

Customer Shopping Behavior: Insights & Strategic Opportunities

- **Scope:** Analyzed 3,900 transaction records to decode specific spending patterns and product preferences.
- **Methodology:** Executed an end-to-end data pipeline using Python (cleaning), SQL (analysis), and Power BI (visualization).
- **Key Insight:** Optimizing 'Loyal' segments and subscription models offers the highest ROI leverage.
- **Strategic Goal:** Shift from reactive reporting to proactive decisions in marketing and inventory management.

Total Customers

3,900

Variables Tracked

18

Data Health

**Imputed
Missing
Values**

Primary Tools

**Python &
SQL**

Project Objectives & Strategic Value



Goal: Uncover hidden trends in demographics, product preferences, and subscription utility to **guide strategy**.

Dataset: Comprehensive review of **18 variables** including demographics, purchase details, and shipping methods.

Business Value: Validates data integrity to **support decisions** on inventory positioning and customer lifecycle management.

Dataset Profile & Feature Distribution

Who (Customer)

- Age
- Gender
- Location
- Subscription Status

What (Product)

- Item Purchased
- Category
- Size
- Color
- Review Rating

How (Transaction)

- Purchase Amount
- Season
- Shipping Type
- Discount Applied
- Promo Code Used
- Payment Method
- Frequency of Purchases

- **Volume:** 3,900 Rows, 18 Columns (robust sample size for trend analysis).
- **Key Dimensions:** Demographics (Age/Gender/Location) and Transactional (Spend/Season/Shipping).
- **Data Health:** High integrity dataset; identified only 37 missing values in 'Review Rating' (imputed).

Analytical Workflow: From Raw Data to Insights



An end-to-end pipeline ensures data accuracy, scalability, and reproducible insights.

Ensuring Data Integrity & Usability (Python)

Before: Raw Data

	Item Purchased	Review Rating	Promo Code Used
1	Blouse	null	Yes
2	Jeans	3.5	No

Issues: Missing ratings, redundant columns, inconsistent casing.

After: Cleaned & Engineered

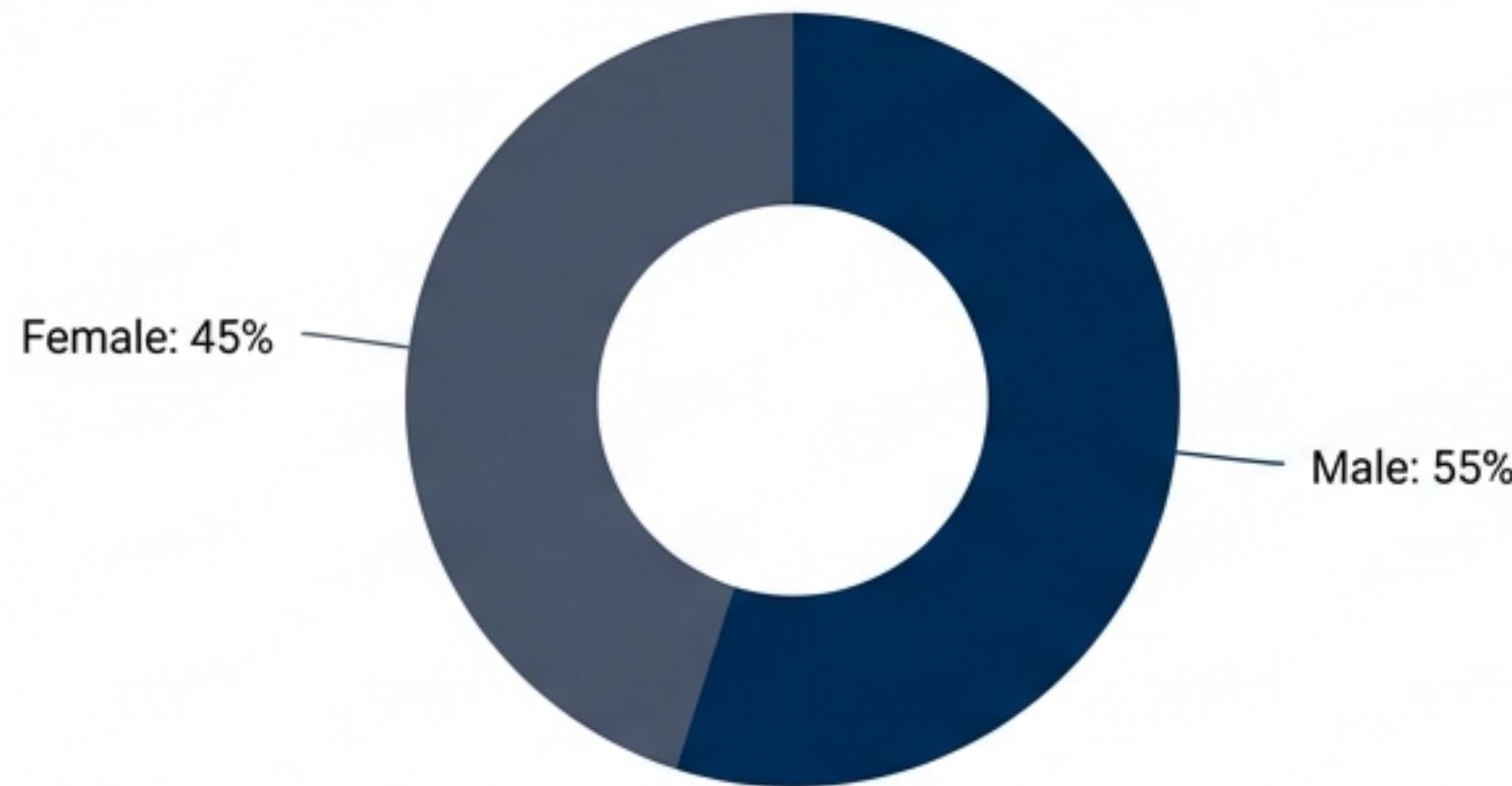
	item_purchased	review_rating	age_group
1	blouse	3.9 (Imputed)	20-30
2	jeans	3.5	30-40

