

Age and Gender Differences in Online Consumer Behavior:

Insights from an Amazon Customer Survey

An exploratory analysis based on survey data

Student: AZOUZOU MANAL

Course: Data analysis and Softwares

Instructor: PETERLE EMMANUEL

Institution: Marie and Louis Pasteur University

Academic Year: 2025–2026

Introduction

Online shopping has become part of everyday life. Over the past years, the development of the internet and digital technologies has changed the way consumers search for products, compare prices, and make purchasing decisions. E-commerce platforms now offer many features that simplify the shopping experience, such as easy access to product information, customer reviews, fast delivery options, and flexible return policies.

Among these platforms, Amazon plays a major role in online retail. It offers a very large variety of products and allows users to browse and purchase items with only a few clicks. Amazon also relies on personalized recommendations and data-driven tools to improve the user experience and encourage purchases. While these mechanisms are designed by the platform, consumers do not all react in the same way, and their behavior remains diverse.

I chose this topic because online shopping is something I personally use on a regular basis, and Amazon is a platform that most people are familiar with. I found it interesting to better understand how consumers behave in an online environment and whether factors such as age, gender, satisfaction or reliance on reviews influence purchasing decisions. This topic also seemed relevant beyond personal interest, as understanding consumer behavior is important for businesses that operate in digital markets.

The aim of this study is to analyze consumer behavior on Amazon using survey data. The analysis focuses on purchasing frequency, product category preferences, browsing behavior, customer reviews, and satisfaction with the platform. By examining how these elements vary across individuals, this study seeks to provide insights into the behavioral factors that influence online purchasing decisions.

Methodology

Data source

The data used in this study comes from a publicly available dataset on Kaggle. The dataset is based on a survey created using Google Forms and focuses on consumer behavior on Amazon. The survey was conducted mainly in the Indian subcontinent, as well as among Indian residents living in European and Middle Eastern regions. Responses were collected by sharing the survey link on various social media platforms.

The dataset contains individual level responses related to demographic characteristics, purchasing habits, browsing behavior, customer reviews, satisfaction, and perceived areas for improvement. As the data is self-reported, it reflects consumers' perceptions and declared behaviors rather than observed purchasing actions.

Research objectives

The main objective of this study is to understand how consumer behavior on Amazon varies across individuals, with a particular focus on the role of demographic characteristics and online shopping habits. More specifically, this study seeks to answer the following main research question:

How do age and gender shape purchase frequency, review behavior, and satisfaction in online shopping on Amazon?

To address this main question, several sub-questions are explored:

- What is the demographic composition of Amazon customers in the dataset in terms of age and gender?
- How do purchasing frequency and product category preferences vary by age and gender?
- Is there a relationship between browsing frequency and purchasing frequency across gender (female and male)?
- Do age and gender influence review behavior, including leaving reviews and reliance on customer reviews?
- Does customer satisfaction and service appreciation affect purchasing frequency?

To answer these questions, the analysis relies primarily on descriptive statistics and group comparisons, complemented by a simple regression model to examine the relationship between purchasing frequency, customer satisfaction, and service appreciation.

Data preparation and challenges

Several challenges arise from the nature of the data. First, most variables in the dataset are qualitative and ordinal, such as purchasing frequency, browsing frequency, satisfaction levels, and reliance on reviews. In order to conduct meaningful analysis, these variables were carefully encoded, either as ordered factors or numeric variables when appropriate.

Second, some variables allow multiple responses. For example, the variable related to purchase categories contains multiple selections for a single respondent. This required transforming the variable into separate binary indicators in order to analyze product category preferences.

Additionally, minor data cleaning was necessary, including handling placeholder values, and removing non-informative fields such as timestamps and other variables. These transformations were performed directly in R to ensure transparency and consistency.

Overall, given the nature of the survey, the dataset limits causal interpretation, but it remains well suited for descriptive and exploratory analysis of online consumer behavior.

Results

1. What is the demographic composition of Amazon customers in the dataset in terms of age and gender?

The demographic composition of the sample shows a clear imbalance across gender groups. Overall, the respondents are mostly female, with 352 female respondents compared to 142 male respondents. The remaining participants identify as “Others” (19 respondents) or “Prefer not to say” (88 respondents).

