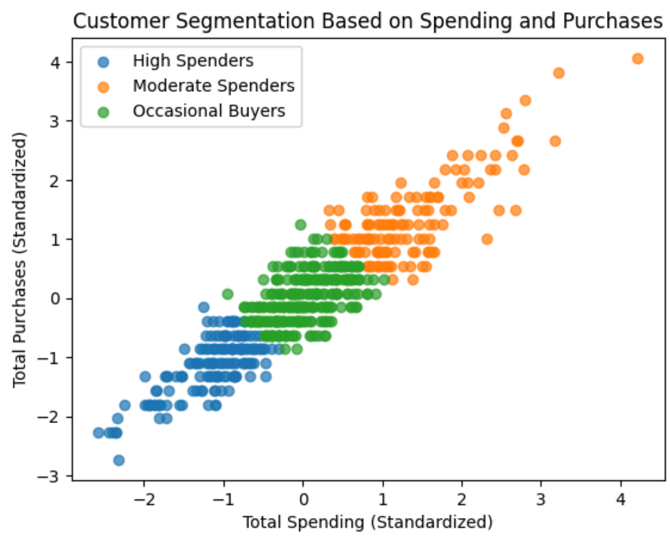
```
Basic Dataset Information:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10000 entries, 0 to 9999
Data columns (total 6 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   cust_id        10000 non-null  object
 1   prod_id        10000 non-null  object
 2   product_name   10000 non-null  object
 3   category       10000 non-null  object
 4   purchase_amt   10000 non-null  float64
 5   purchase_date  10000 non-null  object
dtypes: float64(1), object(5)
memory usage: 468.9+ KB
None

Summary Statistics:
       purchase_amt
count  10000.000000
mean      57.084050
std       25.125422
min        6.990000
25%       43.490000
50%       56.490000
75%       78.990000
max       98.490000
```

5. Methodology

This section describes the machine learning models and methodologies used for customer classification and product recommendation, along with hyperparameter tuning and complexities encountered.

- **Customer Classification (Clustering with K-Means):**
  - K-Means clustering was used to segment customers based on spending behavior, frequency of purchases, and product preferences. The optimal number of clusters was determined using the elbow method.
  - Customer Segments
    - High Spenders - Frequent buyers with high overall spending.
    - Moderate Buyers - Regular shoppers with moderate purchase amounts.
    - Occasional Buyers - Infrequent buyers with lower spending.

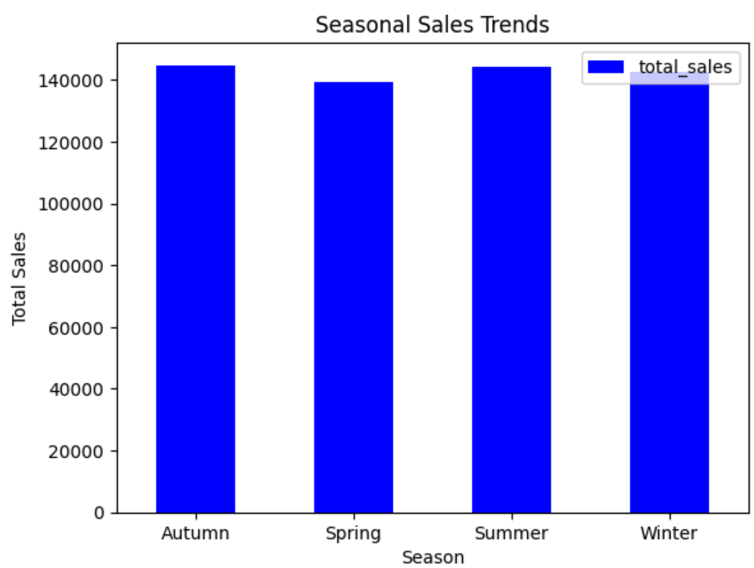


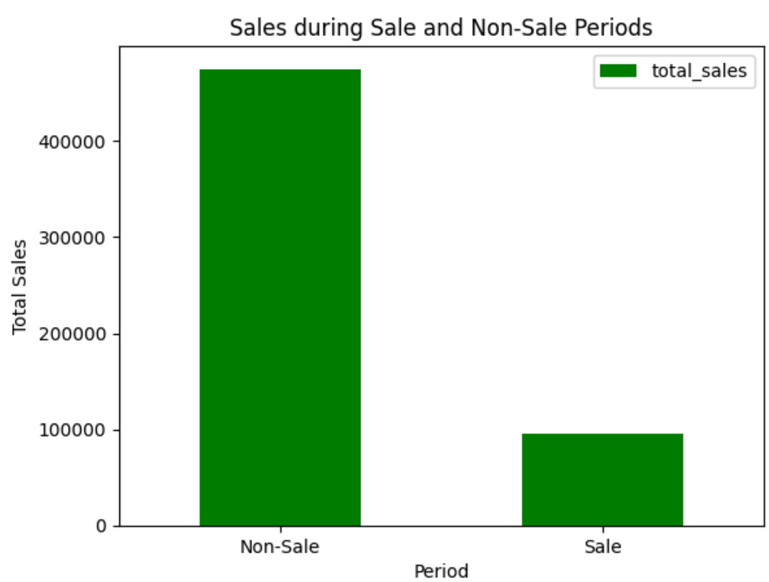
- **Product Recommendation Approaches:**
  - **Collaborative Filtering (K-Nearest Neighbors, KNN):**
    - KNN was applied to find customers with similar purchasing behaviors.
    - The choice of k=5 was made based on performance in similarity detection.
    - A distance metric (cosine similarity) was used to measure customer similarity in purchasing habits.

