EXPLORATORY DATA ANALYSIS OF AMAZON CONSUMER BEHAVIOR DATASET

Purchasing Patterns, Consumer Intent & User Engagement

TOOLS USED

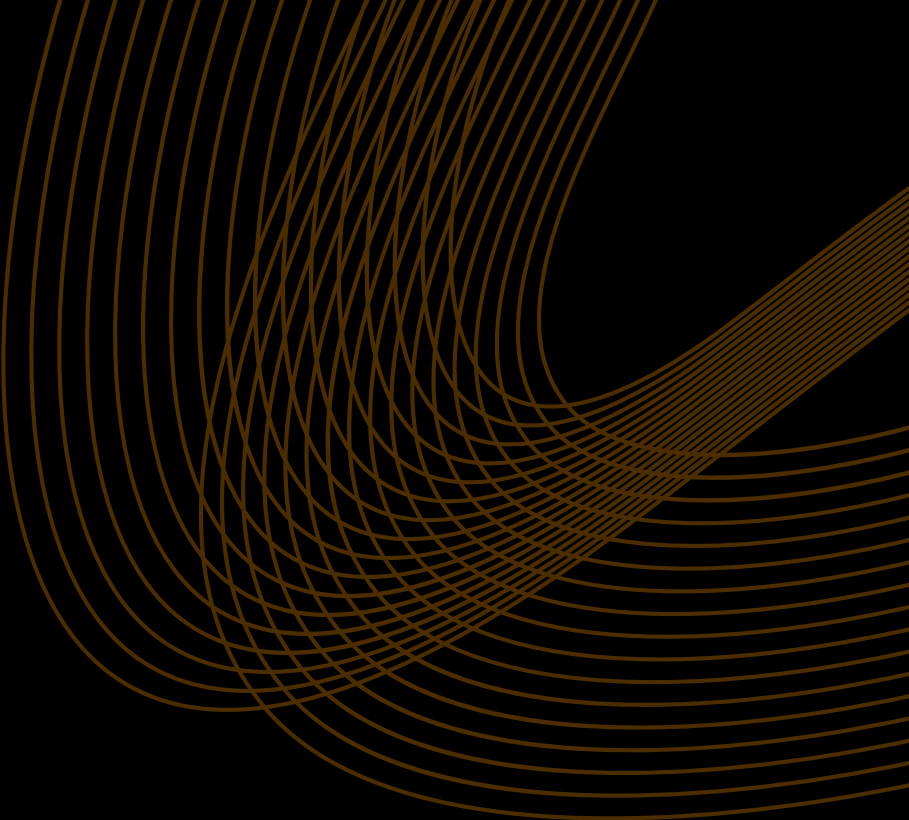Python, Pandas, Matplotlib, Seaborn & Plotly

PRESENTED BY

Tejas Jadhav

TABLE OF CONTENTS

PROBLEM STATEMENT & OBJECTIVE

Problem Statement:

The Amazon platform hosts millions of products across diverse categories. With massive variations in purchases, ratings & reviews, it becomes challenging to understand consumer trends, customer preferences & what drives product success.

Objective:

To explore the Amazon consumer behaviour dataset and uncover patterns, insights, and trends that help businesses and sellers make data driven decisions.

EDA WORKFLOW

For this analysis, a structured workflow was followed, involving data collection, understanding, cleaning, exploration, and summarization of insights to allow a clear understanding of the dataset and its trends.

01 Data Collection

- Gathered the Google Play Store dataset
  from Kaggle

02 Data Understanding & Anomaly Detection

- Looked at data distributions
- Found missing values, outliers, and unusual patterns

03 Data Cleaning & Treatment

- Fixed missing or incorrect values
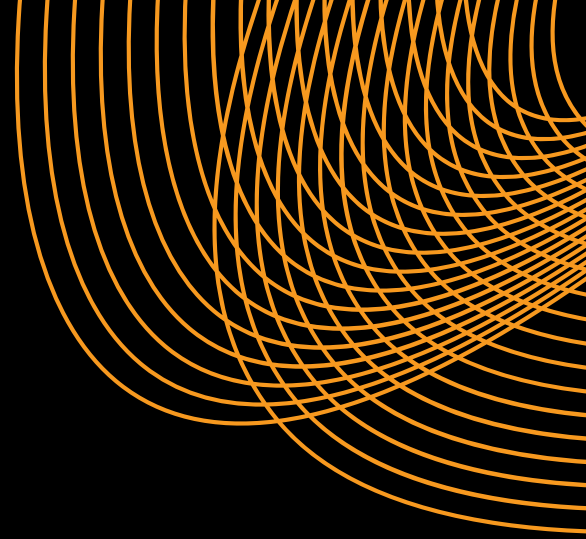- Standardized formats for consistency

04 Exploratory Analysis

- Univariate Analysis: e.g Purchase frequency, Search Method, Reviews
- **Bivariate Analysis:** Age vs Purchase Frequency
- Multivariate Analysis: Shopping Satisfaction vs Purchase Frequency vs Age Groups

05 Insights & Reporting

- Summarized patterns and trends
- Highlighted key findings for developers and businesses

KEY QUESTIONS EXPLORED

- How do user demographics (age and gender) influence shopping behavior and engagement patterns on the platform?

- Which stages of the shopping journey (browsing, cart, checkout) contribute most to drop-offs and missed conversions?

- Do occasional shoppers present a greater growth opportunity compared to already frequent buyers?

- How effective are product recommendations in influencing purchase decisions, and where do they fall short?

- What role does pricing perception play in cart abandonment and cross-platform comparison behavior?

- How do customer support responsiveness and delivery reliability impact overall satisfaction and trust?

- How do reviews and post-purchase experiences influence future purchase intent and platform credibility?

- Which satisfaction drivers most strongly contribute to repeat purchases and long-term loyalty?

DATA OVERVIEW

The dataset provides insights into Amazon consumer behavior, covering demographics, purchasing and browsing patterns, product search behavior, cart interactions, and the influence of reviews and personalized recommendations. It also captures overall shopping satisfaction and user feedback for service improvement.

Data Source: Kaggle

Dataset Size

602

Records

23

Features

Purchase Diversity

29

Product Categories

4

Search Methods

5

Satisfaction Indices

DATA OVERVIEW

Below is a detailed description of the feature set:

Dataset Features	Type	Feature Description
Timestamp	Date	Response time
Age	Numerical (Discrete)	Age of the user
Trending_date	Categorical	Date the video trended on YouTube
Purchase_Frequency	Categorical	How often the user buys on Amazon
Purchase_Categories	Categorical	Types of products the user buys
Personalized_Recommendation_Frequency	Categorical	How often purchases are influenced by recommendations
Browsing_Frequency	Categorical	How often the user browses Amazon
Product_Search_Method	Categorical	How the user searches for products
Search_Result_Exploration	Categorical	Whether the user explores multiple pages of results
Customer_Reviews_Importance	Categorical	Importance of reviews in purchase decisions
Add_to_Cart_Browsing	Categorical	Whether the user adds items to the cart while browsing
Cart_Completion_Frequency	Categorical	How often cart items are purchased
Cart_Abandonment_Factors	Categorical	Reasons for leaving items in the cart
Saveforlater_Frequency	Categorical	Frequency of using "Save for Later
Review_Left	Categorical	Whether the user has left a product review
Review_Reliability	Categorical	Trust in product reviews
Review_Helpfulness	Categorical	Usefulness of other customers' reviews

DATA OVERVIEW

Continued..

