

Exploratory Data Analysis

A Case Study in E-commerce Customer
Behavior

by: Jose Mosse

The "Why": What is EDA?

What is EDA?

Exploratory Data Analysis (EDA) is the crucial process of investigating datasets to summarize their main characteristics. It's how we uncover patterns, spot anomalies, test hypotheses, and check assumptions

before formal modeling begins.

Goal

To analyze 22,049 e-commerce transactions to understand customer demographics, identify key purchasing habits, and uncover the primary drivers of revenue. This process forms the foundation for actionable, data-driven strategy.

The Dataset: At a Glance



Volume

22,049 transactions were analyzed, providing a robust sample size for identifying significant trends.



Features

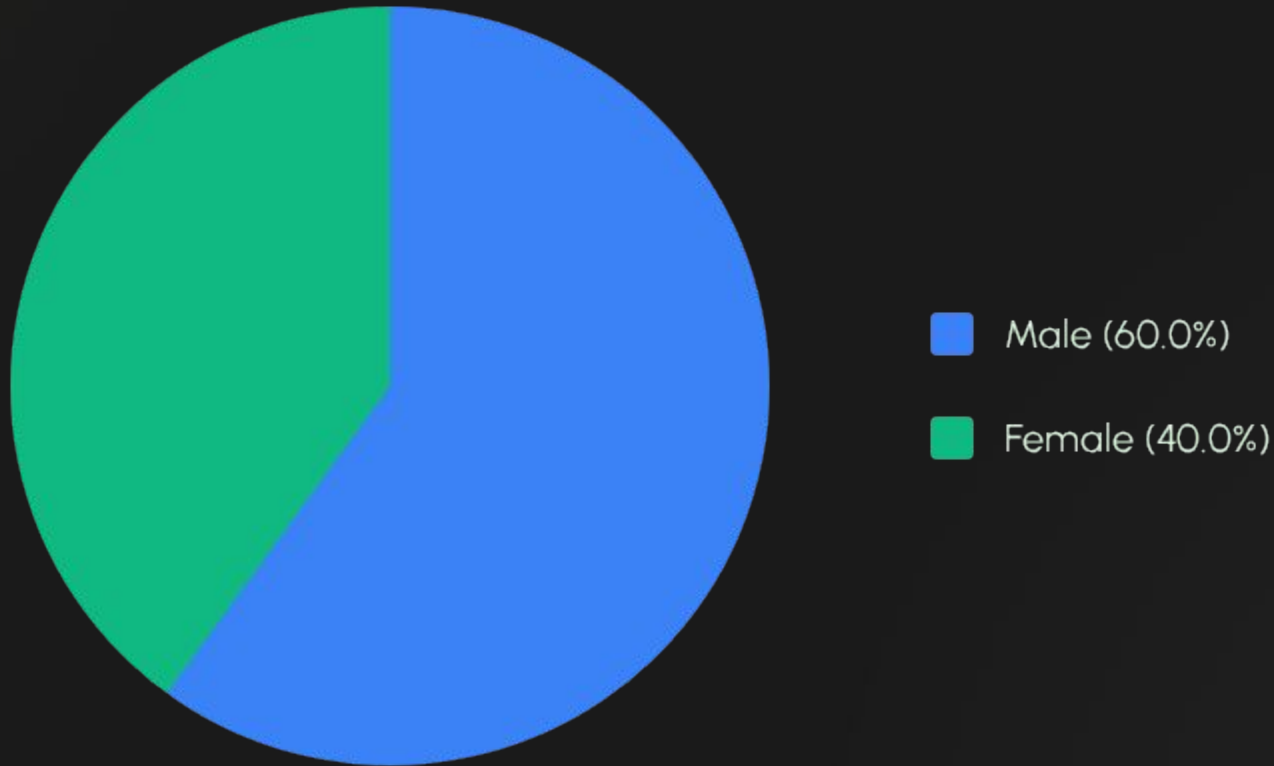
18 columns, including demographics (Age, Gender), order details (Category, Price), and behavior (Session Duration).



