Customer Shopping Behavior Analysis

This project analyzes 3,900 customer purchases to uncover insights into spending patterns, product preferences, and subscription behavior. Our goal is to guide strategic business decisions.

Dataset Summary

Our dataset comprises 3,900 rows and 18 columns, detailing customer demographics, purchase specifics, and shopping behaviors.

- **Customer Demographics:** Age, Gender, Location, Subscription Status
- **Purchase Details:** Item, Category, Amount, Season, Size, Color
- **Shopping Behavior:** Discount, Promo Code, Previous Purchases, Frequency, Review Rating, Shipping Type

Missing Data: 37 values in the Review Rating column were imputed using the median rating per product category.

Exploratory Data Analysis (Python)

We initiated our analysis with data preparation and cleaning using Python, ensuring data quality and consistency.

Data Loading & Exploration

Imported data with `pandas`, checked structure with `df.info()`, and summarized statistics with `.describe()`.

Data Cleaning

Handled 37 missing `Review Rating` values by imputing with median per category. Standardized column names to `snake_case`.

Feature Engineering

Created `age_group` and `purchase_frequency_days`. Removed redundant `promo_code_used` column.

Database Integration

Loaded the cleaned DataFrame into PostgreSQL for advanced SQL analysis.

SQL Analysis: Revenue & Discounts

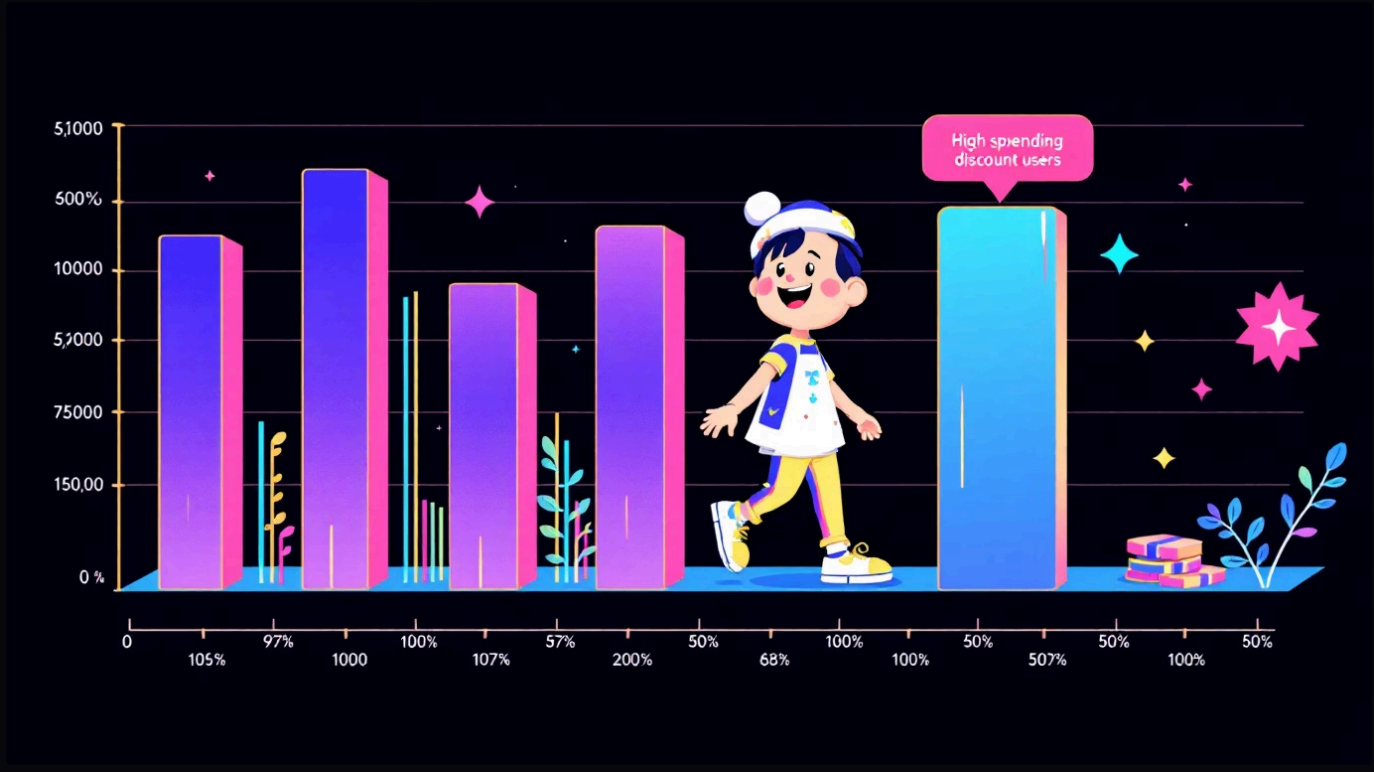
Revenue by Gender

Male customers generated significantly more revenue than female customers.

Female	75,191
Male	157,890

High-Spending Discount Users

Identified 839 customers who used discounts but still spent above the average purchase amount, indicating effective discount strategies.



SQL Analysis: Product & Shipping

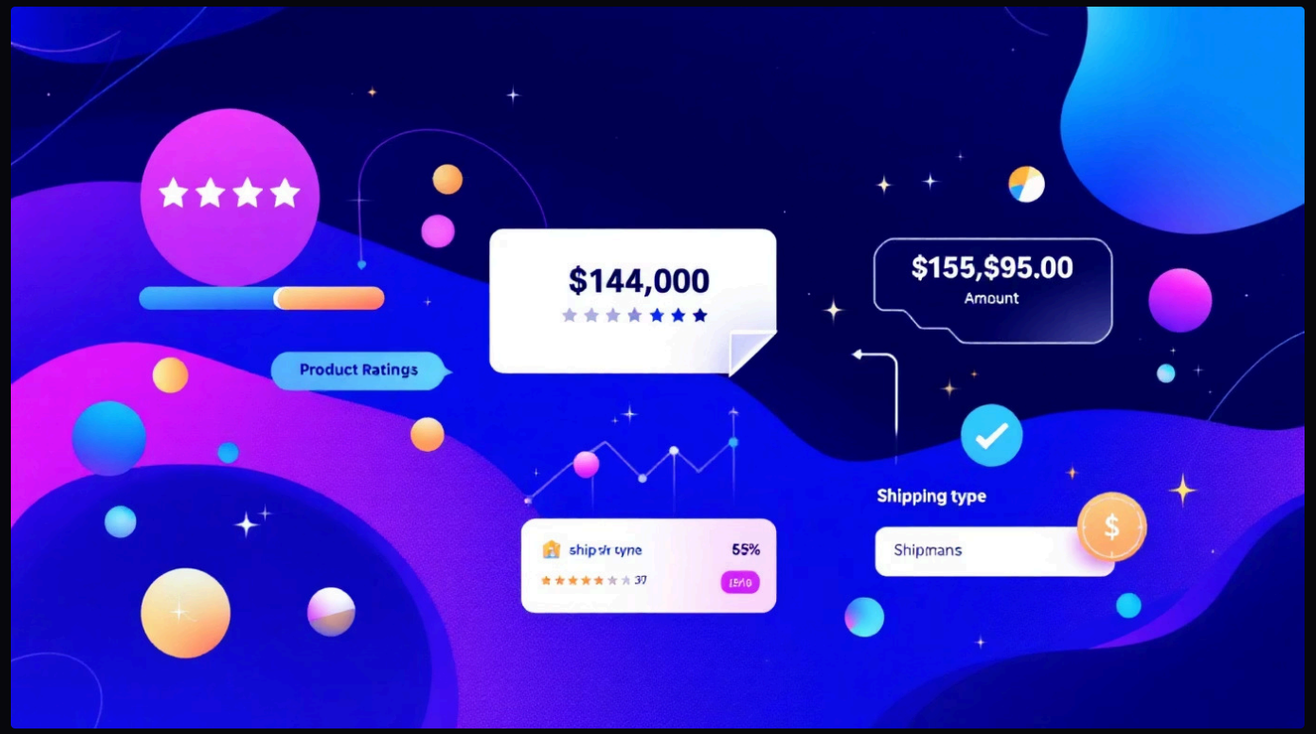
Top 5 Products by Rating

Gloves	3.86
Sandals	3.84
Boots	3.82
Hat	3.80
Skirt	3.78

Shipping Type Comparison

Express shipping correlates with higher average purchase amounts, suggesting customers are willing to pay more for faster delivery.