Dataset Features	Type	Feature Description
Personalized_Recommendation_ Frequency	Categorical	How often user receives personalized recommendations
Recommendation_Helpfulness	Categorical	USefulness of personalized recommendations
Rating_Accuracy	Categorical	Relevance of recommendations
Shopping_Satisfaction	Categorical	Overall satisfaction with Amazon
Service_Appreciation	Categorical	Most Appreciated Amazon Services
Improvement_Areas	Categorical	Suggested areas for Amazon to improve

DATA QUALITY CHALLENGES & ANOMALIES

Few inconsistencies were found in the dataset, which could have affected the analysis if left unaddressed.

DATA ANOMALIES

- **Columns requiring cleaning or type conversion**: timestamp, Gender, Purchase_Categories, Product_Search_Method, Search_Result_Exploration, Customer_Reviews_Importance, Add_to_Cart_Browsing, Cart_Abandonment_Factors, Review_Left, Service_Appreciation, Improvement_Areas, Rating_Accuracy, Shopping_Satisfaction
- **Personalized_Recommendation_Frequency** appears twice with different types (str and int64) resolve duplication (rename or remove).
- **Columns with notable anomalies or invalid entries**:
  - Improvement_Areas & Service_Appreciation (has an entry like '.')
- Two missing values in **Product_Search_Method** column.

DATA CLEANING & TREATMENT

Inconsistent and missing values were addressed, and key features were cleaned and standardized for analysis.

DATA CLEANING SUMMARY

- Features such as t**imestamp, Gender, Purchase_Categories, Product_Search_Method, Search_Result_Exploration, Customer_Reviews_Importance, Add_to_Cart_Browsing, Cart_Abandonment_Factors, Review_Left, Service_Appreciation, Improvement_Areas, Rating_Accuracy, Shopping_Satisfaction** were standardized with respect to data type conversion.
- **Improvement_Areas** & **Service_Appreciation** were cleaned to handle invalid entries (. replaced with mode).
- Personalized_Recommendation_Frequency were resolved and renamed as:
    - '**Personalized_Recommendation_Frequency** ' to Number_of_times_Personalized_Recommendation_Received
    - '**Personalized_Recommendation_Frequency**' to Purchase_made_on_Personalized_Recommendation
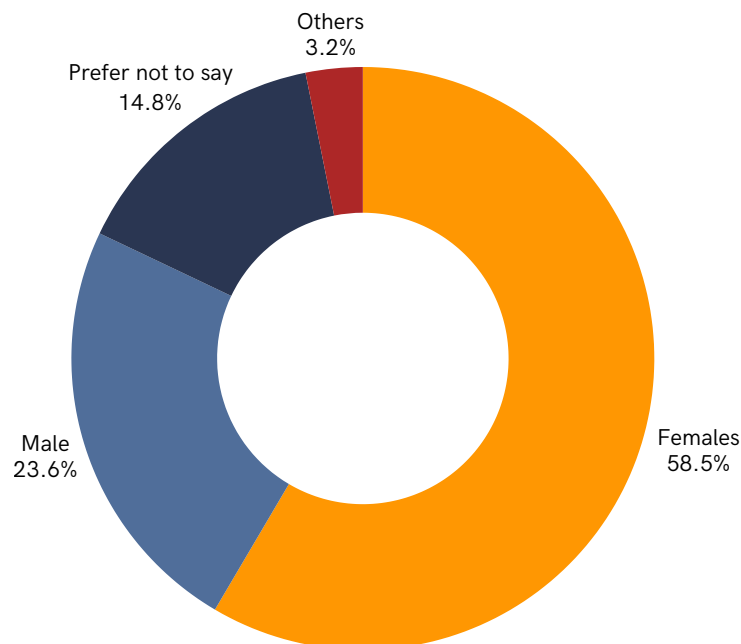- Missing values in **Product_Search_Method** were replaced with mode.

INSIGHTS

Customer Demographics & Engagement Overview

Female users form the majority of Amazon shoppers

58.5%

Of the Customers are Females

Others
3.2%

Prefer not to say
14.8%

Males
23.6%

Females
58.5%

Key observations

- Female users make up the largest share, clearly outnumbering male users.
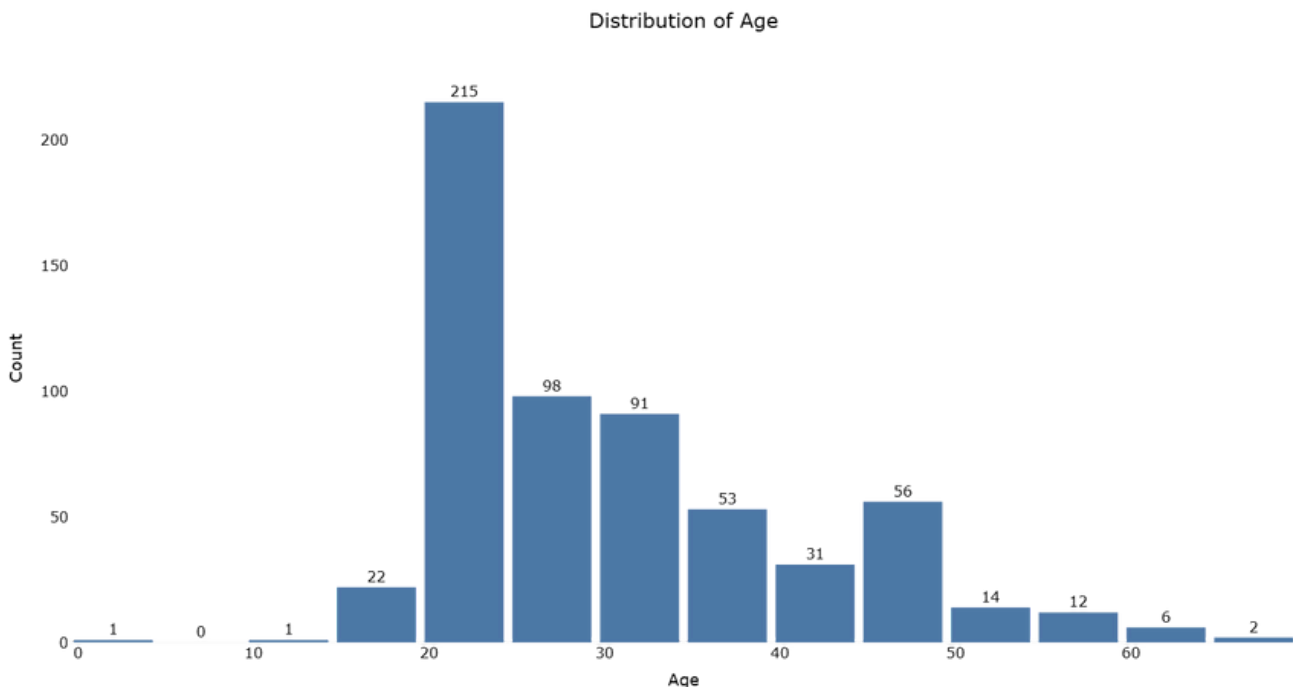- A noticeable group chose not to disclose gender, while very few fall under other categories.

