

# CUSTOMER SHOPPING BEHAVIOR ANALYSIS

3,900 purchases — 18 features — goal: uncover spending, segments, preferences, subscriptions





# DATASET SUMMARY

## ROWS & COLUMNS

3,900 rows · 18 columns

## KEY FEATURES

Demographics, purchase details,  
behavior, shipping

## MISSING DATA

37 missing review ratings



# EXPLORATORY DATA ANALYSIS (PYTHON)

## LOAD & INSPECT

pandas: df.info(), .describe()

## CLEAN & IMPUTE

Imputed review\_rating by category median

## ENGINEER FEATURES

age\_group, purchase\_frequency\_days

## DB INTEGRATION

Loaded cleaned DF into PostgreSQL



# REVENUE BY GENDER

## MALE

\$157,890 total revenue

## FEMALE

\$75,191 total revenue



# DISCOUNTS & HIGH-SPENDING USERS

839 rows of discount users identified who still spent above average (examples: customer\_id 2→64, 3→73, 4→90...)

## DISCOUNT-DEPENDENT PRODUCTS

Top five: Hat (50%), Sneakers (49.66%), Coat (49.07%), Sweater (48.17%), Pants (47.37%)



# TOP PRODUCTS & RATINGS

## GLOVES

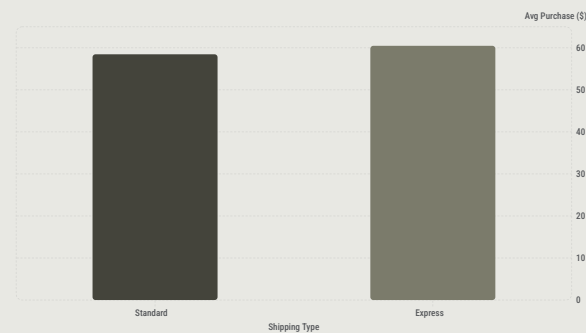
# SANDALS

## BOOTS





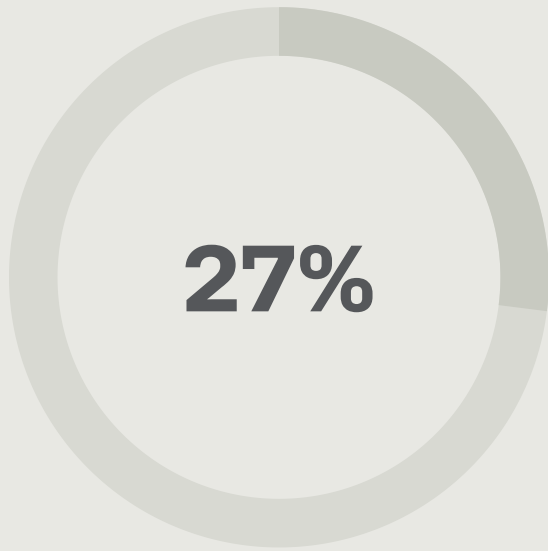
# SHIPPING TYPE COMPARISON



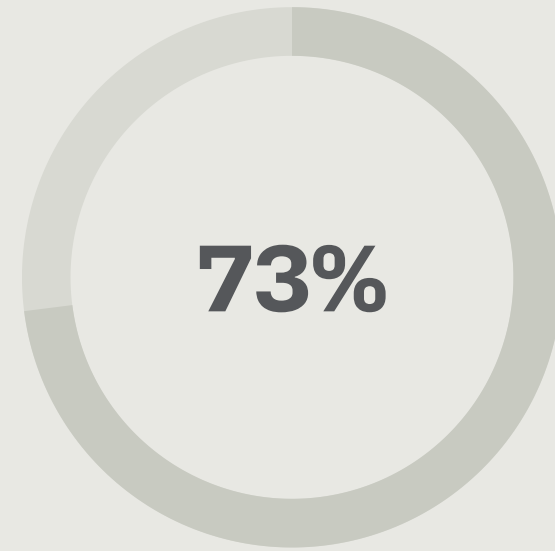
Express slightly higher avg spend  
(60.48) vs Standard (58.46)