Standard	58.46
Express	60.48



SQL Analysis: Subscriptions & Discounts

Subscribers vs. Non-Subscribers

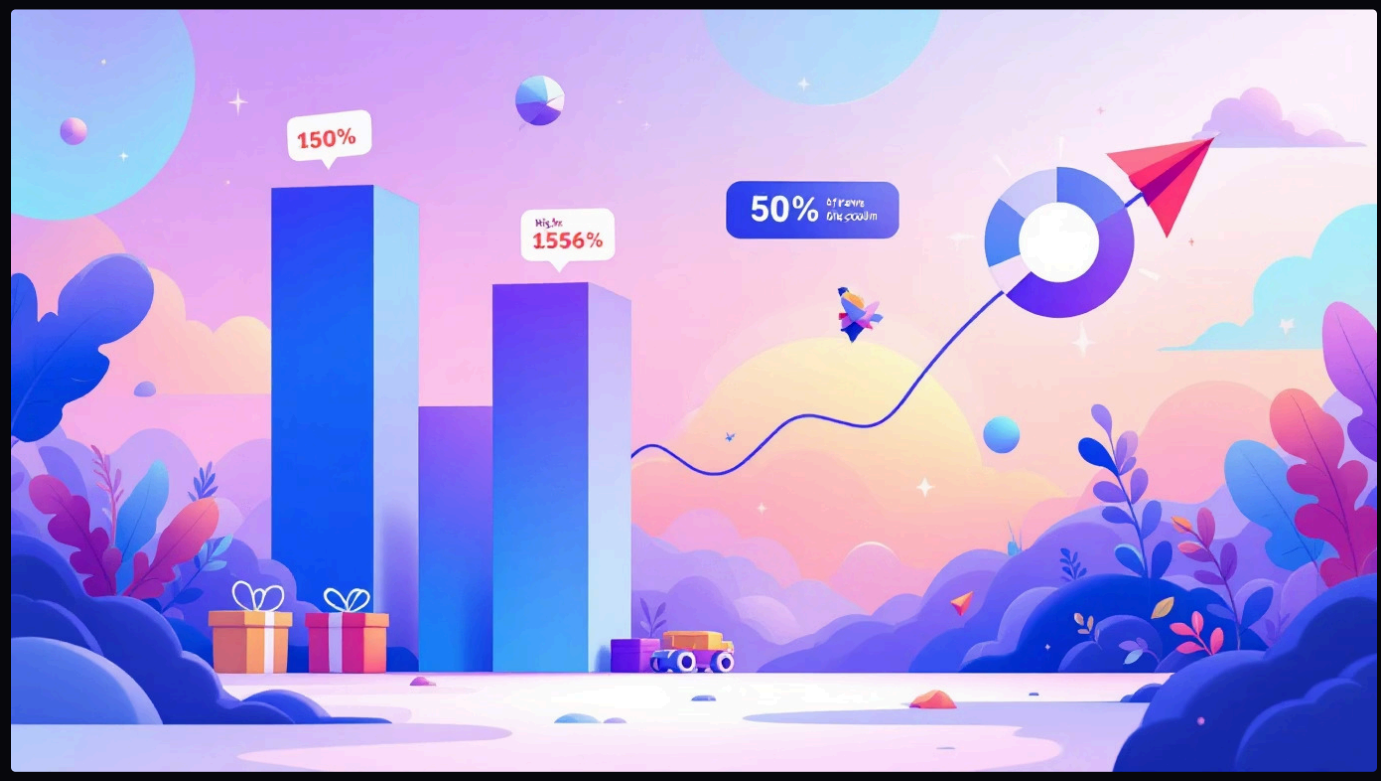
Non-subscribers contribute more to total revenue, but subscribers show consistent engagement.

