

Customer Shopping Behavior Analysis

Uncovering strategic insights from transactional data to guide business decisions.

PROJECT OVERVIEW

Analyzing 3,900 Purchases



Transactional Data

Analyzed 3,900 purchases across various product categories.



Key Goals

Uncover spending patterns, customer segments, and product preferences.



Business Impact

Guide strategic decisions on subscriptions, loyalty, and marketing.

Dataset and Key Features

3,900

Total Rows

Number of purchases analyzed.

18

Total Columns

Features including demographics and purchase details.

- ❑ Missing Data: 37 values were missing in the Review Rating column.

Key Feature Categories

- **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.

Data Preparation and Cleaning (Python)



Data Loading & Exploration

Imported dataset using pandas and checked structure with `df.info()`.



Missing Data Handling

Imputed 37 missing Review Rating values using the median rating per product category.



Standardization & Engineering

Renamed columns to snake case. Created `age_group` and `purchase_frequency_days` features.



Database Integration

Loaded the cleaned DataFrame into MySQL for structured SQL analysis.

SQL Analysis: Revenue and Spending

The screenshot shows a 'Result Grid' interface with two columns: 'gender' and 'total_revenue'. There are three rows: a header row, a row for 'Male' with a value of 157890, and a row for 'Female' with a value of 75191. A 'Filter Rows:' button is visible at the top right.

	gender	total_revenue
▶	Male	157890
	Female	75191

Revenue by Gender: Compared total revenue generated by male vs. female customers.

The screenshot shows a 'Result Grid' interface with four columns: 'subscription_status', 'total_customers', 'average_spend', and 'total_revenue'. There are two rows: one for 'Yes' with values 1053, 59.49, and 62645 respectively, and one for 'No' with values 2847, 59.87, and 170436 respectively. A 'Filter Rows:' button is visible at the top right.

	subscription_status	total_customers	average_spend	total_revenue
▶	Yes	1053	59.49	62645
	No	2847	59.87	170436

Subscribers vs. Non-Subscribers: Compared average spend and total revenue across subscription status.

Analysis also identified customers who used discounts but still spent above the average purchase amount (e.g., Customer 4 spent \$90).

SQL Analysis: Product and Shipping Insights

Top-Rated Products

Identified products with the highest average review ratings, crucial for positioning.

Result Grid		Filter Rows:
	item_purchased	avg_review_rating
▶	Gloves	3.86
	Sandals	3.84
	Boots	3.82
	Hat	3.8
	Skirt	3.78

Shipping Type Impact

Compared average purchase amounts between Standard and Express shipping types.

Result Grid			Filter Rows:
	shipping_type	avg_amount	
▶	Express	60.48	
	Standard	58.46	

Discount Dependency

Found 5 products with the highest percentage of discounted purchases, indicating price sensitivity.

Result Grid			Filter Rows:
	item_purchased	discount_rate	
▶	Hat	50.00	
	Sneakers	49.66	
	Coat	49.07	
	Sweater	48.17	
	Pants	47.37	

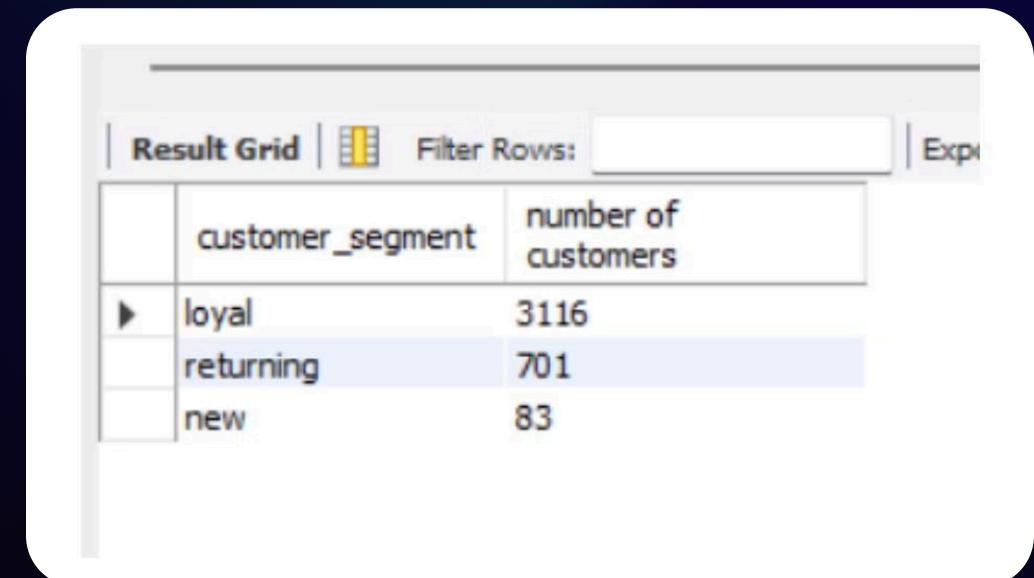
Customer Segmentation and Loyalty

Segmentation by Purchase History

Customers were classified into three segments based on their purchase history:

- **New**
- **Returning**
- **Loyal**

This allows for tailored marketing and loyalty programs.