Data Quality

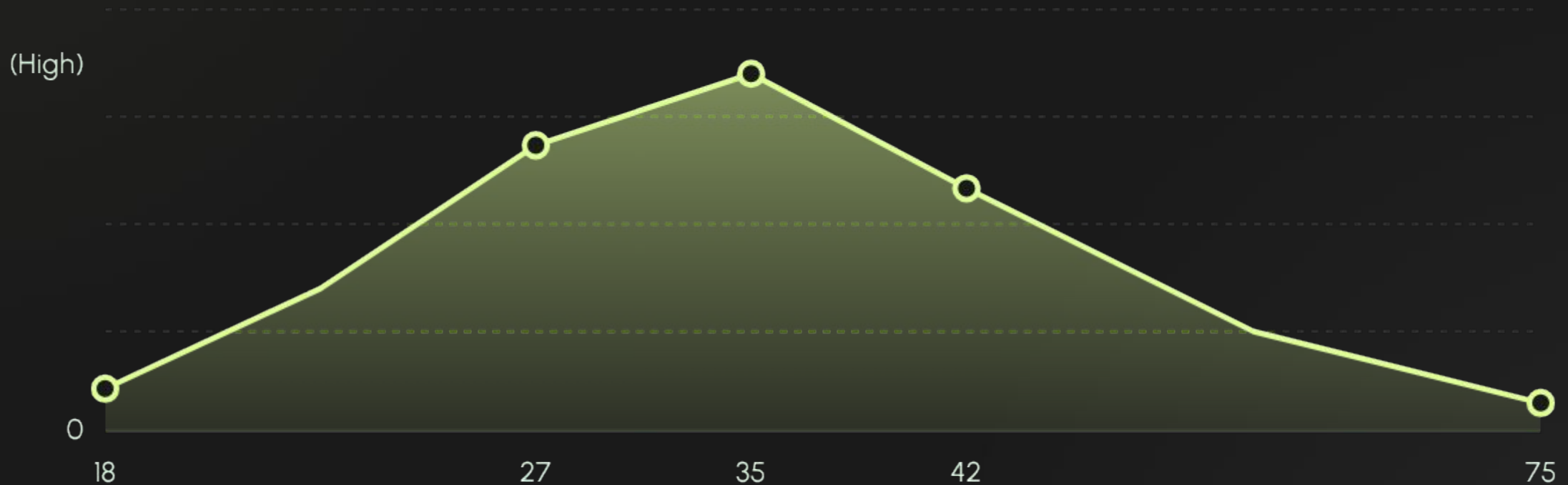
Initial checks confirmed 0 missing values. The data is clean, complete, and ready for analysis, ensuring reliable insights.

Customer Demographics: Gender



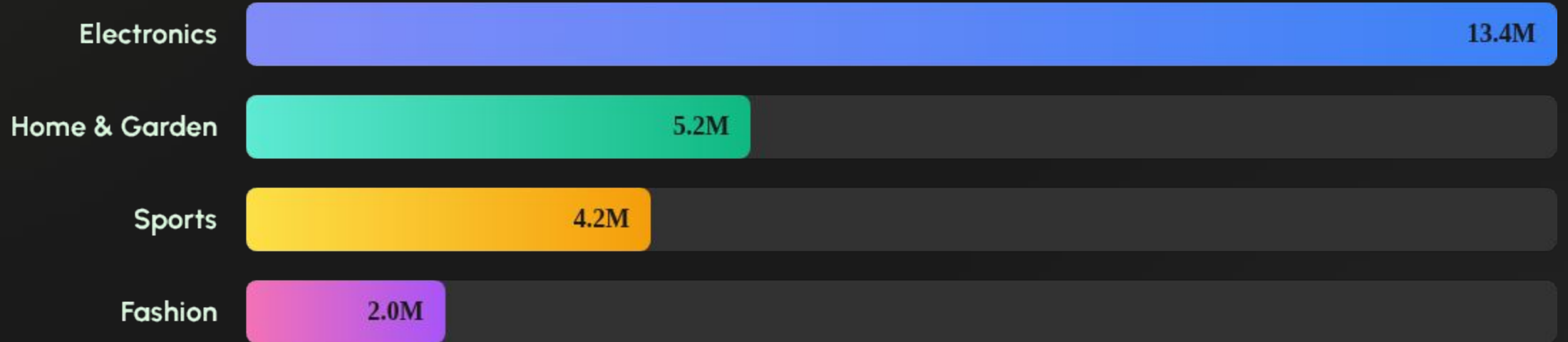
The customer base is predominantly male, accounting for 60% of all transactions analyzed.