Yes	1,053	59.49	62,645
No	2,847	59.87	170,436

Discount-Dependent Products

Products like Hats and Sneakers show a high percentage of discounted purchases, indicating their sensitivity to promotions.

Hat	50.00
Sneakers	49.66
Coat	49.07
Sweater	48.17
Pants	47.37



SQL Analysis: Customer Segmentation

Customer Segments

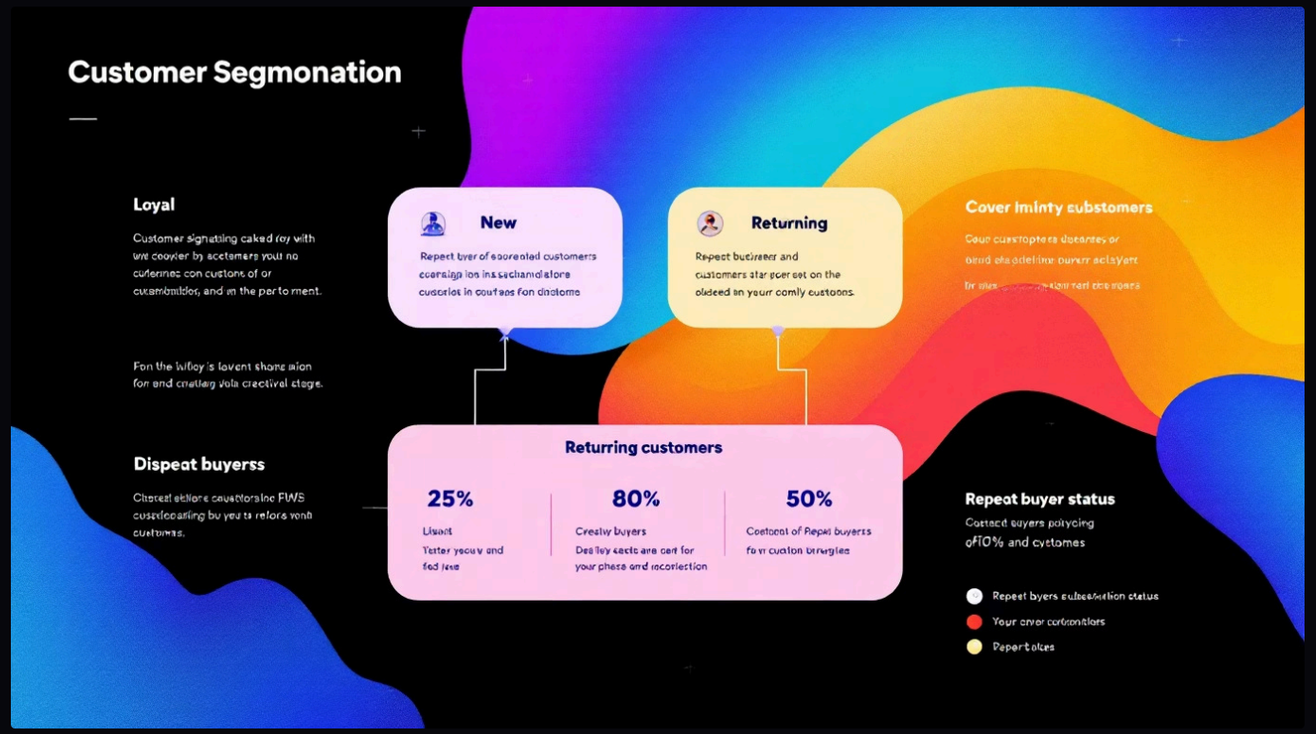
Customers are segmented into Loyal, New, and Returning based on their purchase history, revealing a strong base of loyal customers.

Loyal	3,116
New	83
Returning	701

Repeat Buyers & Subscriptions

Repeat buyers (more than 5 purchases) are more likely to subscribe, indicating a link between loyalty and subscription potential.