Improvements: standardized snake_case, nulls filled with category median, new 'age_group' feature created.

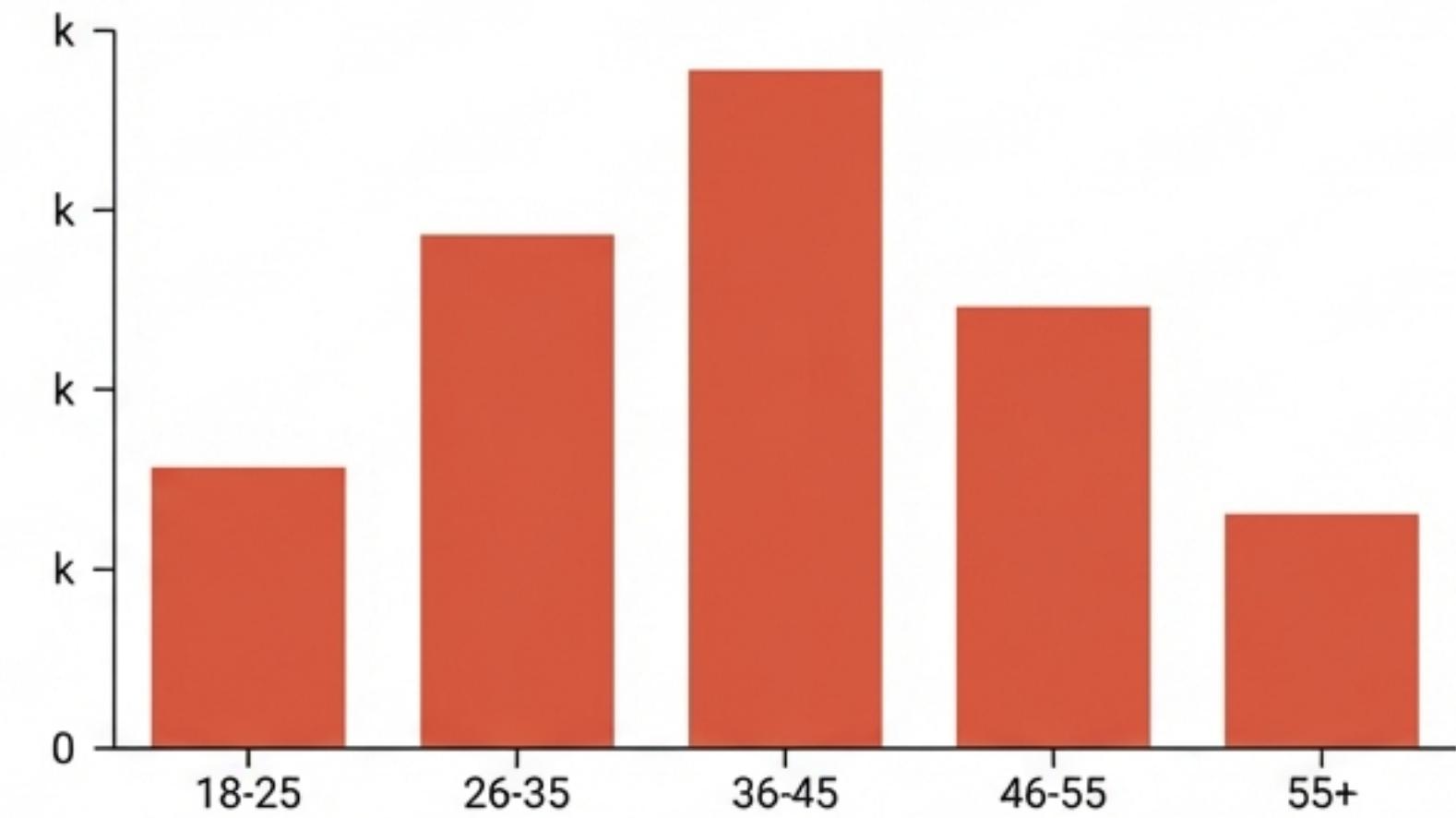
- **Standardization:** Renamed columns to `snake_case` for consistency; removed redundant fields.
- **Imputation:** Filled 37 missing 'Review Ratings' using category medians to preserve data volume.
- **Feature Engineering:** Created 'age_group' and 'purchase_frequency_days' for deeper segmentation.

