

Customer Shopping Behavior Analysis



1. Project Overview

This document presents a comprehensive analysis of customer shopping behavior, leveraging a multi-faceted approach encompassing exploratory data analysis, SQL-based insights, business intelligence dashboards, and predictive modeling. The ultimate goal is to provide actionable recommendations that drive strategic business decisions and enhance overall customer experience.

The primary challenge addressed by this project is the lack of a unified and deep understanding of how customers interact with our products and services. Without clear insights into purchasing patterns, preferences, and influencing factors, businesses often struggle with inefficient marketing campaigns, suboptimal inventory management, and missed opportunities for personalized customer engagement. This leads to wasted resources and a potential decline in customer satisfaction and loyalty.

Specifically, this project seeks to answer critical business questions such as: What are the distinct segments within our customer base? What products or categories are most popular among different groups? How do seasonal trends or promotional activities impact purchasing decisions? What factors contribute to customer churn or repeat purchases? By addressing these questions, we aim to transform raw transaction data into strategic intelligence.

The overarching objective is to empower stakeholders with data-driven insights to optimize sales strategies, refine product offerings, personalize marketing efforts, and ultimately foster long-term customer relationships. This foundational understanding will serve as a cornerstone for future growth and competitive advantage in the retail landscape.



2. Dataset Summary

- Rows: 3,900
- Columns: 18

Key Features:

- Customer demographics (Age, Gender, Location, Subscription Status)
- Purchase details (Item Purchased, Category, Purchase Amount, Season, Size, Color)
- Shopping behavior (Discount Applied, Promo Code Used, Previous Purchases, Frequency of Purchases, Review Rating, Shipping Type)
- Missing Data: 37 values in Review Rating column

3,900

Total Purchases

18

Data Columns

37

Missing Values

3. Exploratory Data Analysis using Python

We began with data preparation and cleaning in Python:

- 1 Data Loading
Imported the dataset using `pandas`.
- 2 Initial Exploration
Used `df.info()` to check structure and `.describe()` for summary statistics.
- 3 Missing Data Handling
Checked for null values and imputed missing values in the **Review Rating** column using the median rating of each product category.
- 4 Column Standardization
Renamed columns to **snake case** for better readability and documentation.
- 5 Feature Engineering
Created **age_group** column by binning customer ages. Created **purchase_frequency_days** column from purchase data.
- 6 Data Consistency Check
Verified if **discount_applied** and **promo_code_used** were redundant; dropped **promo_code_used**.
- 7 Database Integration
Connected Python script to PostgreSQL and loaded the cleaned DataFrame into the database for SQL analysis.

Data Visualization

Data visualization is the art and science of transforming raw data into understandable and actionable insights. In the context of customer shopping behavior analysis, it is crucial for identifying trends, patterns, and outliers that might otherwise remain hidden within vast datasets. By presenting complex information graphically, we empower stakeholders across the organization to grasp key findings quickly and make informed decisions, whether it's optimizing marketing campaigns, refining product strategies, or improving the customer experience.

While Power BI has been instrumental in our dashboard creation, exploring advanced visualization techniques and tools can further enhance our ability to communicate complex behavioral patterns effectively. The goal is to move beyond simple charts to create rich, interactive experiences that invite deeper data exploration.



Clarity & Simplicity

Visualizations should distill complex datasets into easily digestible formats, ensuring that key messages are conveyed without overwhelming the audience. This promotes faster comprehension and reduces misinterpretation.



Interactivity & Exploration

Modern visualization tools allow users to interact with data, filtering, drilling down, and exploring different facets of information. This self-service capability fosters deeper understanding and enables personalized analysis.



Storytelling with Data

Effective data visualization goes beyond mere presentation; it tells a compelling story, guiding the viewer through the insights and implications of the data. This helps in communicating findings persuasively and driving engagement.



Driving Actionable Insights

The ultimate goal of data visualization is to facilitate decision-making. Well-designed dashboards and reports highlight critical metrics and trends, empowering businesses to take timely and effective actions based on reliable data.

4. Data Analysis using SQL (Business Transactions)

We performed structured analysis in PostgreSQL to answer key business questions:

1. Revenue by Gender

Compared total revenue generated by male vs. female customers.