After creating age groups, the distribution of respondents by age and gender reveals similar patterns across most age categories. As shown in Figure 1 (grouped bar chart), females represent most respondents in all age groups. The largest group in the sample is females aged 18–24, with 164 respondents, followed by females aged 25–34 (91 respondents). Male respondents are also most represented in the 25–34 and 18–24 age groups, although their numbers remain consistently lower than those of females. Older age groups (45–54 and 55+) contain fewer respondents overall, regardless of gender.

This demographic composition should be kept in mind when interpreting subsequent results, as the sample is more representative of younger and female Amazon users.

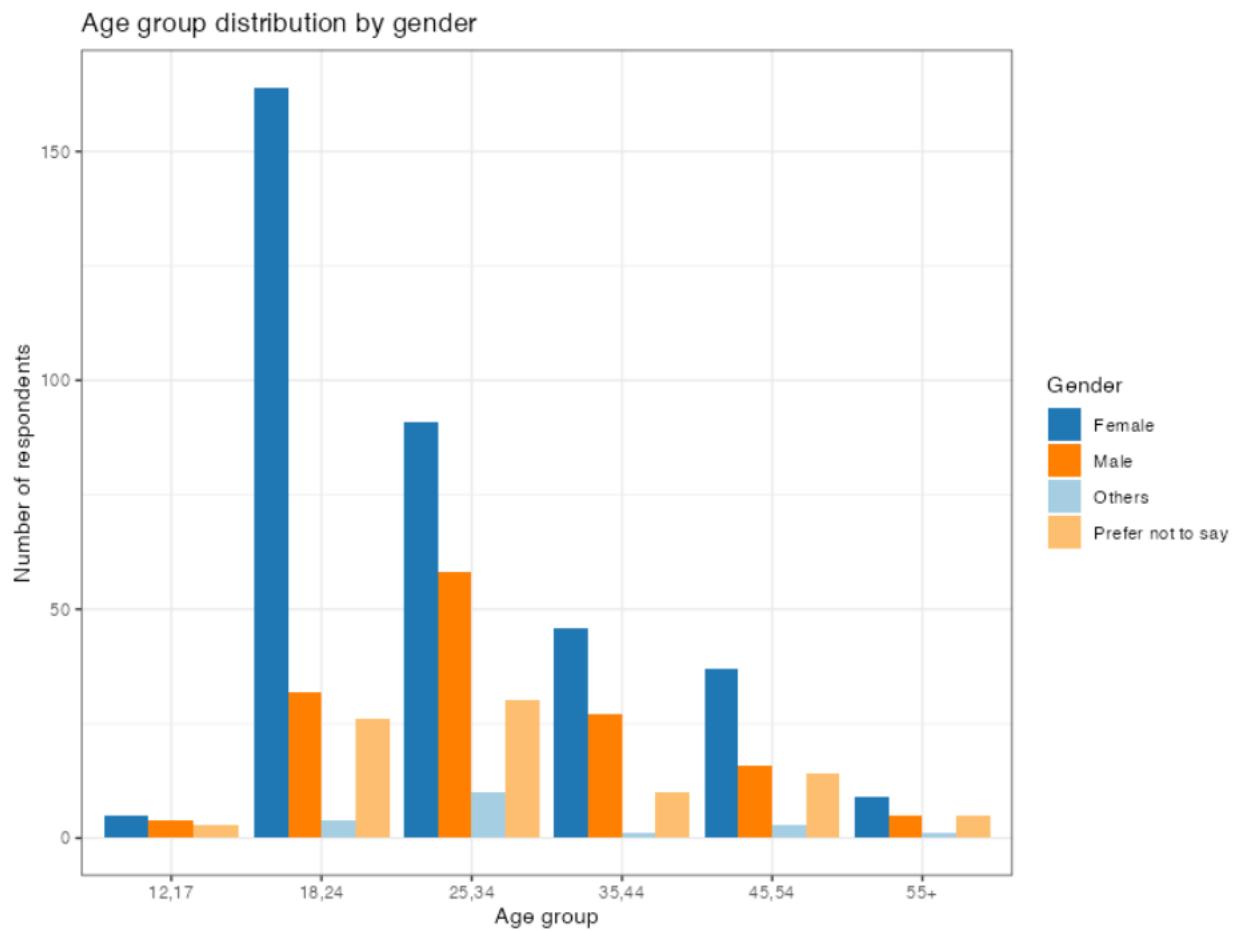


Figure 1: Distribution of respondents by age group and gender

2. How do purchasing frequency and product category preferences vary by age and gender?
 - How does purchasing frequency vary by age and gender?