Business Insights

- Observed shopping patterns are more reflective of female user behavior.
- Gender-neutral personalization remains important due to the sizable undisclosed segment.

Amazon users are concentrated in the 20-35 age range

67%

Of the Customers are Young Adults

Distribution of Age

Key observations

- The largest concentration of users falls in their early to mid-20s, forming a clear peak.
- User representation steadily declines beyond the mid-30s, with fewer users in older age groups.
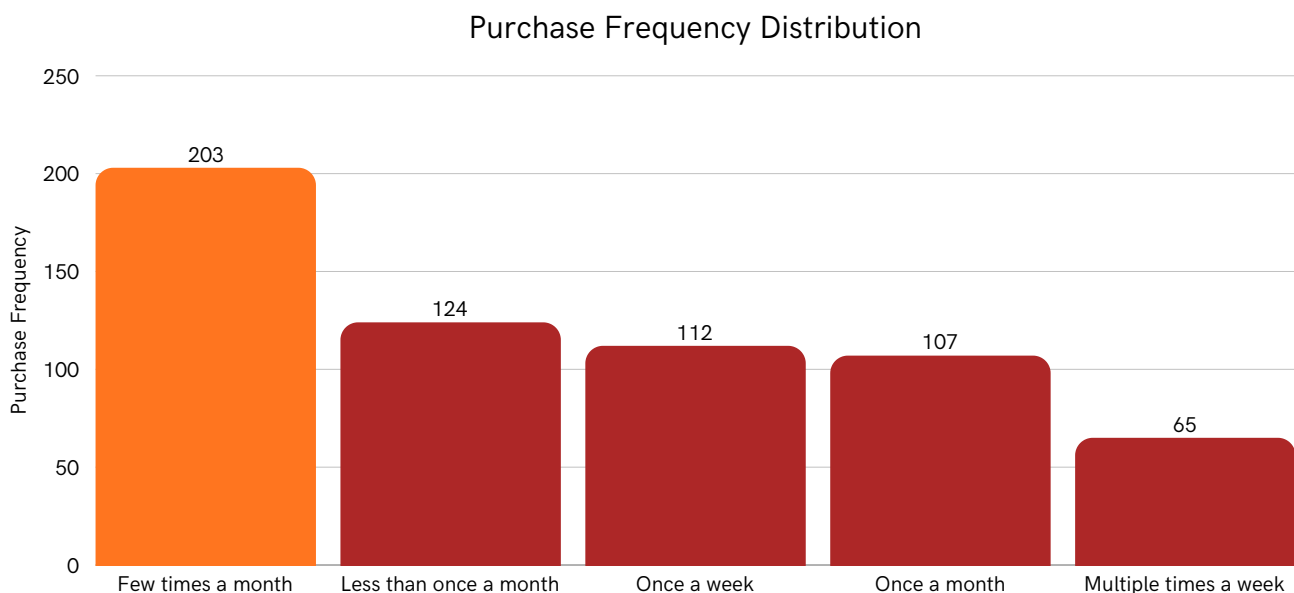
Business Insights

- Products, messaging, and promotions should primarily target young adults to match the dominant user base.
- Simplified navigation and trust signals can help improve engagement among older users.

Most users shop on Amazon occasionally rather than frequently

33.72%

Customers Make Purchases on a Monthly Basis

Purchase Frequency Distribution

Key observations

- The largest share of users purchase a few times a month, followed by those shopping less than once a month.
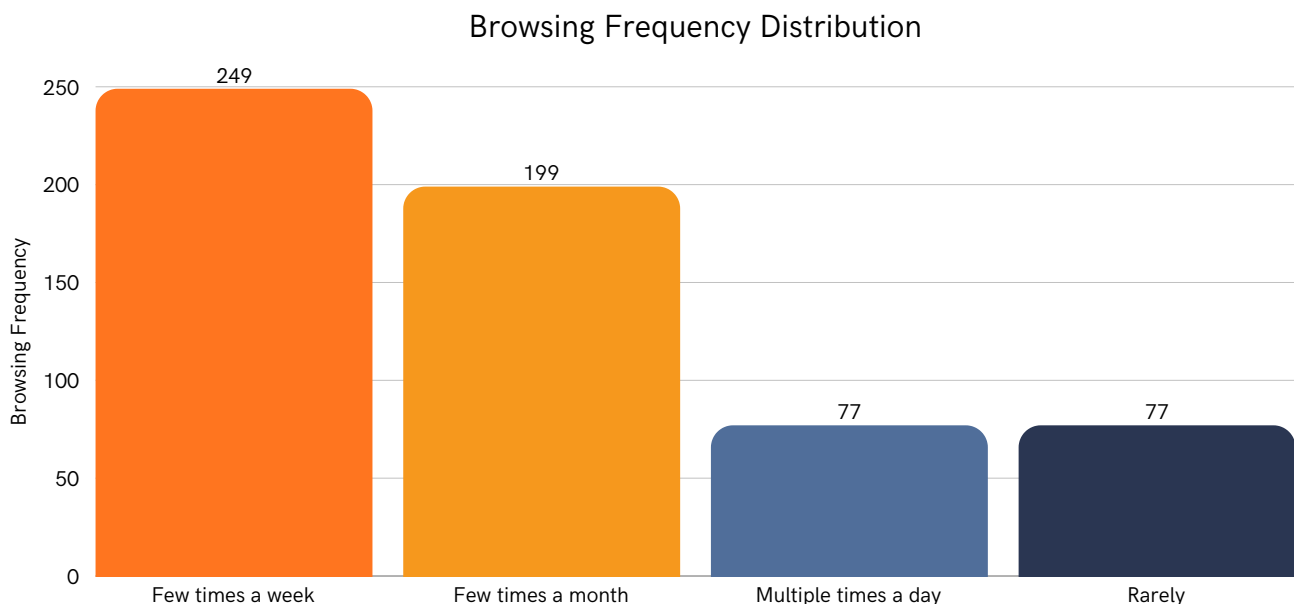- High-frequency shoppers (multiple times a week) form a relatively small segment.

Business Insights

- There is strong potential to increase purchase frequency through reminders, deals, and subscription-based incentives.
- Targeted campaigns for occasional shoppers may yield higher returns than focusing only on already frequent buyers.

Users browse Amazon frequently, even when they don't purchase often

41.36%

Customers Frequently Return to Amazon Each Week

Browsing Frequency Distribution

Key observations

- Most users browse Amazon a few times a week, making it the dominant behavior.
- A sizable group browses a few times a month, while very frequent and very rare browsing are less common.

