

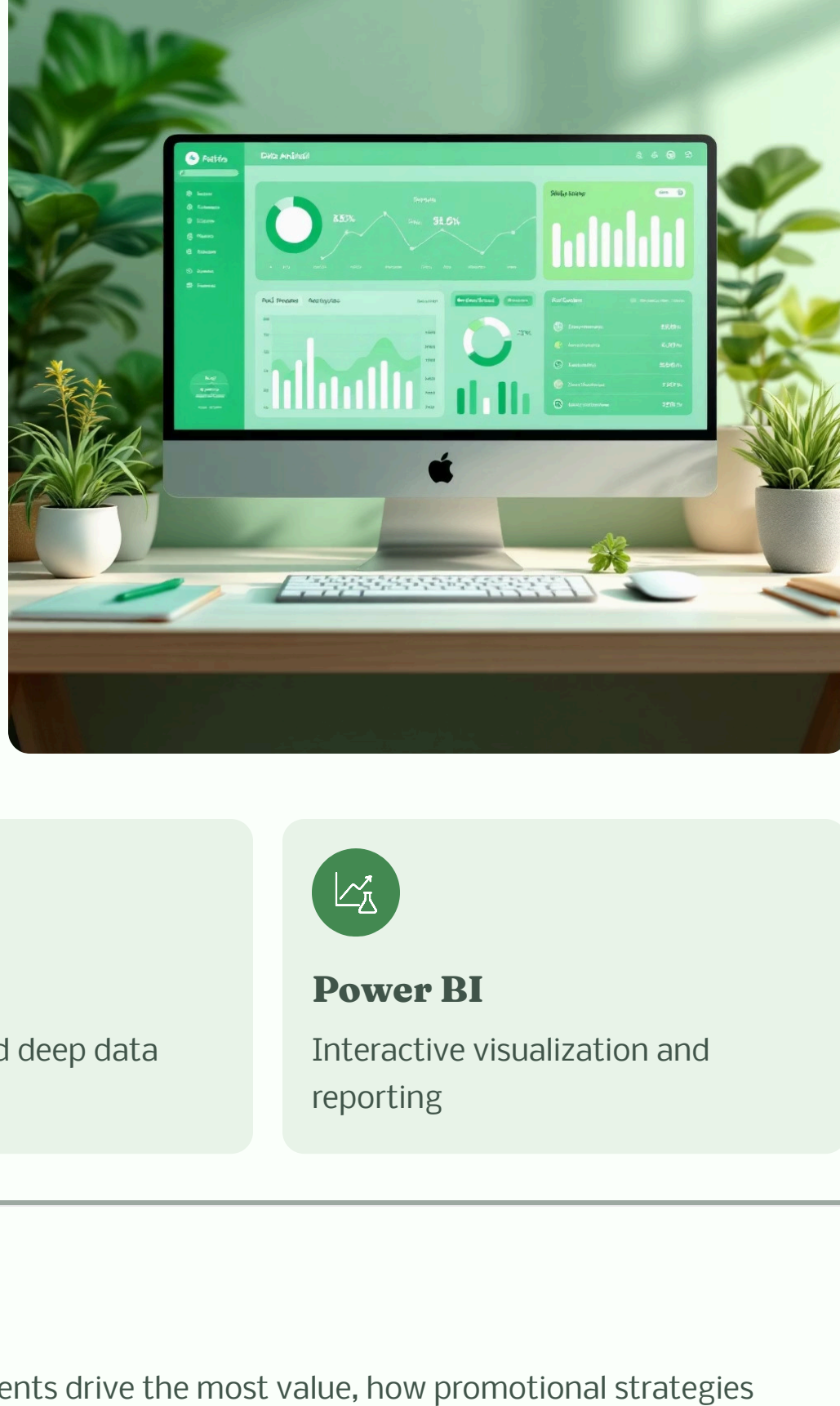
Retail Customer Behavior Analysis

Transforming Data into Strategic Business Insights

This comprehensive analysis explores customer purchasing patterns, discount utilization, subscription impact, and shipping preferences for a retail company. Using Python for data preprocessing, SQL for business intelligence queries, and Power BI for visualization, we uncovered actionable insights to optimize marketing strategies and drive revenue growth.