Revenue Drivers: Demographics & Age Groups

Total Revenue by Gender



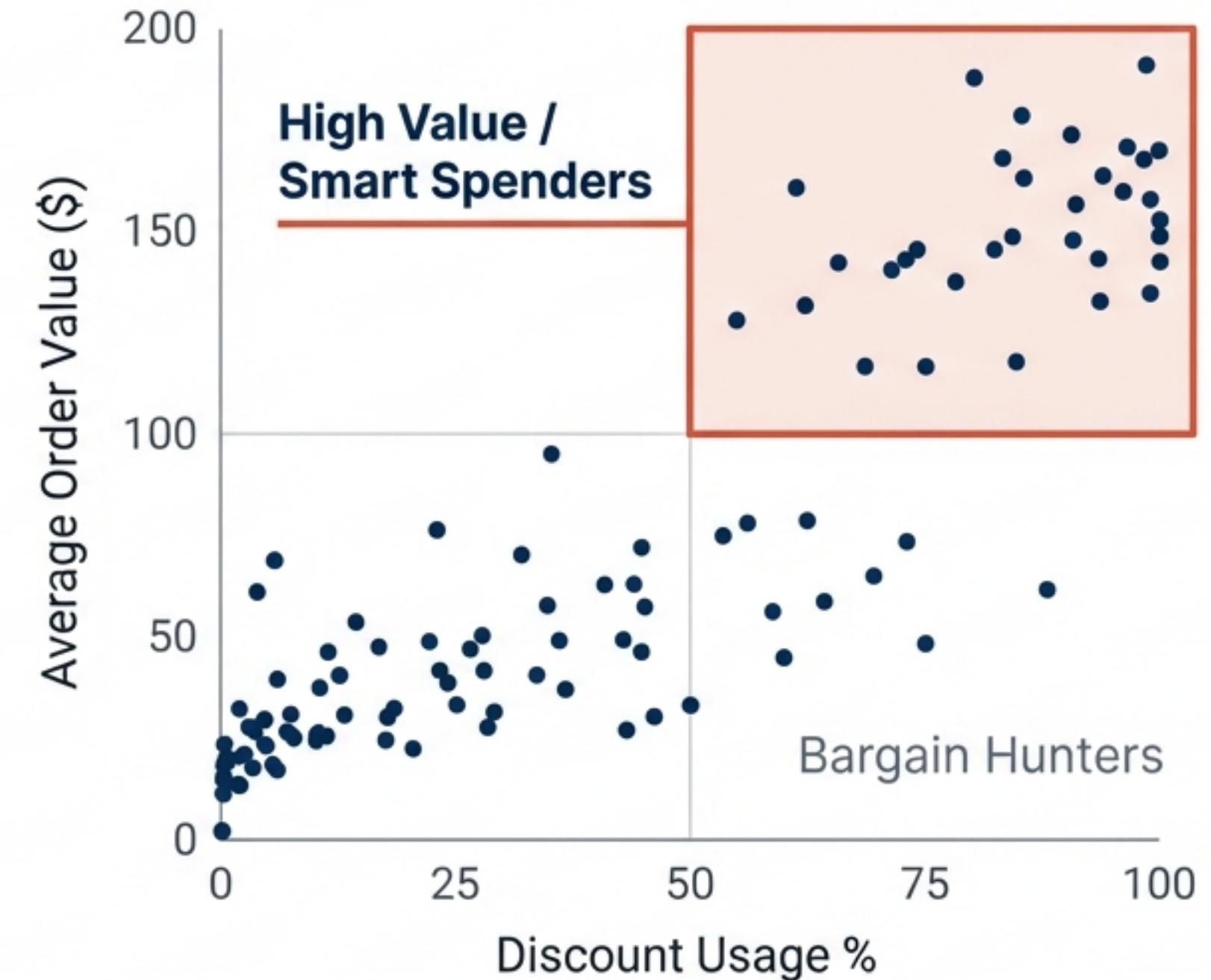
Revenue by Age Group



- **Gender Analysis:** Comparative view of total revenue contributions.
- **Generational Spend:** Revenue broken down by the newly engineered 'age_group' feature.
- **Strategic Fit:** Identifies which demographic cohorts yield the highest ROI for ad spend.

Discount Sensitivity & High-Value Users

- **High-Spenders:** Identified users leveraging discounts while maintaining **above-average** basket sizes.
- **Product Dependency:** Isolated **top 5 products** with the highest percentage of discounted sales.
- **Impact:** Distinguishes between '**bargain hunters**' and '**smart high-spenders**' to refine discount strategies.