- Female: **\$75,191**
- Male: **\$157,890**

2. High-Spending Discount Users

Identified customers who used discounts but still spent above the average purchase amount.

Result: 839 customers

3. Top 5 Products by Rating

Found products with the highest average review ratings.

- Gloves (3.86)
- Sandals (3.84)
- Boots (3.82)
- Hat (3.80)
- Skirt (3.78)

4. Shipping Type Comparison

Compared average purchase amounts between Standard and Express shipping.

- Standard: \$58.46
- Express: \$60.48

5. Subscribers vs. Non-Subscribers

Compared average spend and total revenue across subscription status.

- Yes:** 1,053 customers, \$59.49 avg, \$62,645 total
- No:** 2,847 customers, \$59.87 avg, \$170,436 total

6. Discount-Dependent Products

Identified 5 products with the highest percentage of discounted purchases.

- Hat (50.00%)
- Sneakers (49.66%)
- Coat (49.07%)
- Sweater (48.17%)
- Pants (47.37%)

7. Customer Segmentation

Classified customers into New, Returning, and Loyal segments based on purchase history.

- Loyal:** 3,116 customers
- Returning: 701 customers
- New: 83 customers

8. Top 3 Products per Category

Listed the most purchased products within each category.

Accessories: Jewelry (171), Sunglasses (161), Belt (161)

Clothing: Blouse (171), Pants (171), Shirt (169)

Footwear: Sandals (160), Shoes (150), Sneakers (145)

Outerwear: Jacket (163), Coat (161)

9. Repeat Buyers & Subscriptions

Checked whether customers with >5 purchases are more likely to subscribe.

- No: 2,518 repeat buyers
- Yes: 958 repeat buyers

10. Revenue by Age Group

Calculated total revenue contribution of each age group.

- Young Adult: \$62,143
- Middle-aged: \$59,197
- Adult: \$55,978
- Senior: \$55,763

5. Dashboard in Power BI

Finally, we built an interactive dashboard in Power BI to present insights visually.



Dashboard Features

Key Metrics

- 3.9K Number of Customers
- \$59.76 Average Purchase Amount
- 3.75 Average Review Rating

Filters Available

- Subscription Status (No/Yes)
- Gender (Female/Male)
- Category (Accessories, Clothing, Footwear, Outerwear)
- Shipping Type (6 options)

Visualizations

- % of Customers by Subscription Status
- Revenue by Category
- Sales by Category
- Revenue by Age Group
- Sales by Age Group

6. Business Recommendations



Boost Subscriptions

Promote exclusive benefits for subscribers.



Customer Loyalty Programs

Reward repeat buyers to move them into the "Loyal" segment.



Review Discount Policy

Balance sales boosts with margin control.



Product Positioning

Highlight top-rated and best-selling products in campaigns.



Targeted Marketing

Focus efforts on high-revenue age groups and express-shipping users.

7. Model Deployment

Following the comprehensive analysis and business recommendations, the next crucial step involves deploying the predictive models developed from this customer shopping behavior data. Model deployment transforms analytical insights into actionable, automated solutions that can directly impact business operations. This phase is critical for operationalizing the data science effort, moving beyond static reports to dynamic systems that continually learn and adapt to customer interactions.

Effective deployment ensures that the insights gained from our data analysis are not just theoretical but are integrated into the daily fabric of the business, driving personalized experiences, optimizing marketing spend, and enhancing overall customer satisfaction.



Operationalization

Transitioning insights into automated systems for immediate business impact.



Personalization

Enabling real-time, tailored customer experiences and recommendations.



Scalability

Ensuring the models can handle increasing data volumes and user demands.



Continuous Improvement

Setting up feedback loops for ongoing model performance monitoring and retraining.

The successful deployment of these models will allow the organization to leverage predictive analytics to anticipate customer needs, optimize inventory, and target promotions with unprecedented precision, ultimately leading to improved revenue and stronger customer relationships.



Alternative Models

Even after we launch our first models, we must keep exploring other machine learning methods. This is vital for staying competitive and finding new patterns in how customers shop. This ongoing process helps us confirm our initial findings, make predictions more accurate, and answer new business questions as they come up. By testing different model types and algorithms, we can find the best, most flexible, and easiest-to-understand solutions for our changing business needs.