Key Focus Areas:

- Customer segmentation and revenue drivers
- Discount and subscription effectiveness
- Product category performance analysis
- Shipping method impact on behavior



Python
Data cleaning, preprocessing, and normalization

SQL (MySQL)
Business queries and deep data insights

Power BI
Interactive visualization and reporting

Business Challenge

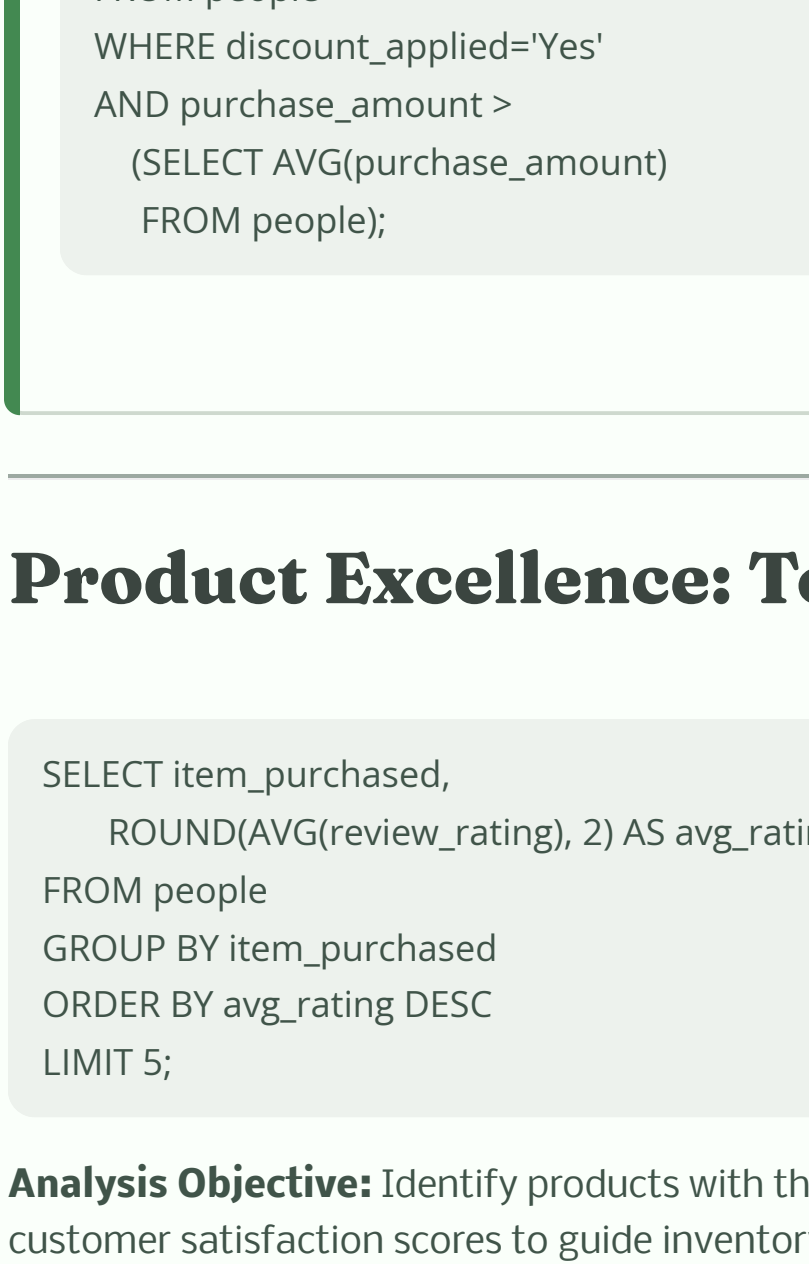
The retail company needed to understand which customer segments drive the most value, how promotional strategies impact purchasing decisions, and which operational factors influence buying behavior. This analysis aims to provide data-driven recommendations for optimizing marketing spend, improving customer retention, and maximizing revenue.