Figure 2 represents a heatmap illustrating the distribution of purchasing frequency across age groups. Across all age categories, purchasing “a few times a month” appears to be the most common behavior, as indicated by the consistently higher shares in this category.

While some variation across age groups is observed, differences remain relatively moderate. Weekly and very frequent purchasing occurs across all ages but represent a smaller share of respondents, particularly when compared to monthly purchasing patterns.

Overall, purchasing frequency appears broadly similar across age groups, with no age specific concentration in very frequent purchases.

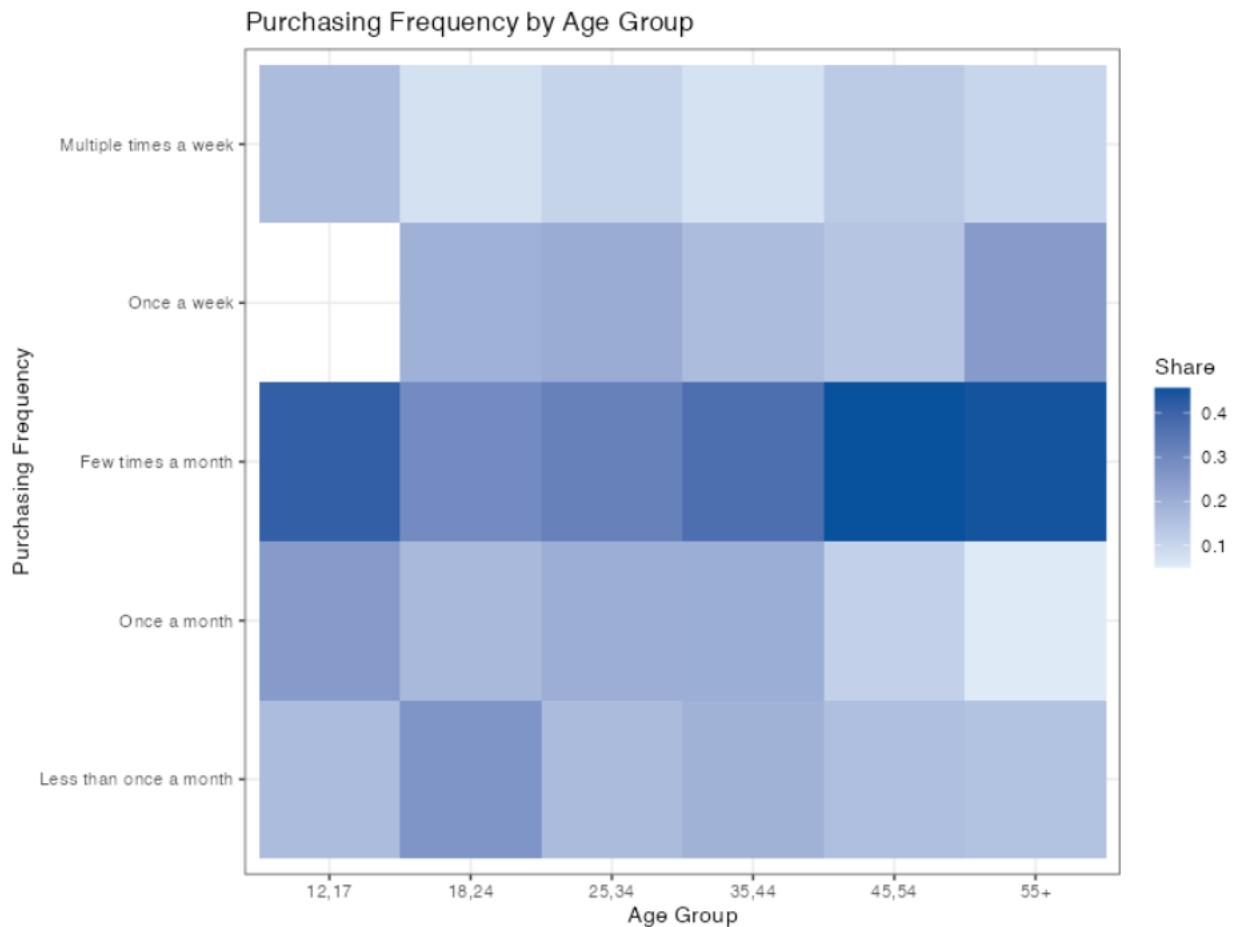


Figure 2: Purchasing frequency by age groups

Consistent with the descriptive results mentioned above, female respondents generally report slightly more frequent purchasing behavior than male respondents.

