

Instructions to Find our Assignment Components

Our ReadMe contains our written analysis. The final code can be found in the folder “Project 1 G5” and is titled “Amazon_Code_Final.” Our individual codes can be found in the branches corresponding to our names. All charts and graphs were put into the “Output” folder within the repo. Our slide deck is also within the repo folder titled “Northwestern Group 5 presentation.”

Written Analysis

Introduction

Our database examines the behaviors of people on Amazon that are part of or influence their decisions on the platform. We proposed our questions based on what we believe the average seller may be thinking when they list items for purchase on Amazon. Our data allows us to examine who and how people purchase goods on one of the largest e-commerce platforms of the modern age.

Age Group vs Purchasing Frequency

For this statistical analysis we looked at age to see how different groups compared in their purchasing behaviors on Amazon. We created a bar chart to see a comparison between age groups and a boxplot to see the variance in behaviors among individual age groups. The purchasing frequencies were provided in the data set and the age groups used were determined based on the US Census’ standardization of age groups.

Majority of people in this sample size purchase items on Amazon once a month. For a seller on the platform, this could imply that customers are taking a more considered approach to their purchases, possibly engaging in planned shopping rather than frequent, spontaneous buys. In response, sellers should consider optimizing their product listings to highlight features that resonate with customers who make less frequent purchases, such as product durability, long-term value, or unique selling points.

The age group that purchases the most frequently on Amazon, according to this data frame are individuals between the ages of 18-24. While the individuals within this age group weren’t primarily purchasing the most frequently on Amazon, when looking at who does the most purchasing it was them. This is a signal to a vendor on the platform to cater and advertise to this age group since they are the most likely to purchase items often.

Finally, based on the box plot analysis, age groups 18-24 and 35-44 had the most variance as indicated by the size of their respective boxes. This suggests distinct

shopping patterns within these demographics that go beyond just age. As we know not all individuals within an age group think the same. As a seller on Amazon, it's crucial to recognize the need for tailored marketing and product strategies within each demographic beyond age. For the 18-24 age range, characterized by higher variability, a seller should explore trend-centric marketing approaches. Similarly, for the 44-65 age group, which also displays significant variability, their strategy should focus on stability and reliability, highlighting product durability, quality, and long-term value.

Cart Abandonment

We decided to analyze the responses the participants gave for cart abandonment, which is when people put items in their cart but do not end up purchasing them. The three main questions we wanted to find answers to were:

Why are people abandoning carts?

Does gender play a role in cart abandonment?

Does search type play a role in cart abandonment?

We wanted to answer these three questions since they are the most relevant for sellers who want to increase the percentage of orders that are completed once buyers add the item to their carts. Knowing the reason why people chose to abandon their carts or discovering that a search type led to a higher percentage of participants saying they abandoned their carts gives the opportunity for sellers to adjust their listings accordingly.

The biggest reasons people stated for abandoning their cart were changing their mind and finding a better price elsewhere. A small percentage of people stated that it was because shipping costs were too high.

The people surveyed were asked to rate how often they abandoned carts, with the answers being "Often," "Sometimes," "Rarely," or "Never." The participants' answers to this question were studied in relation to gender. It was found that a higher percentage of men abandon carts "Often" or "Never." However, the % of women that answered that they abandon carts "sometimes" was nearly 10% higher than the percentage of men.

Search type was also investigated in relation to cart abandonment. People responded to the search type with "Filter," "Keyword," and "Category," and the percentage of people that gave each abandonment response was found for each of these search types. It was found that the percentage of people who searched by categories that stated that they abandon carts "Sometimes" was approximately 10% more than the percentages for the other search types. In addition, the percentage of people who search by category and abandon carts "Often" is significantly lower. There

were no significant differences in abandonment answers for people who searched by “Filter,” or “Keyword.”

The main conclusion that can be drawn from the cart abandonment data is that people mainly abandon carts due to changing their mind or finding a better price elsewhere, not due to shipping costs. This would suggest that sellers should not be concerned with the shipping price listed on their item, but should instead focus on the item price. Although there were some differences in responses by gender or search type, a lack of numerical data from this survey makes drawing conclusions about these topics difficult (for example, there is a 10% difference between participants that selected “Sometimes” and “Never” for gender, but what percentage difference in abandonment does this actually correlate with). Thus, more research should be conducted on these questions with numerical data in order to best help sellers.

Who is most likely to leave an amazon review?

We decided to ask this question because reviews are often what make people want to buy something or not. Reviews are so important that companies have been known to sometimes buy fake reviews in order to make their product more marketable. By looking at who is likely to leave a review, companies can reach these customers through advertising or by offering incentives to leave a review.

We first looked at gender to describe Amazon shoppers. The data included male, female and others. It was first assumed that women would leave more reviews because there are more women than men in our dataset. We plotted the average percent of reviews left by each gender group and found that there is actually little variation. Each group leaves reviews at a very similar rate. We can say that of the people surveyed in our dataset, women leave more reviews than men, but the difference is so small that it is not entirely accurate to say that as a description of amazon shoppers.