1 Data Preprocessing Cleaned and normalized raw customer data using Python pandas, handling missing values, duplicates, and data type conversions to ensure analysis-ready datasets.	2 SQL Analysis Executed strategic business queries to uncover patterns in customer spending, product preferences, and behavioral segments across demographics.	3 Visual Insights Created interactive Power BI dashboard to visualize key metrics, enabling stakeholders to explore data and make informed decisions.
--	--	---

Critical Business Questions Answered

Revenue by Demographics Which customer segments generate the highest total revenue and average purchase values?	Discount Impact How do promotional discounts affect spending patterns among different customer tiers?
Subscription Value What is the revenue contribution difference between subscribed and non-subscribed customers?	Shipping Preferences How does shipping method selection correlate with purchase amounts and customer loyalty?

Data Preprocessing Pipeline



The preprocessing phase established a clean, reliable foundation for analysis. Using Python's pandas library, we systematically addressed data quality issues and standardized formats.

Preprocessing Steps:

- Loaded raw CSV dataset into pandas DataFrame
- Identified and removed duplicate customer records
- Imputed missing values using statistical methods
- Normalized categorical variables (gender, subscription status, shipping type)
- Validated and converted numeric fields (purchase amounts, ratings)
- Exported cleaned data to MySQL database for analysis

SQL Analysis: Revenue by Gender

```
SELECT gender,
       SUM(purchase_amount) AS total_revenue
FROM people
GROUP BY gender;
```



Key Finding: Male customers generated slightly higher total revenue compared to female customers, indicating a balanced customer base with a modest male skew. This insight suggests marketing campaigns should maintain gender-neutral messaging while potentially testing male-focused promotions for incremental gains.

High-Value Discount Users

The Query <pre>SELECT customer_id, purchase_amount FROM people WHERE discount_applied='Yes' AND purchase_amount > (SELECT AVG(purchase_amount) FROM people);</pre>	Strategic Insight Contrary to conventional wisdom that discounts primarily attract budget-conscious shoppers, our analysis revealed a significant segment of high-spending customers actively using promotional codes. These customers spend above the average purchase amount even with discounts applied, suggesting that strategic promotions can drive incremental revenue from premium buyers rather than simply cannibalizing full-price sales.
--	---

Product Excellence: Top-Rated Items

```
SELECT item_purchased,
       ROUND(AVG(review_rating), 2) AS avg_rating
FROM people
GROUP BY item_purchased
ORDER BY avg_rating DESC
LIMIT 5;
```

Analysis Objective: Identify products with the highest customer satisfaction scores to guide inventory prioritization, marketing focus, and quality benchmarking across the catalog.

Business Implications:

- Top-rated products demonstrate strong quality and customer alignment
- High ratings correlate with repeat purchase likelihood
- These items should feature prominently in marketing materials
- Product teams can study these successes to improve lower-rated items
- Consider bundling top-rated products to increase average order value

Shipping Speed and Spending Behavior

Standard Shipping Customers selecting standard shipping typically exhibit price-sensitive behavior, with average purchase amounts trending lower. These buyers prioritize value over speed and often bundle items to maximize shipping efficiency.	Express Shipping Express shipping customers demonstrate higher average purchase values, indicating premium buying behavior. These customers value convenience and urgency, suggesting they may be professionals or higher-income segments willing to pay for expedited service.
--	---

```
SELECT shipping_type,
       ROUND(AVG(purchase_amount), 2) AS avg_purchase
FROM people
WHERE shipping_type IN ('Standard', 'Express')
GROUP BY shipping_type;
```

Strategic Recommendation: Maintain competitive express shipping rates to retain high-value customers while using standard shipping promotions to increase conversion among price-sensitive segments.

The Subscription Advantage

654 Total Customers Analyzed Comprehensive dataset coverage	67% Subscription Rate Strong adoption across customer base	₹58.46 Average Purchase Baseline spending per transaction
---	--	---

```
SELECT subscription_status,
       COUNT(customer_id) AS total_customers,
       ROUND(AVG(purchase_amount), 2) AS avg_spend,
       SUM(purchase_amount) AS total_revenue
FROM people
GROUP BY subscription_status;
```

Critical Finding: Subscribed customers contribute disproportionately higher revenue through both increased purchase frequency and larger average transaction values. The subscription model creates predictable revenue streams while deepening customer relationships and reducing churn risk. This validates the strategic importance of expanding subscription offerings and incentivizing non-subscribers to convert.

