Customer Purchase Analysis

The dataset chosen for this analysis is based on Nuuly, a subscription-based clothing rental platform tailored to women's fashion. This dataset was sourced from Hugging Face and offers a unique perspective on how modern businesses are adapting to evolving customer needs in the fashion industry.

Unlike traditional e-commerce platforms focused solely on buying and selling, Nuuly operates a hybrid model where customers can rent clothes and purchase them if they wish after the rental period. This distinctive business model introduces unique customer behaviors and trends, making the dataset particularly intriguing for analysis.

Data Preprocessing:

To ensure the dataset was clean and ready for thorough analysis, several preprocessing steps were undertaken:

1. Data Structuring: The *reviews* column was initially excluded to create a focused product dataset. This step was essential for isolating product-specific attributes and conducting targeted analysis of products and categories.

2. Transforming Data: After constructing the product dataset, the information from the *reviews* column was further processed. This step generated meaningful insights by creating separate datasets of customers, orders, products and reviews.

3. Feature Engineering: Introduced few new columns like product_category which was created by extracting category information from the product_link field. Few other columns were synthesized to reflect realistic purchasing behaviors and revenue patterns.

Clean and structured version of the datasets were created to facilitate different types of analysis. This dataset was saved as CSV files for use in subsequent analysis.

Customer Demographics:

The dataset comprises 33,377 unique customers, with diverse demographics. A detailed analysis was performed to understand customer attributes, such as age distribution and color preferences, which provide valuable insights into customer behavior.

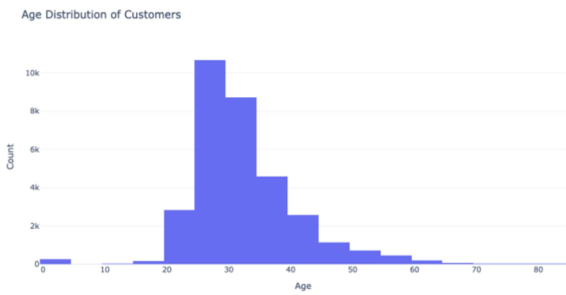
Age:

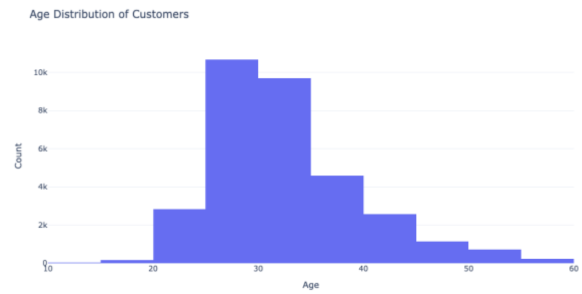
The initial analysis revealed that most customers fall within the **25-35 age** range, which is the most engaged demographic. This suggests that the platform appeals strongly to working professionals with disposable incomes.

However, the initial dataset contained outliers, including unrealistic age values below 15 and above 80, which could distort the analysis.

To address this, the Interquartile Range (IQR) method was applied to clean the data. The outliers were removed by identifying values outside the bounds defined as 1.5 times the IQR below the first quartile and above the third quartile.

The first graph illustrates the raw data, including outliers, while the second graph represents the cleaned dataset. After preprocessing, it was observed that the majority of customers belonged to the 25-40 age group, providing a more representative view of the customer base.