Implication: shipping choice correlates  
with spend

# SUBSCRIPTIONS & SEGMENTS



Subscribers (Yes) – 1,053 customers · Avg spend \$59.49 · Revenue  
\$62,645

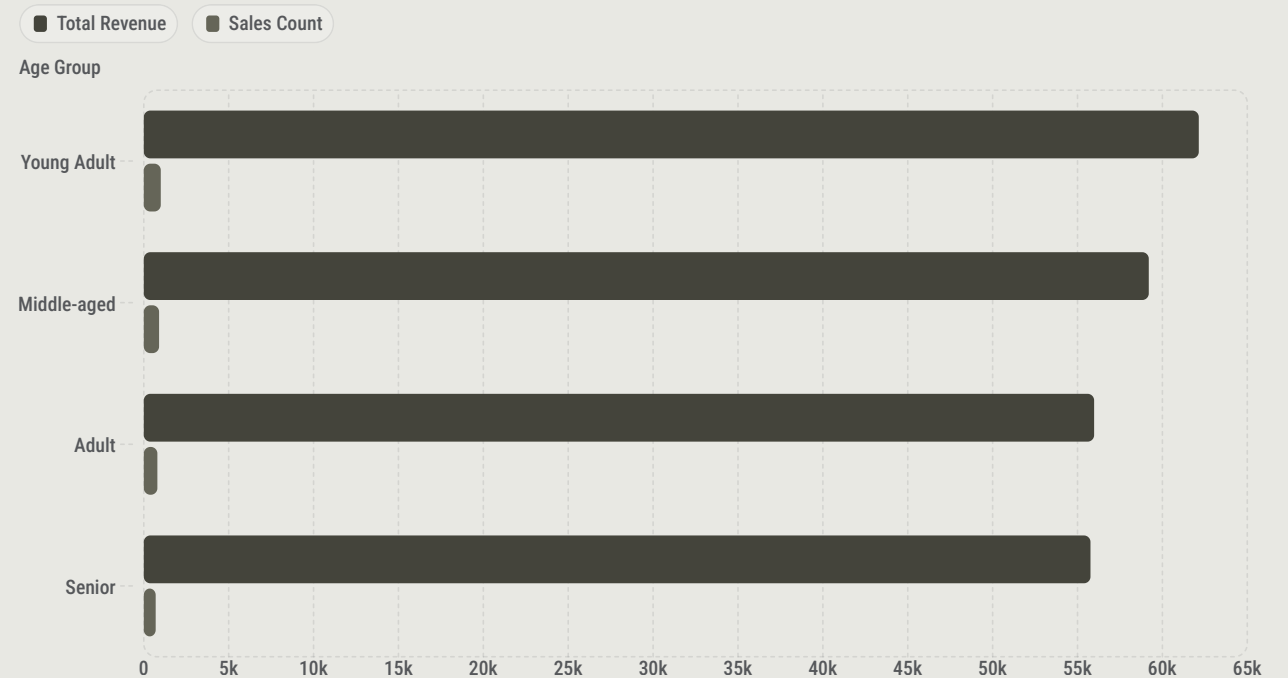


Non-subscribers – 2,847 customers · Avg spend \$59.87 · Revenue  
\$170,436

Customer segments: Loyal 3,116 · Returning 701 · New 83



# REVENUE & SALES BY AGE GROUP



Focus: Young Adult and Middle-aged segments contribute highest revenue and sales

# BUSINESS RECOMMENDATIONS

## **BOOST SUBSCRIPTIONS**

Promote exclusive subscriber benefits

## **LOYALTY PROGRAMS**

Reward repeat buyers to grow Loyal segment

## **DISCOUNT POLICY**

Balance volume gains with margin control

## **PRODUCT POSITIONING**

Highlight top-rated & best-selling items

## **TARGETED MARKETING**

Focus on high-revenue age groups & express-shipping users