Discount-Sensitive Product Categories

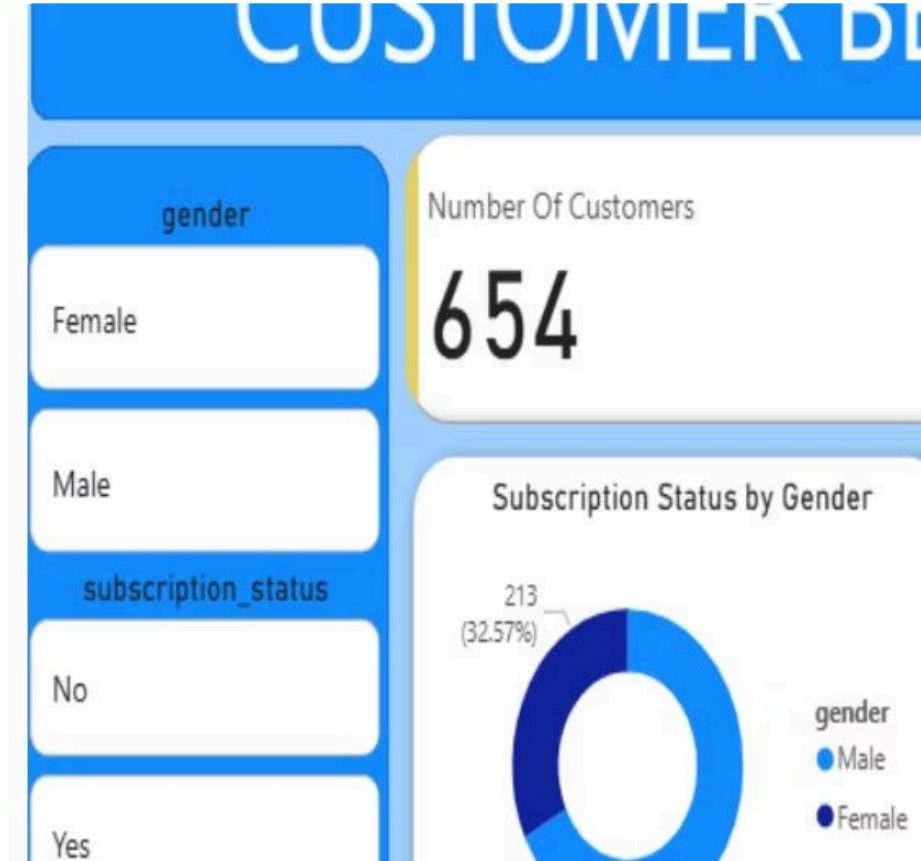
Understanding which products drive the most discount redemptions helps optimize promotional strategies and inventory planning.

```
SELECT item_purchased,
       ROUND(SUM(CASE WHEN discount_applied='Yes'
                       THEN 1 ELSE 0 END) * 100 / COUNT(*), 2)
       AS discount_rate
FROM people
GROUP BY item_purchased
ORDER BY discount_rate DESC
LIMIT 5;
```

Apparel Items Highest discount sensitivity, suggesting price-conscious shopping behavior in fashion categories	Accessories Moderate discount usage, indicating opportunities for strategic seasonal promotions	Electronics Lower discount dependency, reflecting brand loyalty and feature-driven purchasing decisions
--	---	---

Marketing Application: Focus aggressive discount campaigns on high-discount-sensitivity categories while maintaining premium positioning for electronics and accessories through value-added bundles rather than price reductions.

Customer Segmentation Analysis



```
SELECT CASE
  WHEN previous_purchases BETWEEN 0 AND 1
  THEN 'New'
  WHEN previous_purchases BETWEEN 2 AND 20
  THEN 'Returning'
  ELSE 'Loyal'
END AS customer_segment,
COUNT(*) AS segment_count
FROM people
GROUP BY customer_segment
ORDER BY segment_count;
```

Insight: The majority of customers fall into the "Returning" segment, indicating moderate brand retention. This presents an opportunity to develop targeted loyalty programs that convert returning customers into loyal advocates through personalized experiences and exclusive benefits.