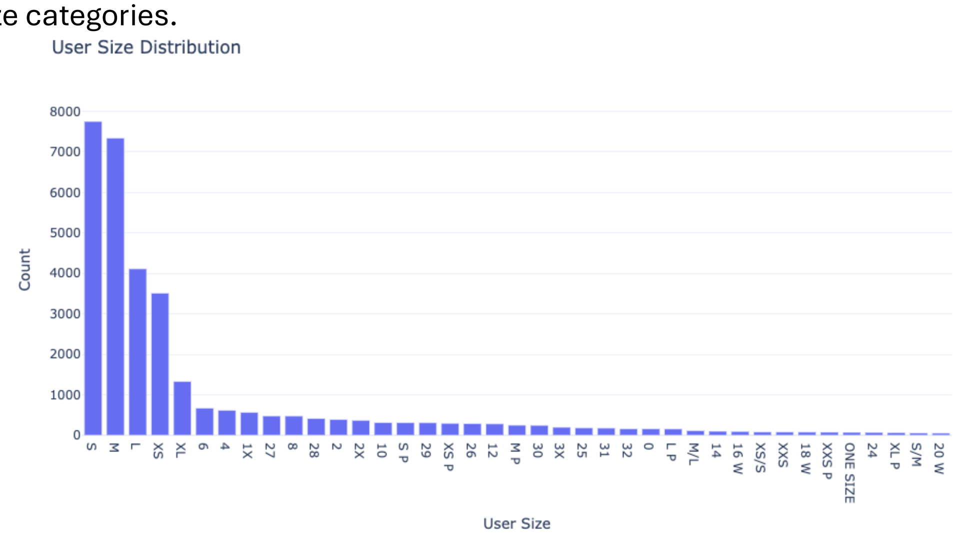
Expanding affordable options could attract younger customers, while focusing on value and longevity may appeal to older demographics.

Color Preferences:
An analysis of color preferences revealed that black was the most popular color, chosen by 5,876 customers, accounting for approximately 17% of the total customer base across 124 unique color options

Other top colors included multi-color (2,552 customers), white (1,785 customers), blue (1,628 customers), and brown (1,617 customers). These findings highlight a strong preference for classic and versatile colors among customers.

Size:
The dataset includes a detailed analysis of size distribution, focusing on sizes with at least 50 recorded sales. This threshold was chosen to ensure that only sizes with significant customer demand were analyzed, eliminating noise from sizes with minimal sales and providing a clearer picture of customer preferences

The most popular sizes among the customers were Small (S) ~ 23% of our customers, Medium (M) ~ 22% of our customers, Large (L) ~ 12% of the customers and extra small (XS) ~10% of the customers out of 151 unique sizes offered by Nuuly. Together, these four sizes accounts for 67% of all customer orders, indicating a strong concentration of demand within these size categories.



Body Type:
The most common body types among customers are:

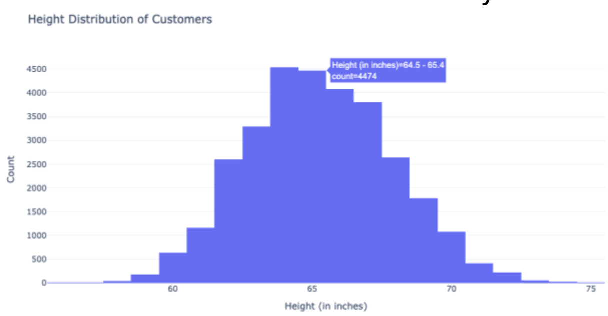
- **Hourglass**: ~42% of the total
- **Straight**: ~24% of the total

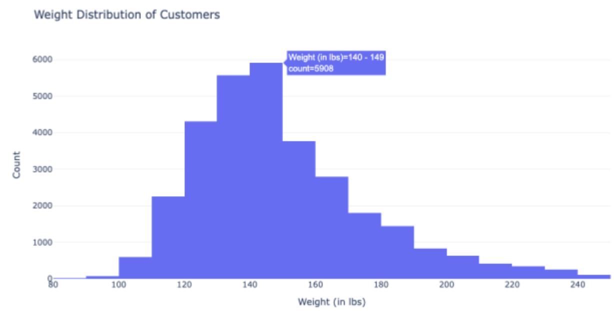
This indicates strong demand from customers with Hourglass and Straight body types, suggesting a focus on designs that cater to these groups.

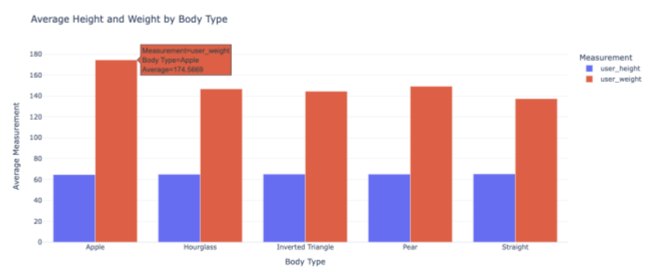
Height and Weight:
The average customer falls within:

- **Height**: 5’3” to 5’9” (160–175 cm)
- **Weight**: 120–180 lbs (54–82 kg)
-

Outliers outside the ranges of height (<4’10” or >6’5”) and weight (<90 lbs or >250 lbs) were excluded to maintain data accuracy.