Business Insights

- High browsing activity presents opportunities to convert browsers into buyers through timely nudges and offers.
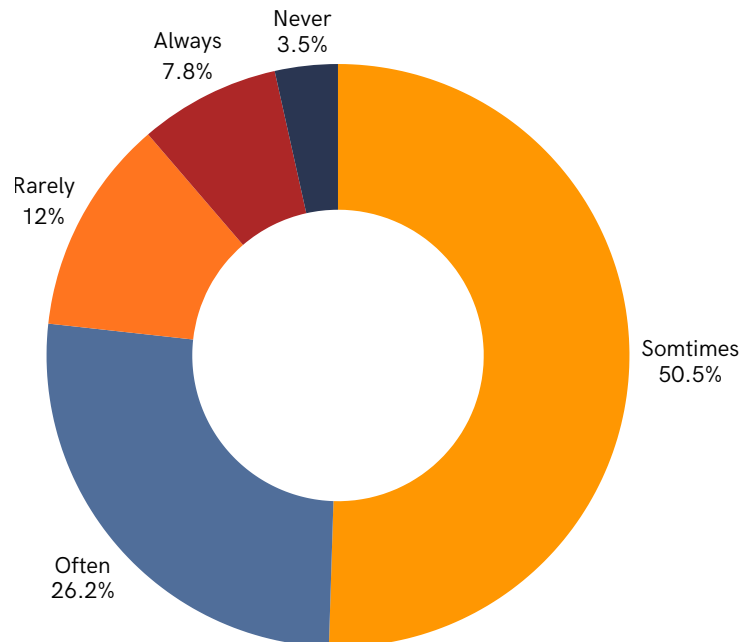- Improving product discovery and recommendations during browsing sessions can increase purchase intent.

Purchase Funnel &
Cart Behavior

Cart completion is inconsistent, with many purchases left undecided

50.5%

Cart Drop-Offs Occur Frequently

Never
3.5%

Always
7.8%

Rarely
12%

Somtimes
50.5%

Often
26.2%

Key observations

- Most users complete purchases only sometimes after adding items to the cart.
- A smaller group completes purchases often, while consistent completion or complete drop-off is less common.
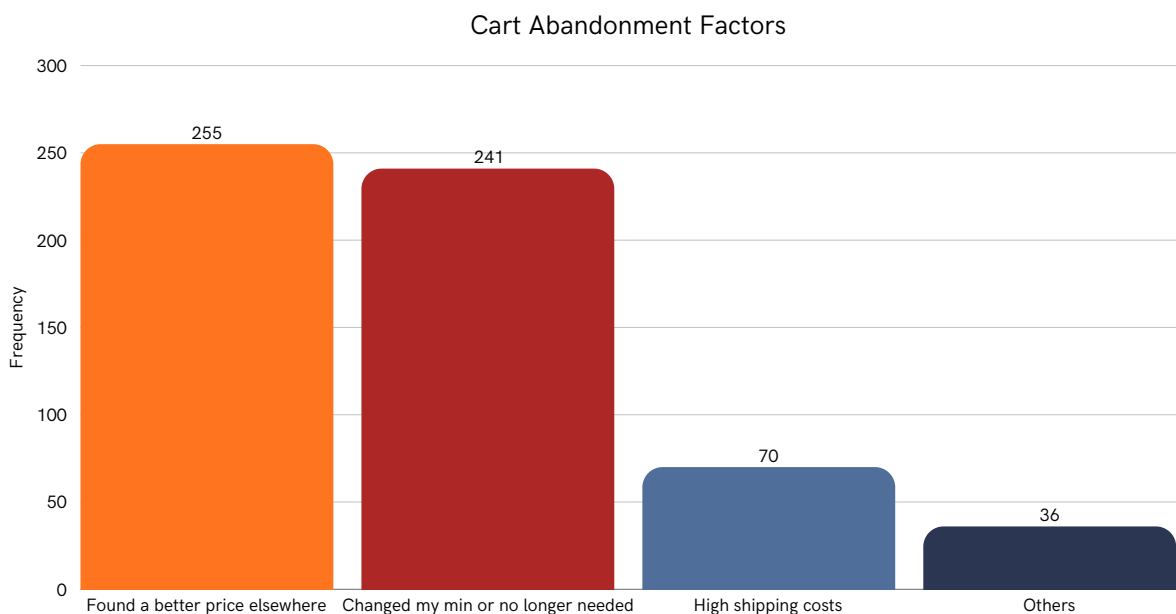
Business Insights

- Reducing friction at checkout can help convert occasional completions into successful purchases.
- Cart reminders, price clarity, and trust signals may improve completion rates among undecided users.

Price sensitivity and indecision are the main drivers of cart abandonment

42.35%

Portion of Consumers Prioritize Price Over Brand Loyalty
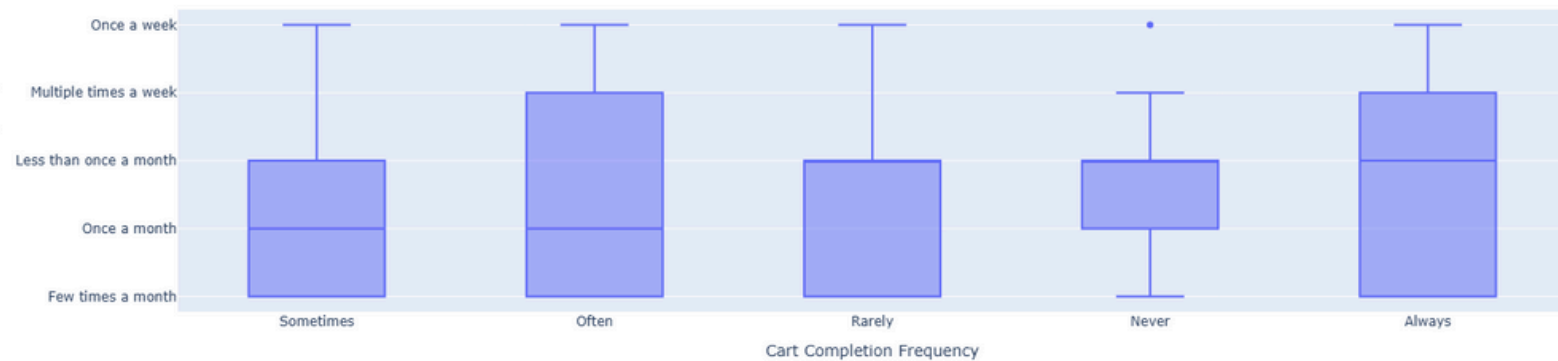
Cart Abandonment Factors

Key observations

- Finding a better price elsewhere is the most common reason for abandoning the cart.
- Many users also abandon purchases after reconsidering or no longer needing the item.

Business Insights

- Competitive pricing and price-match cues can help reduce abandonment due to external comparison.
- Timely nudges, limited-time offers, or reminders may help convert users who are undecided rather than price-driven.