Category Performance Deep Dive

Analyzing the top three products within each category reveals best-sellers and guides strategic inventory allocation.

```
SELECT category, item_purchased, total_count
FROM (
  SELECT category, item_purchased, COUNT(*) AS total_count,
         ROW_NUMBER() OVER (PARTITION BY category
                             ORDER BY COUNT(*) DESC) AS rankk
  FROM people
  GROUP BY category, item_purchased
) AS temp
WHERE rankk <= 3;
```

Clothing Dominance Skirts, shirts, and pants lead clothing sales, with shirts showing particularly strong cross-demographic appeal	Accessories Strength Jewelry and watches drive accessories revenue, suggesting opportunities for premium product expansion	Footwear Trends Sneakers and sandals dominate, reflecting casual fashion preferences among the customer base
--	--	--

Subscription and Repeat Purchase Correlation

```
SELECT subscription_status,
       COUNT(customer_id) AS repeat_buyers
FROM people
WHERE previous_purchases > 5
GROUP BY subscription_status;
```

Critical Discovery: The vast majority of high-frequency repeat buyers maintain active subscriptions, demonstrating a strong causal relationship between subscription status and customer lifetime value.



This correlation validates the subscription model's effectiveness in building lasting customer relationships. Repeat buyers with subscriptions not only purchase more frequently but also demonstrate higher tolerance for price increases and lower price sensitivity, making them the most valuable long-term customer segment.

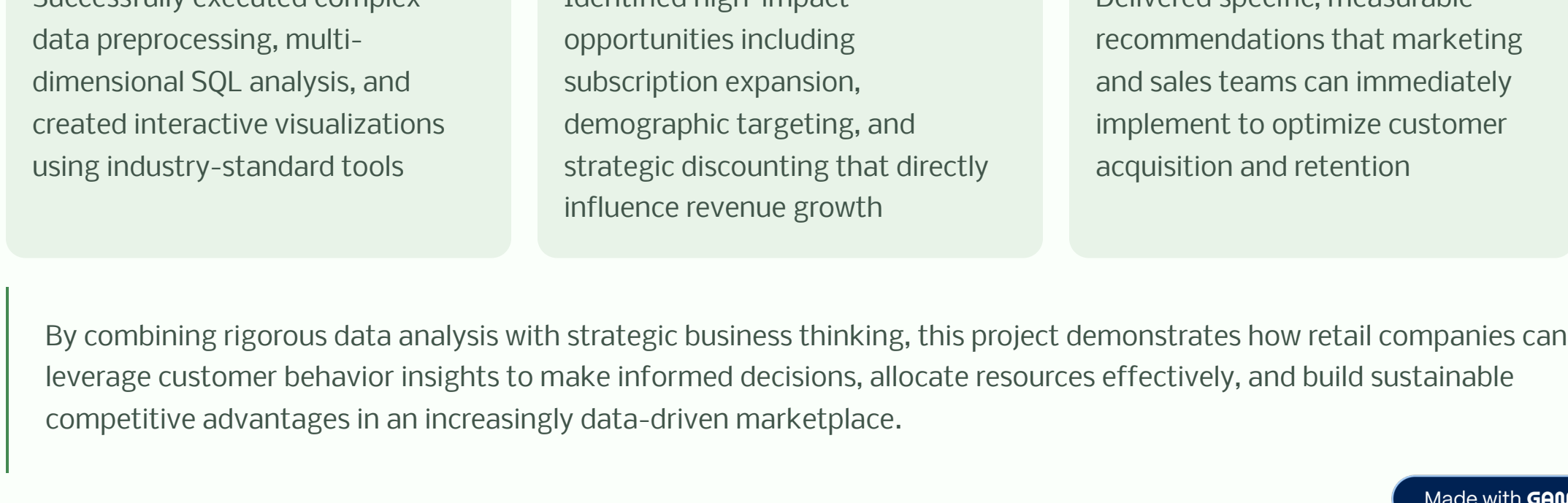
Age Demographics and Revenue Contribution

```
SELECT age_group,
       SUM(purchase_amount) AS total_revenue
FROM people
GROUP BY age_group
ORDER BY total_revenue DESC;
```