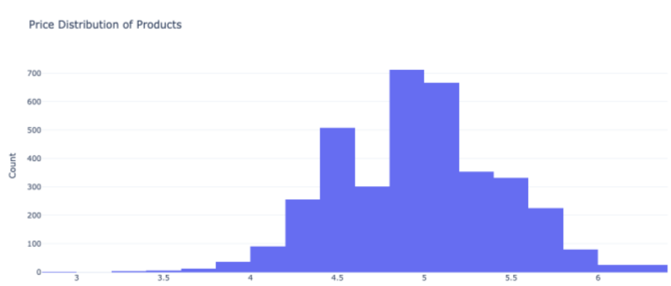




Customers with Apple and Pear body types, associated with higher weights, may benefit from more size-inclusive designs. Offering well-fitted options for all body types can improve customer satisfaction and attract a broader audience.

<u>Product Insights:</u>

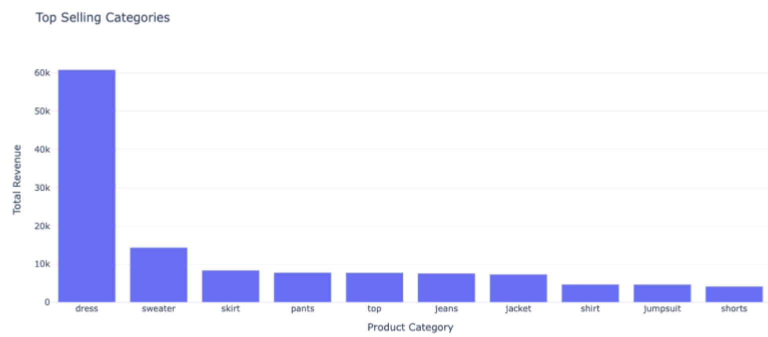
Price Distribution:
The dataset includes a diverse range of product prices, with a significant concentration of items priced between $50 and $150. These mid-range products dominate the inventory, reflecting a strong appeal to value-conscious customers.
High-end products, priced above **$300**, constitute a smaller portion of the inventory, targeting a niche customer base with premium preferences.



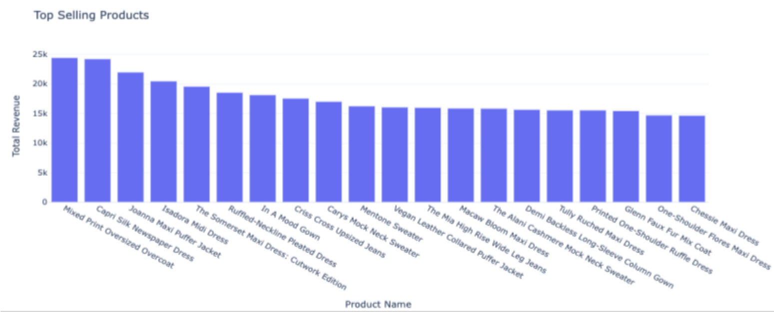
This distribution suggests that most customers prefer affordable, mid-range options. Maintaining a robust inventory in the $50–$150 price range, while offering premium products strategically, can balance accessibility and exclusivity, catering to diverse customer segments effectively.

Top Selling Products and Categories:

The analysis highlights that **Dresses** are the top revenue-generating category, contributing over **$60,000**—significantly higher than any other category. This indicates strong customer preference for dresses, making them the leading category by a wide margin.



Other categories, such as **Sweaters**, **Skirts**, and **Pants**, contribute moderately to overall revenue. These categories exhibit steady demand but are likely more seasonal, with customer interest fluctuating based on specific trends or needs.



High-performing products, such as the **Mixed Print Oversized Overcoat**, generate approximately **$24,000 annually** per product, underscoring their importance in driving sales.