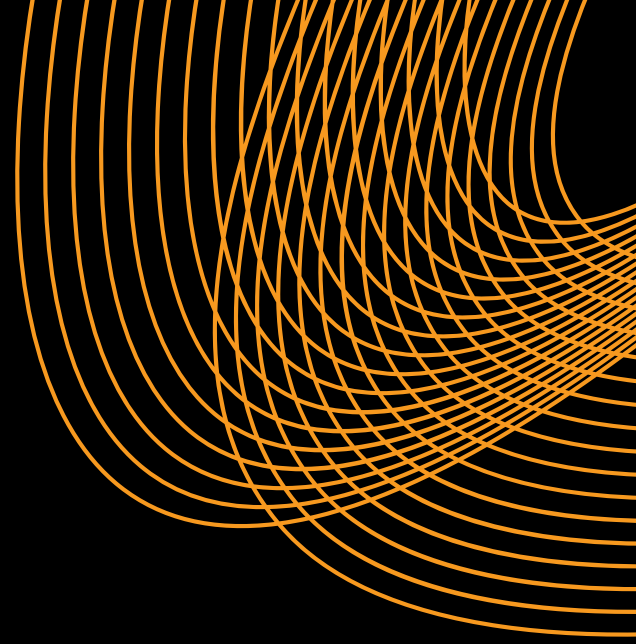# Frequent shoppers are more consistent at completing purchases

Key observations

- Users who purchase multiple times a week or weekly tend to fall under often or always completing their carts.
- Infrequent shoppers are more spread across sometimes and rarely, showing less consistency in checkout behavior.

Business Insights

- High-frequency shoppers can be nurtured with loyalty benefits and faster checkout to sustain completion habits.
- Reducing friction and uncertainty for low-frequency shoppers may help convert browsing into completed purchases.

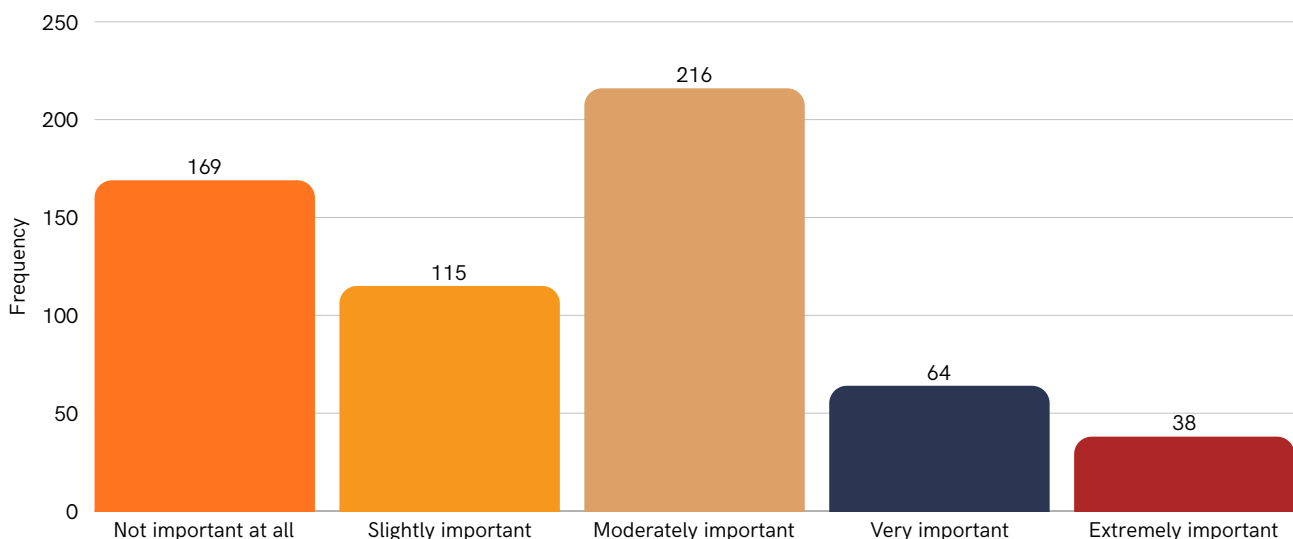# Influence of Customer Reviews on Buying Decisions

Customer reviews play a moderate but meaningful role in purchase decisions

16.94%

customers report that online reviews affect their buying choices

Customer Review Importance Distribution

Key observations

- Most users rate the importance of customer reviews at a mid-level, indicating balanced consideration rather than absolute reliance.
- Fewer users consider reviews extremely important, while a notable group assigns them low importance.

Business Insights

- Reviews should be clear, concise, and informative to influence users who are moderately review-sensitive.
- Highlighting the most helpful and relevant reviews can increase their impact without overwhelming users.

Review participation is evenly split between reviewers and non-reviewers

51.5%

Of the Customers are Review Contributors

Yes
51.5%

No
48.5%

Key observations

- Nearly half of the users leave reviews, while a comparable share choose not to provide feedback.
- Review activity is present but not dominant, indicating mixed engagement after purchase.

Business Insights

- Simple nudges like reminders or small incentives can push passive users to leave reviews.
- Improving the post-purchase experience may increase organic review participation without heavy incentives.

Satisfied shoppers are more likely to leave reviews they find helpful

33.72%

Customers Make Repeat Purchases
Several Times a Month

Review Left vs Review Helpfulness vs Shopping Satisfaction

Key observations

- Higher shopping satisfaction (3-5) clusters with "Yes" reviews marked as helpful or sometimes helpful.
- Low satisfaction users (1-2) are less engaged and more likely to skip reviews or mark them unhelpful.

Business Insights

- Improving overall shopping satisfaction directly boosts both review volume and review quality.
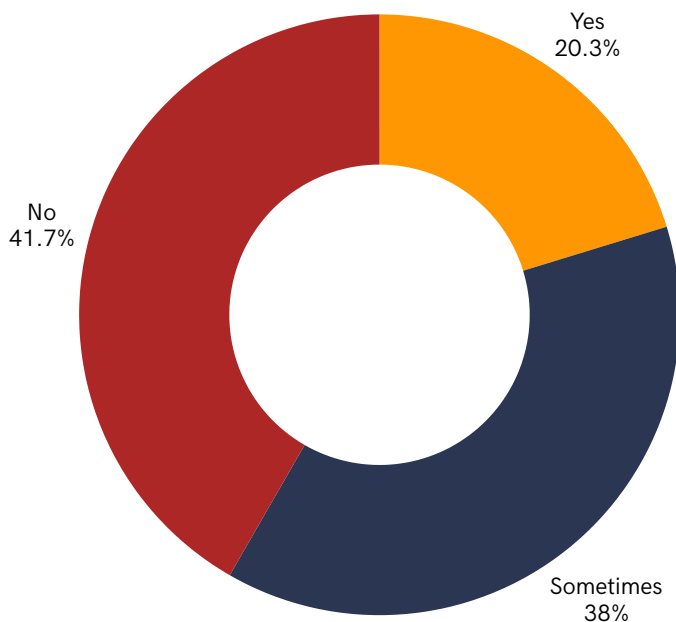- Encouraging happy customers to review can amplify helpful content and influence future buyers.