Seniors (60+) Highest revenue contributors with premium purchasing power	Middle-Aged (40-59) Strong secondary segment with established buying patterns
Adults (25-39) Growing segment with increasing purchasing capacity	Young Adults (18-24) Emerging market with future growth potential

Strategic Implication: Marketing budgets should prioritize seniors and middle-aged demographics for immediate revenue impact while cultivating younger segments through brand-building initiatives that will pay dividends as their purchasing power increases.

Power BI Dashboard Overview



The interactive Power BI dashboard consolidates all key metrics into a single view, enabling stakeholders to explore customer behavior patterns, filter by demographics, and drill down into product performance. The visualization layer transforms complex SQL query results into actionable insights that drive decision-making across marketing, operations, and product teams.

Strategic Business Insights

01 Subscription Model Success Subscribed customers demonstrate significantly higher lifetime value through increased purchase frequency and larger transaction sizes, making subscription growth the highest-impact revenue lever	02 Clothing Category Leadership Apparel items dominate revenue generation across all customer segments, warranting continued inventory investment and marketing focus in this category
03 Demographic Opportunity Seniors and middle-aged customers contribute the largest revenue share, suggesting targeted campaigns for these segments will deliver maximum ROI	04 Smart Discounting High-value customers use discounts without reducing overall spend, indicating strategic promotions can drive volume without margin erosion
05 Premium Shipping Value Express shipping correlates with higher spending behavior, suggesting these customers represent a premium segment worth retaining through service excellence	06 Retention Focus Returning customers dominate the base, presenting opportunities to strengthen loyalty programs and convert them into advocates

Actionable Business Recommendations

1 Aggressive Subscription Promotion Launch a targeted campaign to convert non-subscribed returning customers, offering first-month incentives and highlighting subscriber-exclusive benefits. Goal: Increase subscription rate from 67% to 75% within six months.	2 Top-Rated Product Marketing Feature highest-rated products prominently in all marketing channels, leveraging customer reviews and ratings as social proof to drive conversion and build trust with new customers.
3 Personalized Discount Strategy Implement tiered discount programs based on customer segmentation, offering exclusive promotions to repeat buyers while maintaining full-price positioning for new customers to maximize margin.	4 Age-Targeted Campaigns Allocate 60% of marketing budget to senior and middle-aged demographics through channels they frequent, while establishing long-term brand presence among younger segments through digital platforms.
5 Express Shipping Optimization Maintain competitive express shipping rates for high-value customers while exploring partnerships with logistics providers to reduce costs and improve delivery speed, enhancing the premium customer experience.	

Project Impact and Methodology

This end-to-end analysis demonstrates a complete data analytics workflow applicable to real-world business challenges:

- Data Preprocessing:** Python-based cleaning and normalization ensuring analysis reliability
- Business Intelligence:** SQL queries extracting actionable insights from customer behavior patterns
- Visualization:** Power BI dashboards enabling stakeholder exploration and decision-making
- Strategic Translation:** Converting technical findings into business recommendations

The methodology mirrors professional data analyst workflows, translating raw transactional data into strategic insights that drive measurable business growth and competitive advantage.

Conclusion: Data-Driven Retail Excellence

Technical Mastery Successfully executed complex data preprocessing, multi-dimensional SQL analysis, and created interactive visualizations using industry-standard tools	Business Value Identified high-impact opportunities including subscription expansion, demographic targeting, and strategic discounting that directly influence revenue growth	Actionable Strategy Delivered specific, measurable recommendations that marketing and sales teams can immediately implement to optimize customer acquisition and retention
--	---	--

By combining rigorous data analysis with strategic business thinking, this project demonstrates how retail companies can leverage customer behavior insights to make informed decisions, allocate resources effectively, and build sustainable competitive advantages in an increasingly data-driven marketplace.