This proactive approach keeps our data analysis up-to-date. It helps us quickly adjust to market changes and improve how we engage, keep, and gain customers. Each new model gives us a fresh way to look at customer data, possibly uncovering connections and trends we missed before.



Boost Prediction Power

Use advanced techniques (like ensemble methods, deep learning) for more accurate forecasts.



Improve Customer Groups

Use smart grouping methods (like K-Means, DBSCAN) to sort customers into smaller, more specific segments.



Make Better Suggestions

Try new models that suggest products (like collaborative filtering) to make recommendations more personal.



Understand Models Better

Use simpler models (like decision trees or rule-based systems) to clearly see why models make certain decisions.

In the future, we will compare these alternative models against our current ones. We will look at key measures like customer lifetime value, how many customers we keep, and conversion rates. This will help guide our long-term data plan.

Tools & Technologies

This project leveraged a robust stack of tools and technologies to ensure efficient data processing, insightful analysis, and effective visualization. Our selection focused on powerful, scalable, and industry-standard solutions, enabling us to extract maximum value from the customer shopping behavior data.



Programming Core

- Python
- Pandas, NumPy
- Matplotlib, Seaborn

Foundation for data cleaning, manipulation, and basic visualization.



Data Management

- SQL
- PostgreSQL

Reliable storage and efficient querying of diverse datasets.



Business Intelligence

- Power BI

Interactive dashboards and visual reports for key insights.