Impact of Personalized Recommendations

Personalized recommendations influence purchases, but not consistently

20.3%

Customers Are Responsive to Personalized Buying Suggestions
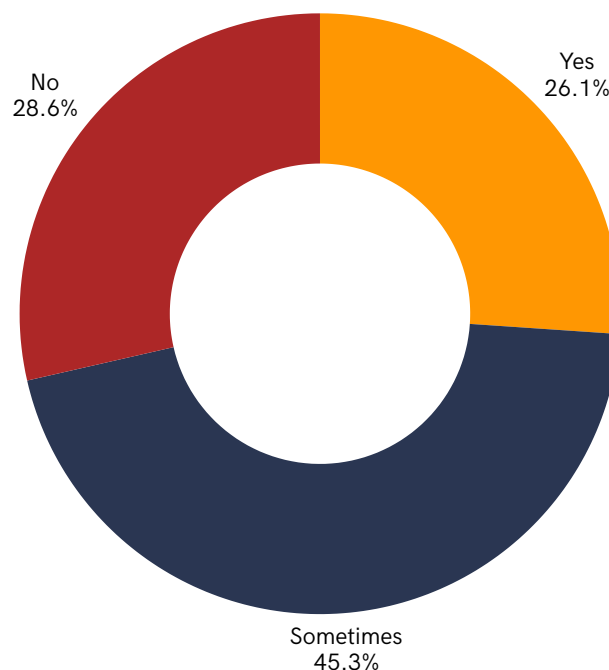

Yes
20.3%

No
41.7%

Sometimes
38%

Key observations

- The largest group of users report not making purchases based on personalized recommendations.
- A substantial portion of users say recommendations influence their purchases sometimes, while fewer users rely on them consistently.

Business Insights

- Improving the relevance and timing of recommendations could convert occasional influence into regular purchases.
- Personalized recommendations should complement, not replace, user-driven search and browsing behavior.

Product recommendations are helpful, but lack consistent reliability

No
28.6%
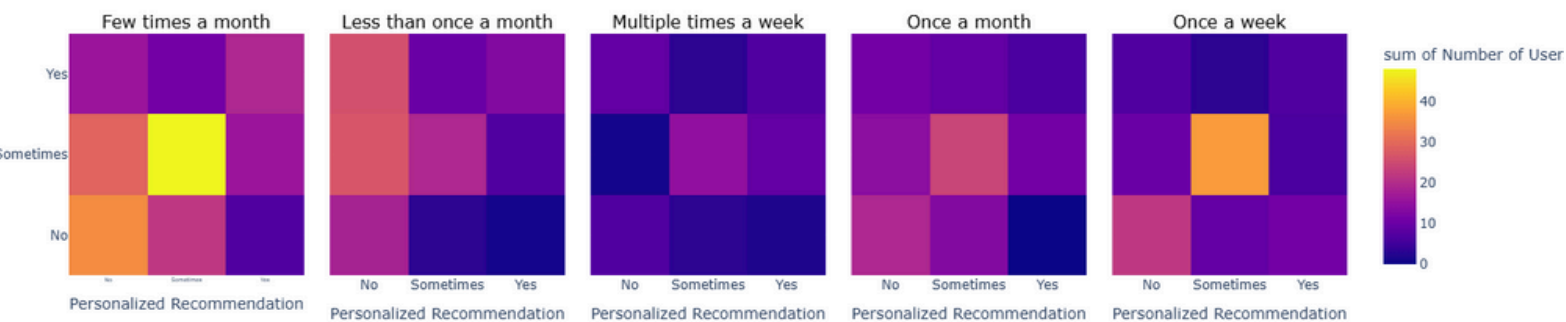
Yes
26.1%

Sometimes
45.3%

Key observations

- Most users report recommendations as sometimes helpful, indicating inconsistent relevance.
- A comparable number of users find recommendations unhelpful versus consistently helpful.

Business Insights

- Improving personalization accuracy can convert occasional usefulness into consistent value.
- Refining recommendation logic can directly increase user trust and engagement.

Personalized recommendations drive purchase frequency only when users perceive them as helpful

Personalized Recommendation vs Recommendation Helpfulness vs Purchase Frequency

Key observations

- Frequent buyers (weekly or multiple times a week) cluster around "Sometimes" to "Yes" helpfulness ratings for personalized recommendations.
- Users who find recommendations not helpful tend to purchase less frequently, regardless of personalization presence.

Business Insights

- Simply offering personalization is not enough—its perceived relevance determines its impact on repeat purchases.
- Improving recommendation accuracy and transparency can convert occasional shoppers into frequent buyers.
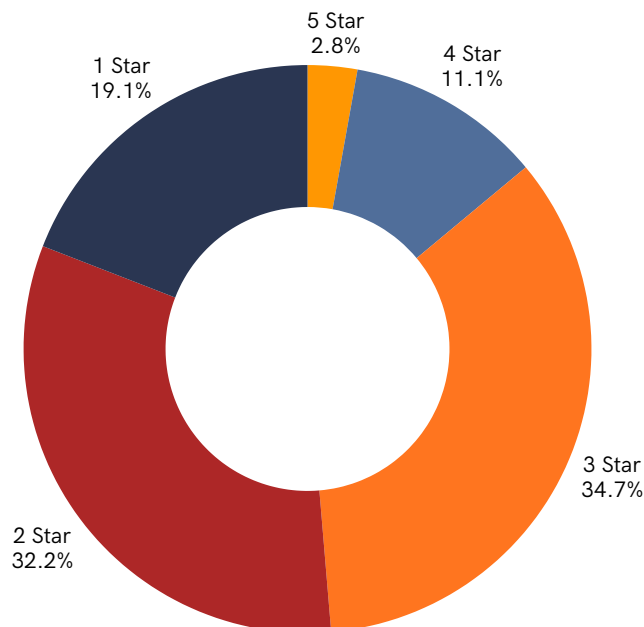
Customer Satisfaction & Experience Evaluation