- How do product category preferences vary by gender?

Table 1 shows that product category preferences differ across gender groups. Female respondents are more likely to buy beauty and personal care products, while male respondents more frequently purchase from home and kitchen as well as grocery related categories. Clothing and fashion purchases are common across all genders. Patterns observed for respondents identifying as “others” and “prefer not to say” should be interpreted with caution due to smaller sample sizes.

Gender	Beauty category	Clothing category	Home category	Groceries category	Other category
Female	59.09	56.82	34.66	16.19	21.59
Male	38.03	55.63	43.66	23.24	45.07
Others	52.63	47.37	31.58	36.84	31.58
Prefer not to say	52.27	62.50	32.95	15.91	22.73

Table 1: Product category purchase rates by gender (%)

3. Is there a relationship between browsing frequency and purchasing frequency for female and male respondents?

A moderate positive correlation is observed between browsing frequency and purchasing frequency among female respondents ($r=0.406$). Among male respondents, the correlation is positive but weaker ($r=0.334$).

4. Do age and gender influence review behavior, including leaving reviews and reliance on customer reviews?

The results indicate differences in review behavior across age groups and gender. Older respondents tend to leave reviews more frequently than younger respondents. In particular, the share of respondents who report leaving reviews increases among individuals aged 45-54 and 45+, reaching its highest level in the oldest age group. At the same time, reliance on customer reviews decreases with age, as older respondents report lower average levels of review reliance compared to younger age groups.

Younger respondents, especially those aged 18-34, show higher reliance on customer reviews, while a smaller proportion reports leaving reviews themselves. This suggests that younger consumers tend to rely on reviews more as a source of information and decision, whereas older consumers are more likely to contribute to reviewing products.

Gender differences are also observed. Female respondents are slightly more likely to leave reviews than male respondents, with 52.6% reporting having left a review compared to 47.2% of males. Differences in average reliance on reviews between females and males are relatively small.

5. Does customer satisfaction and service appreciation affect purchasing frequency?

A linear regression was conducted to examine whether customer satisfaction and service appreciation influence purchasing frequency. At the 5% significance level, none of the explanatory variables were statistically significant. This suggests that, within this sample, purchasing frequency is not strongly explained by customer satisfaction nor service appreciation.

Conclusion

This study examined how age and gender influence purchasing behavior, review engagement, and satisfaction among Amazon users, using survey data collected from users across different demographic groups. By combining descriptive analysis, this study aimed to provide an explanatory understanding of online consumer behavior.

The results indicate that purchasing frequency varies across age groups and genders, with younger respondents and female customers generally reporting more frequent purchases. Product category preferences also differ by gender, particularly for beauty and personal care and home related products. A moderate positive relationship was found between browsing frequency and purchasing frequency, especially among female respondents, suggesting that increased engagement with the platform is associated with more frequent purchases. In contrast, review behavior shows noticeable patterns across age groups, with older respondents more likely to leave reviews, while younger consumers tend to rely more heavily on reviews when making purchasing decisions. Finally, the regression analysis did not reveal a statistically significant relationship between purchasing frequency and either customer satisfaction or service appreciation at the 5% of significance level.

Overall, these findings suggest that demographic characteristics, in particular age and gender, play a role in shaping online shopping behavior on Amazon. However, the survey qualitative nature of the data limits causal interpretation. Future research could rely on larger and more diverse samples or include additional behavioral indicators to better capture the dynamics of online consumer decision making.

Data sources and references

Kaggle: [Amazon Customer behavior Survey](#) [[Dataset](#)]

Wickham, H., et al. (2019). *Welcome to the tidyverse*. Journal of Open Source Software, 4(43), 1686.

R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.