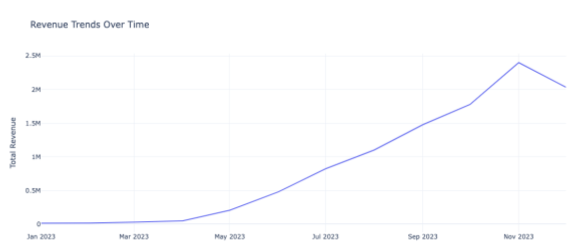
These best-sellers products are predominantly from high-revenue product categories like **Dresses** and **Sweaters**, reflecting customer preferences for timeless and versatile wardrobe staples.

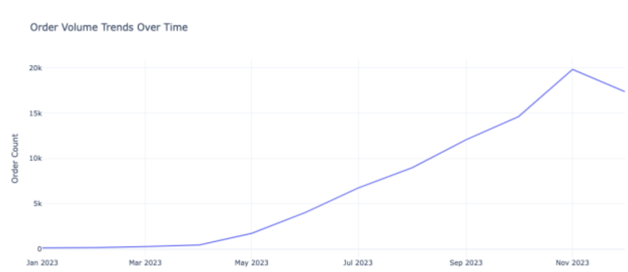
Additionally, products like the **Capri Silk Newspaper Dress** and **Joanna Maxi Puffer Jacket** also generate over **$20,000 annually**, demonstrating their strong appeal. These items combine elegance with contemporary trends, captivating buyers and maintaining their position within the core product lineup.

The analysis indicates that focusing on top-performing categories like Dresses and Sweaters, along with best-selling products, can help maximize revenue. Seasonal items like Pants and Skirts should be strategically managed to capture demand spikes while maintaining inventory efficiency.

Order and Revenue Trends for FY 2023

Both order volume and revenue show a consistent upward trajectory from January 2023 to November 2023, indicating strong business momentum and increasing customer engagement.





The highest spike in both order volume and revenue occurred in November 2023, suggesting a seasonal effect, likely driven by holiday shopping or promotional campaigns such as Black Friday sales.

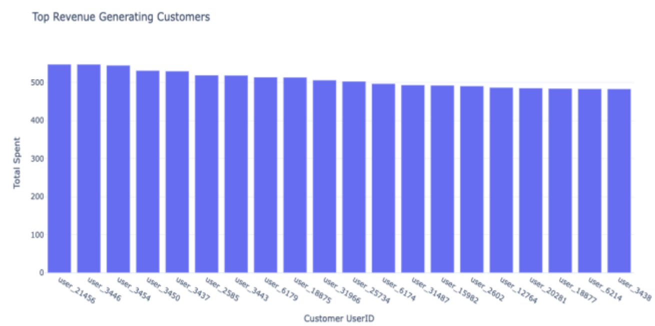
- **Revenue peaked at over $2.5 million**, marking the highest monthly performance.
- **Order volume approached 20,000 orders**, showcasing significant customer activity.

These trends highlight the impact of seasonal promotions and holiday sales on driving business performance. Understanding these patterns can help plan future campaigns and inventory management more effectively.

Average Spending of Customers:

The top customers, such as **user_21456**, **user_3446**, and **user_3454**, each spend an average of over **$500**, reflecting strong engagement and high purchasing power. These individuals represent a core group of high-value buyers who consistently invest in premium or multiple products.

These customers represent a core group of high-value buyers who consistently invest in premium or multiple products.

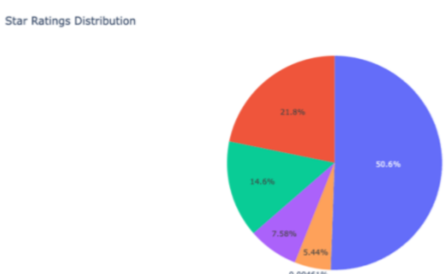


The top 15 customers show a narrow range of average spending, all staying above $450, suggesting a reliable segment of loyal and high-spending customers.
We should use these customers existing spending habits to introduce upselling strategies, such as bundling products or promoting premium collections.

Leveraging these spending habits can enhance revenue while fostering customer loyalty.

Rating Distributions:

Approximately ~ **50%** of reviews are 5-star, highlighting strong customer satisfaction and positive product experiences. Additionally, around ~**21.8%** of reviews are 4-star, while a smaller percentage falls into lower categories. To further improve overall satisfaction, efforts should focus on gathering customer feedback and addressing issues with products rated below **3.5 stars**.