Customer Demographics: Age Distribution



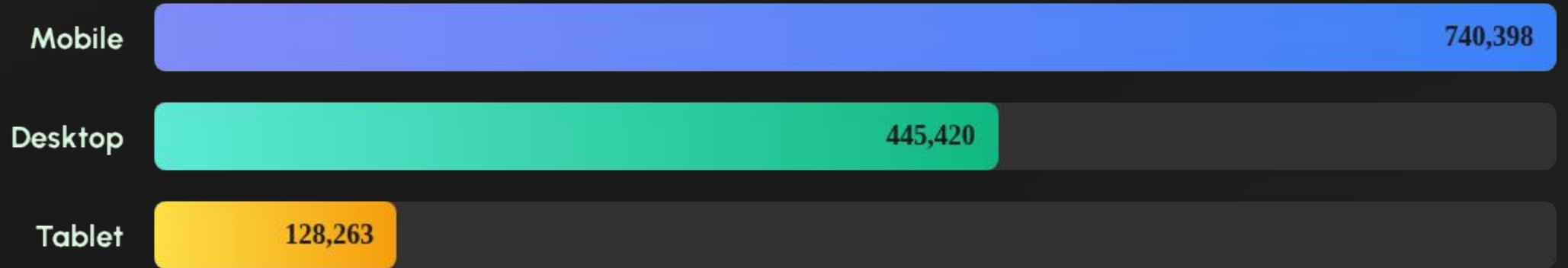
The customer base is primarily young to middle-aged. The median age is 35, with 75% of all customers being 42 or younger.

Key Finding 1: Revenue Drivers by Category



Electronics is the undisputed revenue leader, generating more than 2.5x the revenue of the next closest category, Home & Garden.

Key Finding 2: Discount Strategy by Device



The discount strategy is heavily focused on mobile users, who received 1.6x more in total discounts than desktop users.

Key Finding 3: Payment Method Popularity



Credit Card and Digital Wallet are the dominant payment methods. Cash on Delivery is extremely rare, suggesting a digitally-native customer base.

User Engagement & Loyalty

70%
Returning Customers




Strong Customer Loyalty

A significant 70% of all transactions are made by returning customers. This high loyalty rate indicates a strong product-market fit and a successful customer retention loop.

14.5 Mins

Customers are actively engaged, spending an average of 14.5 minutes per session and viewing an average of 9 pages.

Actionable Insights

-  **Focus on Electronics:** With Electronics driving 2.5x the revenue of the next category, this is the core growth engine. Prioritize inventory, marketing, and cross-selling accessories for these high-value items.
-  **Optimize the Mobile-First Strategy:** The heavy investment in mobile discounts is clear. This strategy should be paired with a best-in-class mobile app experience to maximize conversion and capitalize on the engaged mobile user base.
-  **Streamline Top Payment Methods:** Credit Card and Digital Wallet are virtually tied for #1. A frictionless, "one-click" checkout process for these two methods is essential to reduce cart abandonment.

Next Steps: Deeper Analysis

Deeper Dives

- Perform RFM (Recency, Frequency, Monetary) analysis to segment high-value customers.
- Analyze the impact of discount *percentage* on conversion rates and average order value.
- Map `Customer_Rating` against `Delivery_Time_Days` to find the key drivers of dissatisfaction.

Predictive Modeling

- Build a classification model to predict customer churn based on session data and purchase frequency.
- Create a recommendation engine to cross-sell products (e.g., "Sports" customers may also buy "Electronics").

Thank You

[linkedin.com/in/josemosse](https://www.linkedin.com/in/josemosse) | github.com/mossej154