Next, we looked at how often people left reviews within different age groups. The data in the bar chart includes 6 different age groups. Interestingly, there are 3 age ranges between 18 and 44, but each group left reviews at nearly the same rate. So, someone in their early 20s is just as likely as someone in their early 40s to leave a review based on our dataset. The percent of reviews left by the two oldest age ranges noticeably increases with people over 65 leaving more reviews than any other age group.

Lastly, we looked at shopping satisfaction and service appreciation. Shopping satisfaction can describe a customer’s overall shopping experience and service appreciation describes specific features that a customer particularly value such as competitive prices and quick delivery. Based on the chart, it appears that the more satisfied a customer is, the less likely they are to leave a review. Customers tend to

leave more reviews when they have a neutral experience or worse. Customers with a neutral experience who value product recommendations left the most reviews.

Based on these findings, Amazon customers of all genders tend to leave reviews at similar rates. Amazon customers between the age of 18 and 44 leave reviews at similar rates while customers over 65 leave more reviews than any other group. Satisfied customers leave less reviews than neutral or unsatisfied customers.

Browsing vs. Buying: How do consumers compare when it comes to browsing through Amazon compared to purchasing?

A statistical analysis was performed to create two bar charts representing the different behaviors of customers when it comes to browsing products versus buying. For browsing, there were four options customers could select to describe their behavior: rarely, a few times a month, a few times a week, and a few times a day.

These responses were used to create a bar chart representing how the customers varied in their browsing behaviors. According to the data, most customers browse a few times a week or a few times a month. Few customers fall on the extreme ends of browsing: either rarely or multiple times a day.

For purchasing, there were five options customers could select to describe their behavior: rarely, once a month, a few times a month, once a week, and a few times a week. These responses were used to create a bar chart representing how the customers varied in their purchasing behaviors. Purchasing behaviors were a little more evenly distributed, with the most frequent purchasing behavior being a few times a month. Similar to browsing, most people do not purchase multiple times a week. However, in contrast to browsing, many people do purchase rarely and/or once a month from Amazon.

These same results were then shown as pie charts displayed with percentages. These pie charts provide more exact insights about consumer behavior. For browsing, over 40% of customers browse a few times a week. This is clearly the dominant customer behavior. When combining browsing a few times a week with browsing a few times a day, nearly 65% of customers browse weekly to some extent.

For buying, the dominant behavior is to purchase a few times a month with 33.7% of customers reporting this behavior. Interestingly, the next most common behavior is to purchase once a week. A similar percentage of consumers purchase once a week as they do once a month.

Finally, a chi squared test was performed to compare browsing frequency and purchasing frequency. The p-value for this test was extremely low. With an alpha of 0.05, 0.005, or 0.005, the test would show that there is a statistically significant

relationship between browsing and purchasing. This relationship is demonstrated in the gradient map at the end of the code.

The dominant behaviors show that in general, customers browse more frequently than they purchase. This makes sense, because most people are limited financially. This may restrict customers' wish lists of "nice to have's" to their "must have's." Additionally, people may browse several times when searching for a single product to find the best reviewed product at the best price. A final reason that this may be the case is because one consumer may buy multiple products in a single purchase. This is a limitation of the data set, because it is unknown whether purchase frequency corresponds to each item or each order. Ultimately, as a seller, it may be important to target the most frequent browsers so that when they go to make a purchase, your product is top of mind and more likely to be added to their cart.

Final Conclusions

We were able to establish several key conclusions from this data set. First, the majority of the age groups purchase on Amazon once a month. Second, one of the main reasons that people put stuff in their cart and fail to purchase it is because they find a better price for the item somewhere else rather than due to shipping cost. Next, Amazon reviewers are a diverse group but older groups tend to leave slightly more reviews. Lastly, Amazon customers browse the storefront much more frequently than they buy, and frequent browsing was a dominant characteristic of the survey respondents.

Discussion and Limitations

There are a few considerations that must be taken into account when examining the analyses of this data. The data used is qualitative, subjective data that was voluntary. The age of the sample was skewed young and female, with most of the respondents being women aged 25-34. Additionally, there is no way to measure the truthfulness of the data, since it is a voluntary, electronically administered survey with no identifiers required. The sample size is also quite small with just over 600 respondents. These 600 respondents may not be representative of the population of millions of Amazon consumers in the United States. These are limitations that can be addressed in future research.

Future Research

This data and the analyses we were able to perform provide a useful starting point to understand which types of data are required to answer questions important to Amazon stakeholders and executives. To answer correlation questions about the relationships discussed in this analysis, objective data that uses quantitative measures linked to customer IDs would be essential. To strengthen the validity of the data set, it is recommended to collect data over time and from a heterogeneous sample that is representative of the Amazon consumer population in the United States. It may be

especially important to account for age and gender. Finally, a larger sample size is recommended so that any observed trends may be more confidently presented.