The **Mixed Print Oversized Overcoat**, the highest revenue-generating product, maintains a strong **4.3-star average rating**, reinforcing its popularity and customer appeal.

Other top-performing products, such as the **Capri Silk Newspaper Dress** and **Joanna Maxi Puffer Jacket**, also have consistently high average ratings, aligning with their revenue contributions.

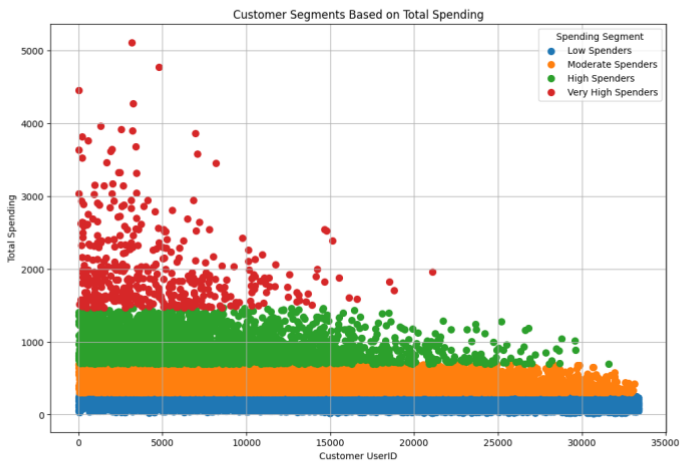
Top-selling categories, such as Dresses and Sweaters, predominantly feature products with average ratings of 3.5 stars or higher, indicating consistent customer satisfaction across these segments.

To enhance customer satisfaction and drive better ratings:

- Focus on improving lower-rated products by analyzing feedback and addressing recurring issues.
- Leverage insights from high-rated products to replicate successful design, quality, and performance characteristics across the inventory.

Customer Segmentation Analysis Based on Purchase Behavior

The Elbow Method was applied to determine the optimal number of clusters, identifying a clear "elbow point" at four clusters. This segmentation balances detail and simplicity, providing actionable insights into distinct customer behaviors.



Low Spenders: Largest group, minimal purchases, suitable for broad-based, cost-effective engagement strategies.
Moderate Spenders: Stable group with potential for conversion into high spenders through targeted campaigns.
High Spenders: Significant contributors to revenue, benefit from loyalty incentives and personalized offers.
Very High Spenders: Small group with the highest revenue, require premium services and exclusive attention.

The segmentation visual (scatterplot) reveals clear distinctions in spending behavior across the groups, with "Low Spenders" densely clustered at the lower end of total spending and "Very High Spenders" spread out with consistently high purchase amounts.

Key Metrics for Segmentation Justification:

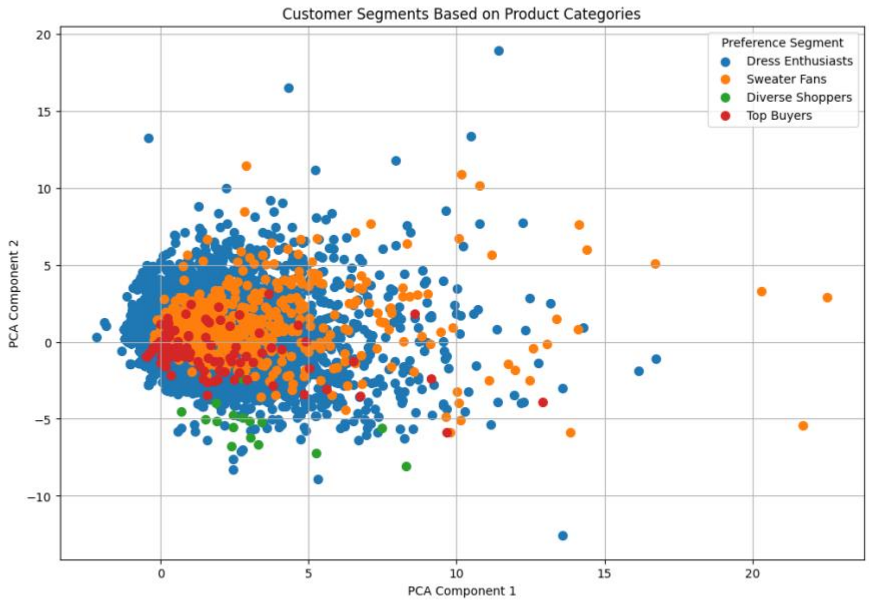
- Average Spending:
  - Low Spenders: <$100 per month.
  - Moderate Spenders: $100–$300 per month.
  - High Spenders: $300–$500 per month.
  - Very High Spenders: >$500 per month.
- Purchase Frequency:
  - Low Spenders: 1–2 purchases per year.
  - Moderate Spenders: 3–6 purchases per year.
  - High Spenders: 7–10 purchases per year.
  - Very High Spenders: 10+ purchases per year.