User satisfaction is largely neutral, with very few highly satisfied shoppers

34.7%

Average Satisfaction Indicates Room for Improvement

5 Star
2.8%

4 Star
11.1%

1 Star
19.1%

3 Star
34.7%

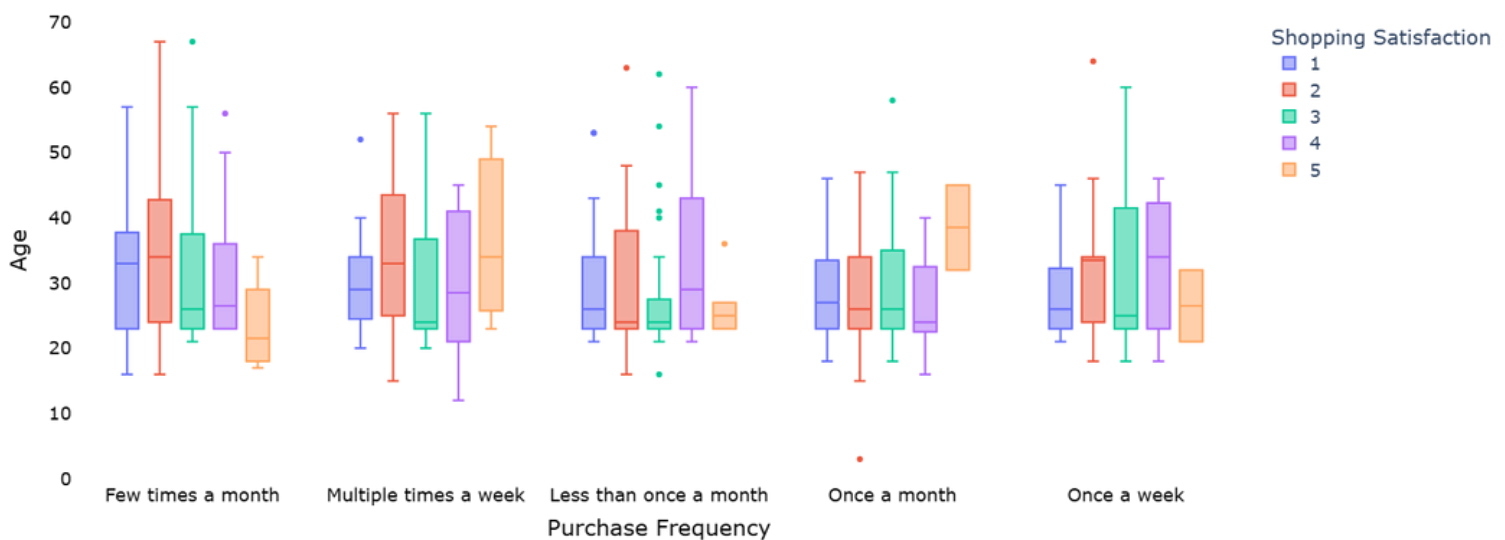2 Star
32.2%

Key observations

- Most users rate their shopping experience at 2-3, indicating average satisfaction rather than strong approval.
- Extreme satisfaction (rating 5) is rare, suggesting limited delight or "wow" experiences.

Business Insights

- There is a clear opportunity to move neutral users toward higher satisfaction through better UX, pricing, or post-purchase support.
- Small improvements in the end-to-end experience could yield disproportionate gains in loyalty and repeat purchases.

Shopping satisfaction varies more by purchase frequency than by age

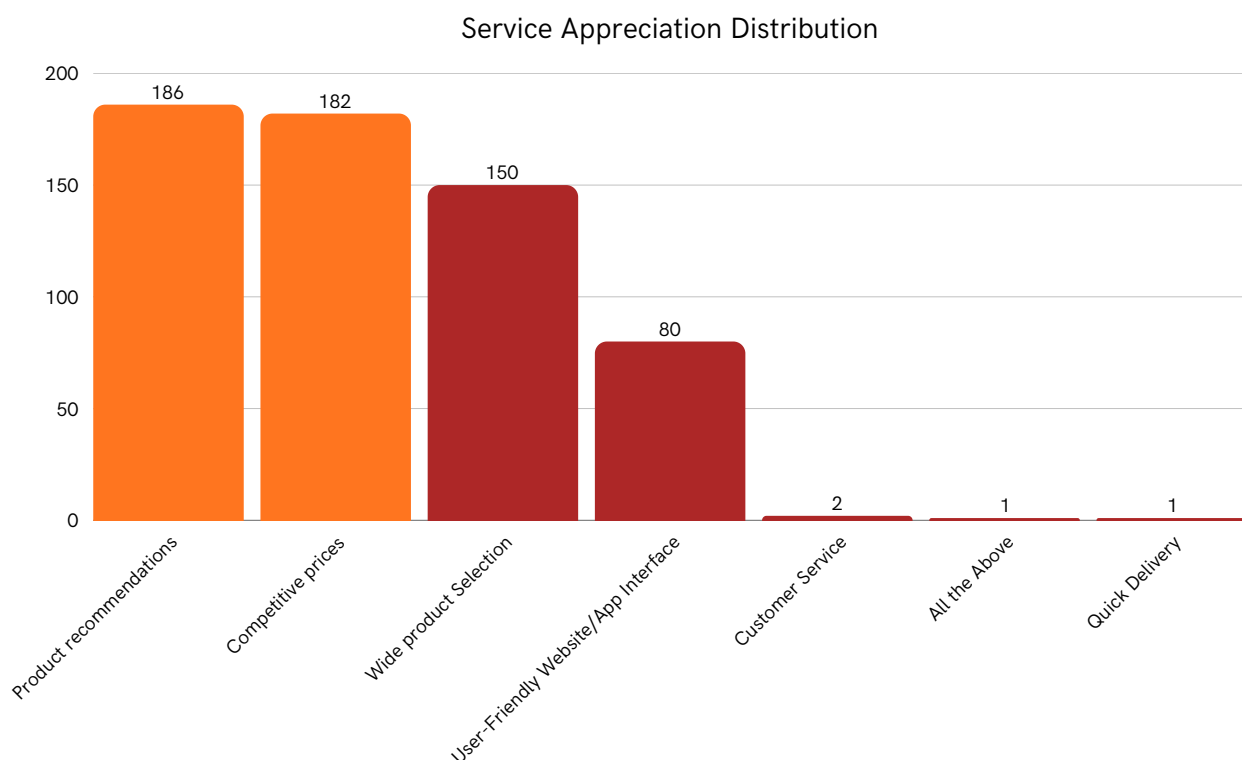Shopping Satisfaction by Purchase Frequency and Age

Key observations

- Across age groups, satisfaction levels cluster similarly within each purchase frequency segment.
- More frequent shoppers tend to show slightly higher and more consistent satisfaction compared to infrequent buyers.

Business Insights

- Improving satisfaction drivers for infrequent shoppers could increase repeat purchase behavior.
- Loyalty and engagement efforts should prioritize purchase frequency over age-based segmentation, as age shows limited differentiation.