No	2,518
Yes	958



SQL Analysis: Top Products & Revenue by Age

Top 3 Products per Category

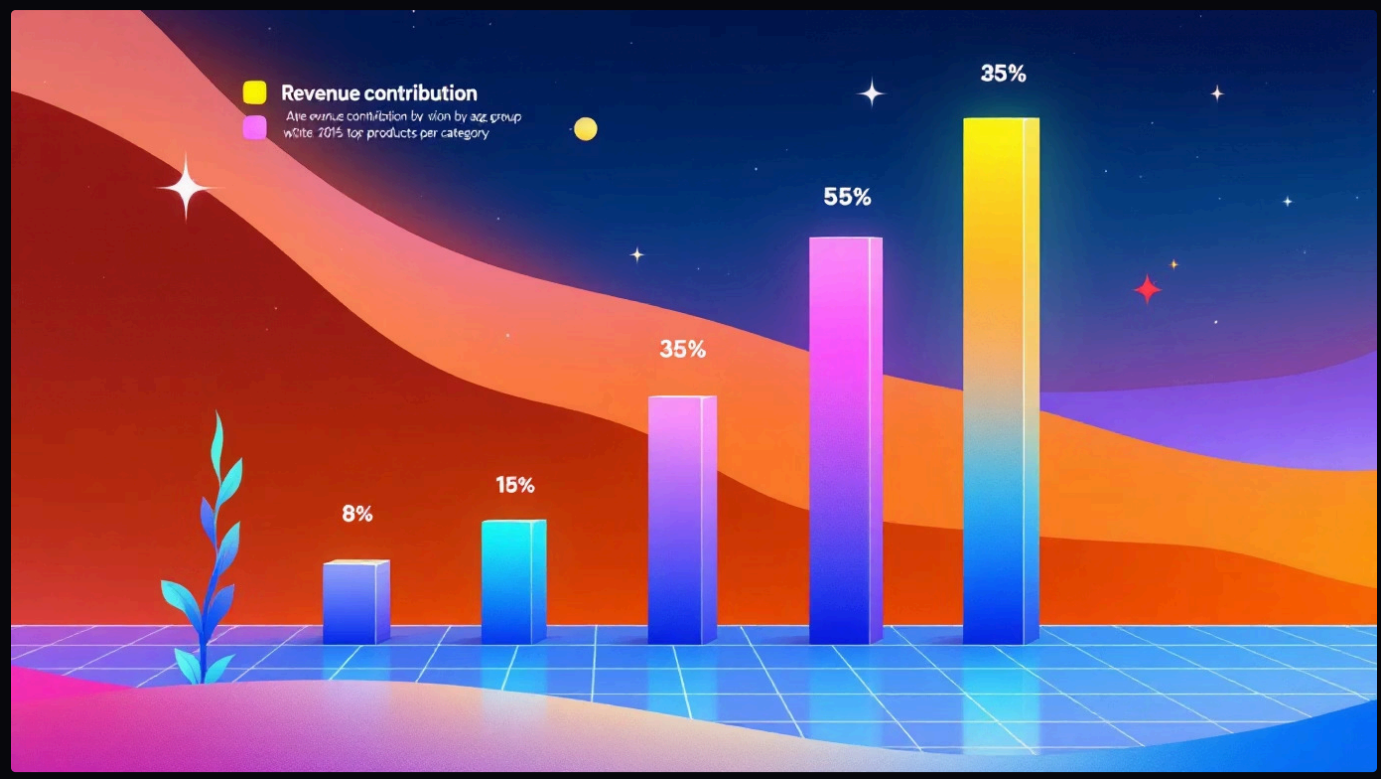
Identified best-selling items within each category, highlighting popular choices like Jewelry, Blouse, and Sandals.