Product Purchasing Pattern

The clustering analysis identifies distinct customer segments based on their product preferences and spending patterns.

Dress Enthusiasts (Blue): This cluster represents customers with a strong preference for dresses and apparel.

Sweater Fans (Orange): This group is primarily focused on knitwear and winter-related products.
Diverse Shoppers (Green): These customers exhibit varied interests, purchasing across multiple product categories.
Top Buyers (Red): High-frequency buyers contributing the most revenue, often purchasing across multiple categories.



1. "Dress Enthusiasts" and "Sweater Fans" are largest segments, indicating dresses and knitwear dominate customer preferences.
2. The wider spread of "Sweater Fans" highlights need for more tailored product offerings within the knitwear category.
3. The homogeneous clustering of "Dress Enthusiasts" indicates consistent preferences, enabling effective targeted marketing strategies.

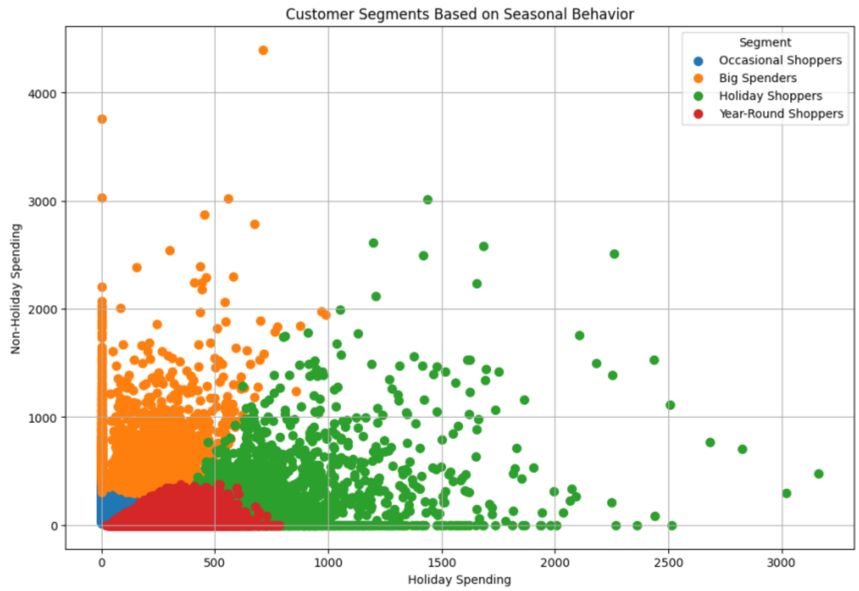
Recommendations:

- For **Diverse Shoppers**, AI can be leveraged to provide personalized based on their purchase and spending history to increase purchase frequency and generate revenue.
- VIP memberships can be offered and early access to sales to foster loyalty and drive revenue growth for **Top buyers**
- Bundles for **Diverse Shoppers**, which combines popular and underperforming products to the boost overall sales.

<u>Customer Segments on Holiday vs. Non-Holiday Spending</u>

The analysis categorizes customers into four distinct groups based on their seasonal spending behavior:

Occasional Shoppers(Blue): This segment represents a minor group with minimal spending and engagement.
Big Spenders(Orange): These customers contribute significantly to both holiday and non-holiday revenue streams, making them a priority for retention strategies.
Holiday Shoppers(Green): This group drives peak revenue during festive periods, showcasing a strong seasonal influence on their purchasing behavior.
Year-Round Shoppers(Red): Customers in this segment maintain stable spending habits throughout the year.