DATA ANALYTICS REPORT

Product recommendations and pricing drive the strongest service appreciation

Service Appreciation Distribution

Key observations

- Product recommendations and competitive prices receive the highest appreciation among users.
- Interface usability and product variety matter, but trail behind value-driven factors.
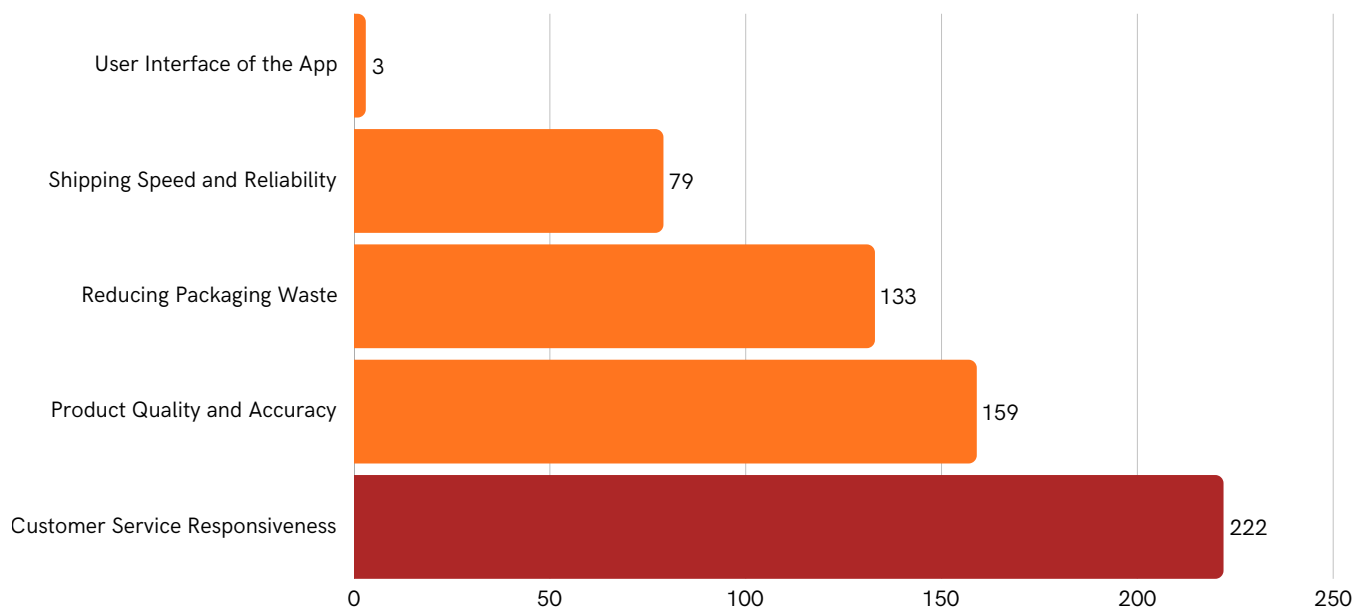
Business Insights

- Strengthening recommendation quality can directly enhance perceived service value.
- Pricing competitiveness remains a critical lever for user satisfaction and retention.

Customer service and product reliability are the most critical improvement drivers

36.87%

Customers Desire Better Support Experience

Improvement Areas Distribution

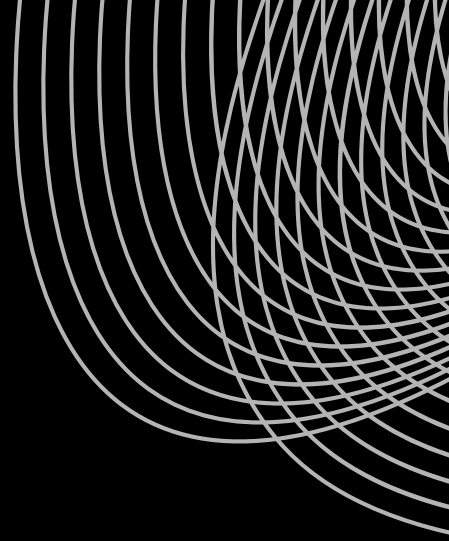| Improvement Area	Value
User Interface of the App	3
Shipping Speed and Reliability	79
Reducing Packaging Waste	133
Product Quality and Accuracy	159
Customer Service Responsiveness	222

Key observations

- Customer service responsiveness and product quality receive the highest number of improvement mentions.
- Shipping reliability and packaging waste follow, while UI-related issues are relatively minimal.

Business Insights

- Enhancing customer support efficiency can significantly improve overall satisfaction.
- mproving product accuracy and delivery reliability is key to building long-term trust.

BUSINESS / DEVELOPER TAKEAWAYS

- **Female-Led Usage with Inclusive Design Needs**: Shopping behaviour is more aligned with female users, but a sizable undisclosed segment makes gender-neutral personalization essential for scale.

- **Young Adults Drive Core Demand**: Young adults dominate the user base, making them the primary target for products, messaging & promotions, while simplified UX and trust signals can help retain older users.

- **Occasional Shoppers Represent the Biggest Growth Opportunity:** Targeted reminders, deals, and subscription incentives can convert occasional shoppers more effectively than focusing only on already frequent buyers.

- **High Browsing, Low Conversion Signals Untapped Potential:** Strong browsing activity highlights opportunities to convert interest into purchases through improved discovery, timely nudges, and smarter recommendations.

- **Checkout Friction Is a Key Drop-Off Point:** Reducing uncertainty at checkout through price clarity, trust signals, and cart reminders can significantly improve completion rates.

BUSINESS / DEVELOPER TAKEAWAYS

- **Price Sensitivity Drives Abandonment:** Competitive pricing, price-match cues, and limited-time offers are effective in converting undecided users who compare across platforms.

- **Loyalty Should Focus on Purchase Frequency, Not Age:** High-frequency shoppers respond well to loyalty benefits and faster checkout, while low-frequency shoppers need friction reduction and reassurance to complete purchases.

- **Reviews Influence Decisions, when curated well:** Clear, concise, and highlighted "most helpful" reviews influence moderately review-sensitive users without overwhelming them.

- **Post-Purchase Experience Fuels Organic Reviews:** Improving satisfaction after purchase encourages higher-quality reviews and greater participation without relying heavily on incentives.

- **Recommendation Quality Determines Real Impact:** Personalization only works when relevance is high. Improving accuracy and timing can turn occasional influence into consistent purchasing behavior.

- **Satisfaction Is the Strongest Loyalty Multiplier** Small improvements in UX, pricing, customer support and delivery reliability can yield outsized gains in repeat purchases and long-term trust.

CHALLENGES & OPPORTUNITIES

Limitations

- The dataset is survey-based and may not fully reflect actual or current Amazon consumer behavior trends.
- Respondents represent a limited subset of Amazon users, potentially introducing bias toward more active or engaged shoppers.
- Timestamped responses provide only a snapshot in time, limiting long-term trend analysis.

Future Work / Opportunities:

- Perform sentiment analysis on customer reviews to understand consumer perception and satisfaction.
- Conduct time-series analysis on purchase and browsing patterns to study how consumer engagement evolves over time.
- Compare frequent buyers with infrequent buyers to identify factors driving purchase behavior.
- Analyze the impact of product categories, personalized recommendations and review metrics on shopping decisions.

CONCLUSION

This analysis of Amazon consumer behavior provided a comprehensive understanding of purchasing patterns, browsing habits, and engagement with platform features. Key insights and trends were identified, demonstrating how data-driven analysis can inform strategic decisions for sellers, marketers and platform managers. Future analysis could include sentiment analysis of customer reviews, time-series tracking of purchase and browsing patterns and comparisons between frequent and infrequent buyers to uncover deeper insights.

This analysis was conducted using Python, Pandas, Matplotlib, Seaborn and Plotly.

Author:

Tejas Jadhav

GitHub: @tejas-jadhav
LinkedIn: Tejas Jadhav