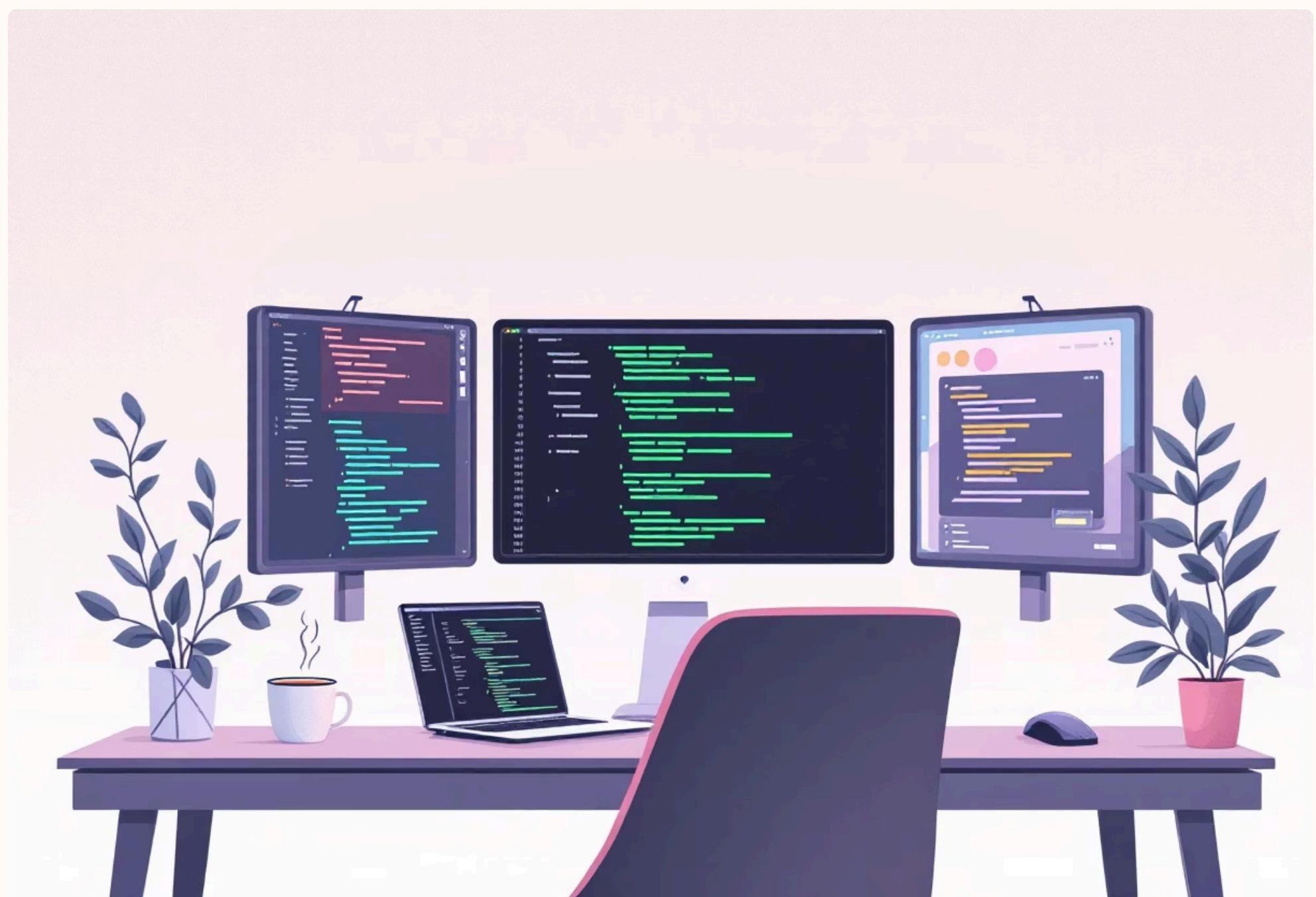


Advanced ML & AI

- Scikit-learn
- Ensemble Methods
- Deep Learning

Predictive modeling, classification, and deep pattern recognition.

The synergy between these tools allowed us to seamlessly move from raw data to actionable business recommendations, providing a comprehensive view of customer behavior and predictive capabilities for future strategies. This robust framework supports continuous improvement and adaptation to evolving market demands.



8. Future Work

Building on the successful deployment of our predictive models, the next phase of this project focuses on continuous innovation and deeper strategic integration. Future work will involve a multi-faceted approach to not only refine our existing analytical capabilities but also expand their scope, ensuring our insights remain relevant and impactful in a dynamic market. This forward-looking strategy aims to unlock even greater value from our customer shopping behavior data, fostering sustainable growth and competitive advantage.



Model Refinement & Optimization

Regularly monitor model performance, retrain with fresh data, and explore advanced algorithms to enhance predictive accuracy for customer segmentation and purchase probability.



Integration of External Data Sources

Incorporate new datasets such as demographic information, social media sentiment, and economic indicators to enrich customer profiles and provide a more holistic view of purchasing influences.



Advanced Predictive Analytics

Develop models for predicting Customer Lifetime Value (CLTV), identifying churn risk, and forecasting demand for new products, moving beyond descriptive analysis to proactive decision-making.



A/B Testing and Experimentation

Design and implement rigorous A/B tests to validate the impact of marketing campaigns, personalized recommendations, and promotional strategies derived from our models, ensuring data-driven optimization.



Expansion to New Customer Touchpoints

Apply our insights and deployed models to optimize customer experiences across additional channels, including in-store interactions, loyalty programs, and post-purchase engagement strategies.

These future initiatives will ensure that our data-driven approach to customer shopping behavior remains at the forefront, continuously delivering actionable intelligence and tangible business outcomes.

Conclusion

This comprehensive analysis of customer shopping behavior has equipped us with invaluable insights, transforming raw data into actionable strategies. From initial exploratory data analysis and SQL-driven transaction insights to the development of predictive models and the vision for future work, every stage of this project has aimed at fostering a deeper understanding of our customers and optimizing business outcomes. We've laid a robust foundation for a data-driven approach that promises continuous enhancement of customer experience and operational efficiency.

The successful deployment of our initial models marks a significant milestone, enabling real-time personalization and improved decision-making. As we look ahead, the strategic exploration of alternative models and the continuous integration of new data sources will ensure our competitive edge in a rapidly evolving market.

1

Enhanced Customer Understanding

Gained deep insights into purchasing patterns, preferences, and motivations across various segments.

2

Actionable Predictive Models

Developed and deployed models for predicting future behavior, optimizing marketing, and personalizing offers.

3

Foundation for Growth

Established a scalable, adaptable framework for continuous data analysis and strategic innovation.

Ultimately, this project culminates in a clear roadmap for leveraging data science to not only meet but exceed customer expectations, driving sustained growth and building stronger, more meaningful relationships with our customer base. The journey of data exploration and innovation is continuous, and this analysis serves as a powerful spring-board for our future endeavors.