Product Excellence: Ratings & Category Leaders



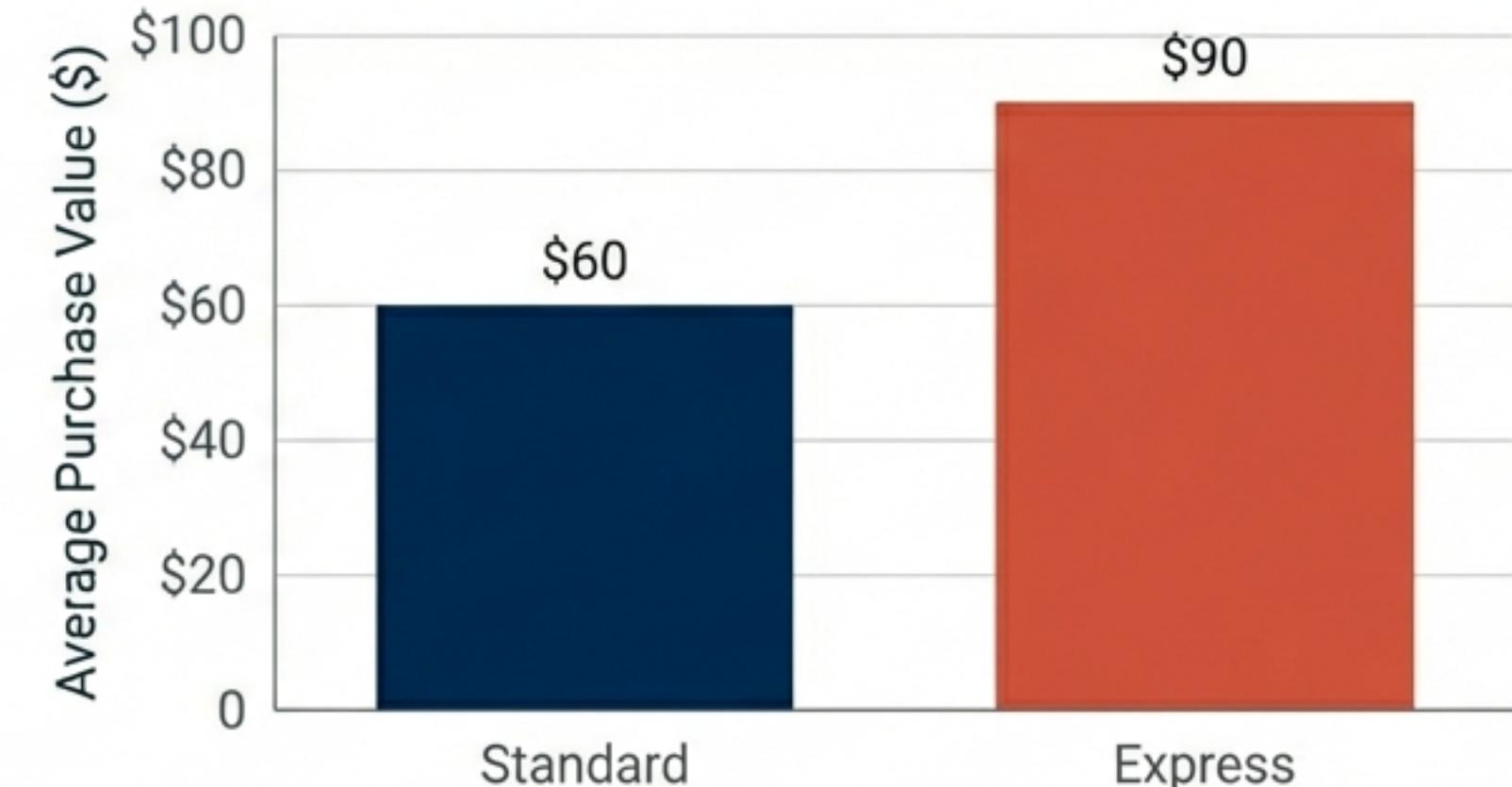
- **Quality Leaders:** Top 5 products identified by highest average customer review ratings.
- **Inventory Focus:** Highlights successful SKUs to feature in upcoming campaigns.

Logistics & Subscription Economics

Average Spend: Subscribers vs. Non-Subscribers



Avg Purchase Value: Standard vs. Express Shipping



- **Shipping:** Analyzed purchase value comparison between 'Standard' vs. 'Express' users.
- **Subscription Value:** Confirmed revenue premium of Subscribers over Non-Subscribers.
- **Retention Link:** Strong correlation found between repeat buyers (>5 purchases) and subscription likelihood.

Segmentation: Defining the Customer Lifecycle



The Framework:
Classified database
based on purchase
history and frequency.

Logic: Differentiates
casual shoppers from
high-LTV advocates.

Application: Enables
tailored messaging (e.g.,
Welcome Series vs. VIP
Rewards) rather than
generic blasts.

Strategic Roadmap & Key Actions

Action Priority Matrix



Executive Takeaway: By leveraging data-driven segmentation and optimizing for subscription conversion, we can significantly increase Customer Lifetime Value (LTV) while reducing reliance on margin-eroding discounts.

Top 3 Actions

Launch Loyalty Tier: Reward the 'Loyal' segment to prevent churn.

Upsell Subscriptions: Target repeat buyers (>5 purchases).

Optimize Inventory: Reposition 'Top Rated' products in campaigns.