A screenshot of a data visualization tool showing a result grid. The grid has two columns: 'customer_segment' and 'number of customers'. The data is as follows:

	customer_segment	number of customers
▶	loyal	3116
	returning	701
	new	83

Age Group Revenue and Subscription Propensity



Revenue by Age Group

Calculated total revenue contribution for each age group to identify high-value demographics.

Result Grid | Filter Rows:

	age_group	revenue
▶	Young Adult	62143
	Middle-Aged	59197
	Adult	55978
	Senior	55763



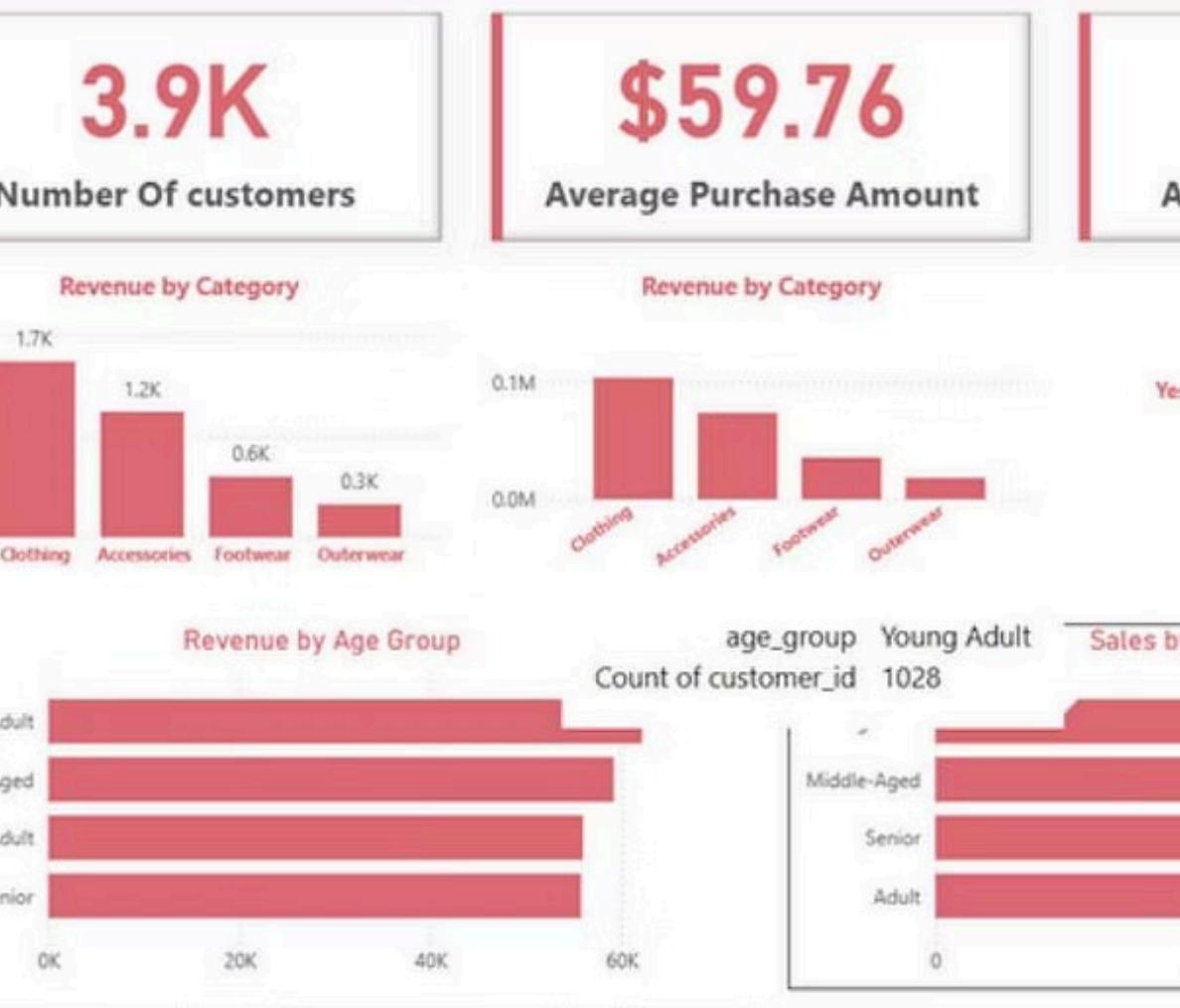
Repeat Buyers & Subscriptions

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

Result Grid | Filter Rows:

	subscription_status	repeat_buyers
▶	Yes	980
	No	2583

Customer Behavior Dashboard



Visualizing Insights in Power BI

An interactive dashboard was created in Power BI to present the key findings visually and allow for deeper exploration of the data.

The dashboard integrates all key metrics, including revenue, category performance, and customer segmentation.

Strategic Business Recommendations

→ **Boost Subscriptions**

Promote exclusive benefits and rewards to increase subscriber base and retention.

→ **Customer Loyalty Programs**

Reward repeat buyers to encourage progression into the highly valuable "Loyal" segment.

→ **Review Discount Policy**

Carefully balance sales boosts from discounts with the need for margin control.

→ **Targeted Marketing**

Focus marketing efforts on high-revenue age groups and customers who prefer express shipping.