The scatterplot illustrates the distribution of holiday and non-holiday spending across customer segments. Holiday Shoppers cluster around high holiday spending, while Year-Round Shoppers and Big Spenders show balanced spending across both periods. Occasional Shoppers are densely packed near the lower spending range, emphasizing their low engagement.

- The majority of customers fall into the "Holiday Shoppers" (Green) and "Big Spenders" (Orange) categories, reflecting a mix of seasonal and year-round spending behavior.
- Holiday Shoppers generate the highest revenue during festive periods, while Big Spenders sustain revenue across the year.
- Occasional Shoppers exhibit low engagement and represent an opportunity to expand the customer base with targeted campaigns.

Recommendations:

1. Engage Occasional Shoppers with entry-level incentives like free shipping or first-purchase discounts to improve activity.
2. Consistently promote offers to Year-Round Shoppers to maintain and increase their stable spending.
3. Leverage data on Holiday Shoppers to preemptively launch campaigns during peak seasons.

AI-Powered Product Recommendation System:

The product recommendation module leverages customer purchase history and ratings to provide personalized product suggestions. This system uses a user-based collaborative filtering approach with cosine similarity to identify similar users and predict potential product interests.

Preparing the Data:

- First, the system loads three datasets: orders (order.csv), reviews (reviews.csv), and product details (product.csv). These are merged using *product_id* and *customer_userid* to create a comprehensive dataset.
- A user-item matrix is created, where each row represents a user, each column represents a product, and the values are the user’s ratings for those products.

Normalizing the Data:

- Since different users may rate on different scales, the system normalizes the data using *StandardScaler* to ensure the ratings are comparable.

Finding Similar Users:

- The system calculates similarity scores between users using cosine similarity. This helps identify which users have similar tastes based on their past ratings.

Making Recommendations:

- For a given customer, the system identifies their most similar users and uses their ratings to predict which products the customer might like.
- It filters out products the customer has already rated and ranks the remaining products based on their predicted scores.

Confidence Levels:

- To make the recommendations more meaningful, the system calculates a confidence score for each suggestion. This score shows how confident the system is that the user will like the recommended product.

Results:

- The top recommendations, along with their product names and confidence percentages, are returned. For example, for a customer ID *user_21456*, the system generated the following recommendations:

```
Recommended products for customer user_21456:
                                 product_id                 product_name  \
1310  5c592d78-04db-42af-aa85-8a4cc917816b    Joanna Maxi Puffer Jacket
2053  4c22194e-2ede-4387-8a33-101186a783e9  Paloma Baggy Corduroy Pants
332   c73609db-b1df-4296-899c-12433eec3d3c           Blanket Mila Skirt
1079  62b40c89-e56b-4672-8b04-4a57a8caf66a        Zadie Mesh Turtleneck
2008  d5dbaab9-962b-43b9-82dc-559bd937ea9a       Ria Turtleneck Sweater

      confidence
1310  100.000000
2053   69.375812
332     5.218588
1079    0.366393
2008    0.000000
```

The algorithm returns the products that it thinks the user will like, ranked by how confident it is in each recommendation.

Implementation Details:

The recommendation system is implemented in Python, utilizing libraries such as *pandas* for data manipulation and *scikit-learn* for normalization and similarity computation. The *recommend_*products function integrates the complete logic, taking the customer ID and the number of desired recommendations as inputs.

This module exemplifies the use of AI-driven techniques to deliver actionable insights, fostering a data-driven approach to customer engagement.

Conclusion:

This analysis effectively addresses the challenge of improving customer experience and business growth through a data-driven approach. By examining purchase behaviors, segmenting customers, and identifying product trends, actionable insights were derived to optimize inventory, refine marketing strategies, and enhance product offerings.

Furthermore, the integration of an AI-powered recommendation system personalizes the shopping experience by delivering tailored product suggestions, enhancing customer engagement and loyalty. The findings provide a practical framework to align business operations with customer needs, driving customer satisfaction, loyalty, and sustained revenue growth in a competitive market.