  - **Content-Based Filtering (Transformers for Text Embeddings):**
    - Product descriptions were embedded using the SentenceTransformer model, which converts textual data into high-dimensional feature vectors.
    - Cosine similarity between product embeddings was used to find similar products.

- This approach allowed the system to recommend products with similar descriptions, even if they had not been frequently purchased together.

- **Seasonal and Sale-Based Recommendation Adjustments:**
  - Purchase trends were analyzed by month and sales periods to identify seasonal effects.
  - Historical sale-based purchases were factored into recommendations to align with customer buying preferences during discount periods.





6. Product Recommendation System

- AI-driven recommendations were generated using multiple approaches:
    - **Collaborative Filtering:** Nearest Neighbors was used to find similar customers and recommend products based on what other similar customers purchased.
    - **Content-Based Filtering:** Cosine similarity of product descriptions (embedded using SentenceTransformer) was utilized to recommend similar products.
    - **Purchase History-Based Recommendations:** Products were recommended based on a customer's past purchases, ensuring continuity in product preferences.
    - **Seasonal Trend-Based Recommendations:** Product suggestions considered seasonal trends, ensuring relevant products were highlighted during peak seasons.
    - **Sale-Based Recommendations:** Customers who made purchases during sales periods were recommended similar discounted items to encourage future purchases.

7. Insights and Findings

- Certain product categories dominated purchases, indicating strong customer preferences.
- High spenders exhibited a pattern of frequent and large-value purchases, while occasional shoppers had sporadic transactions.
- The recommendation system successfully suggested products based on past purchases, improving potential customer engagement.

8. Example Recommendation

For CUST001, the KNN-based recommendation suggests products like Face Scrub, Body Mist, and Nail Strengthener, indicating that customers with similar purchase behavior have bought these items together. The Content-Based Filtering recommendation includes Conditioner, Hair Oil, and Shampoo, as they belong to the same Hair Care category, aligning with the customer's past purchases. This hybrid approach balances collaborative trends with individual preferences.

9. Beauty Retail Dashboard Documentation

<u>Frontend (HTML, CSS, JavaScript)</u>

The frontend consists of a responsive dashboard built with modern web technologies:

Key Components

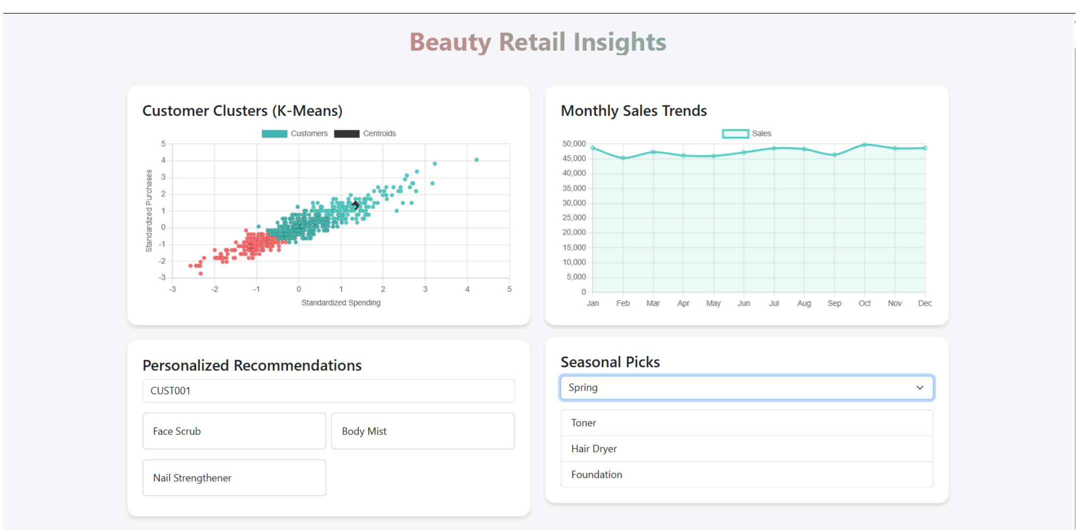
- HTML: A structured layout with cards for each visualization and interaction point
- CSS: Custom styling with hover effects, gradients, and responsive design
- JavaScript: Chart.js integration for data visualization and fetch API for backend communication