Accessories	Jewelry	171
Clothing	Blouse	171
Footwear	Sandals	160

Revenue by Age Group

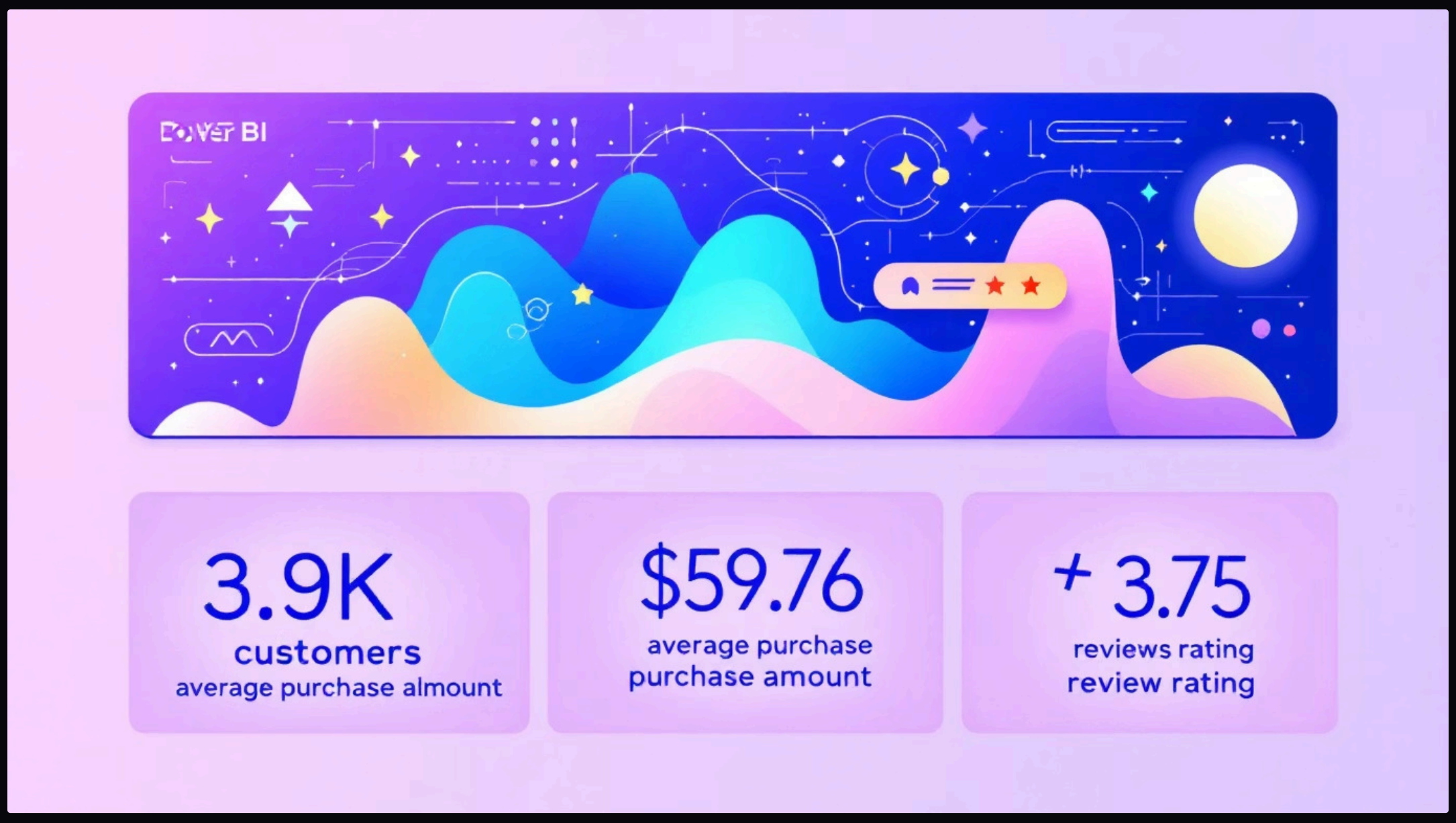
Young Adults contribute the highest revenue, followed by Middle-aged customers, indicating key demographic targets.

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



Interactive Dashboard in Power BI

An interactive Power BI dashboard was developed to visualize key insights, offering dynamic filtering and drill-down capabilities for deeper exploration.



Business Recommendations

Boost Subscriptions

Promote exclusive benefits to increase subscriber base.

Customer Loyalty Programs

Reward repeat buyers to foster long-term loyalty.

Review Discount Policy

Optimize discount strategies to balance sales and profit margins.

Targeted Marketing

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