Features

- Customer Clusters: K-means visualization showing customer segmentation based on spending patterns
- Monthly Sales: Line chart displaying sales trends throughout the year
- Personalized Recommendations: Dynamic product recommendations based on customer ID input
- Seasonal Picks: Dropdown selector for seasonal product recommendations

Implementation Notes

- Uses Bootstrap 5 for grid layout and basic styling
- CSS variables for consistent color theming
- Asynchronous data loading with error handling
- Interactive elements that update in real-time based on user input



Backend (Flask API)

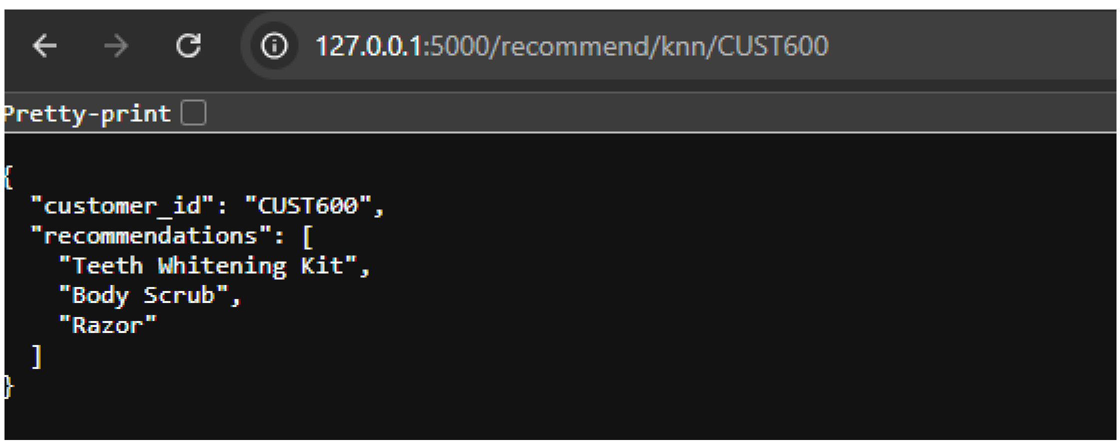
The backend provides data processing and recommendation algorithms through a REST API:

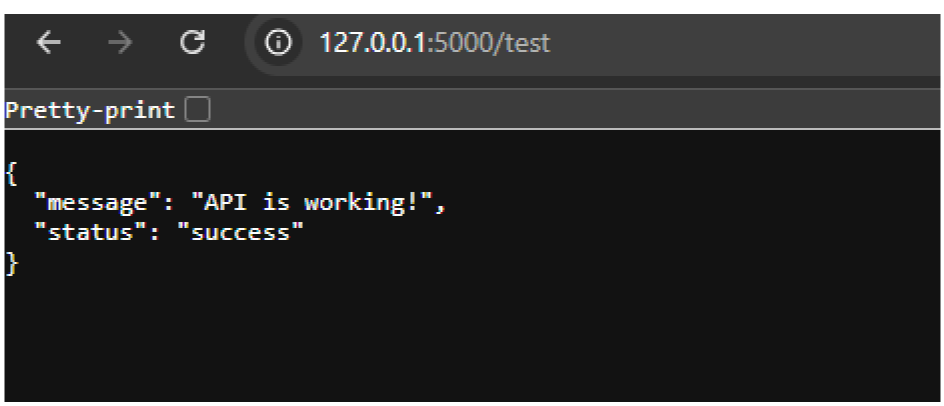
Key Components

- Flask Framework: Lightweight web server handling API requests
- Data Processing: Pandas for data manipulation and analysis
- Machine Learning: Scikit-learn for K-means clustering and nearest neighbors algorithms
- Static File Serving: Direct delivery of the dashboard HTML

Endpoints

- /dashboard: Serves the main dashboard HTML
- /test: Simple health check endpoint
- /analysis/clusters-visual: K-means customer clustering data
- /analysis/monthly-sales: Monthly sales performance metrics
- /recommend/knn/<customer_id>: Collaborative filtering recommendations
- /recommend/seasonal/<season>: Season-based product recommendations
- /recommend/content-based/<customer_id>: Content-based filtering
- /recommend/sale: Sale-period product recommendations





10. Conclusion This analysis and AI-powered recommendation system can enhance customer engagement and sales by offering personalized product suggestions. Future improvements can include advanced deep learning models and real-time recommendation updates.

11. Future Scope

- Implementing deep learning-based recommendation models.
- Enhancing feature engineering for better clustering results.
- Deploying the recommender system as a web service.