The Report of Exploring Amazon Customer Behavior in June 2023

GROUP4

Yuxin GONG & Jiani XU

1. Project Introduction

(1) Project Dataset Description

Data Source: We chose the dataset called Amazon Customer Behavior Survey from the Kaggle and it is mainly about the behavior of Amazon Customer.

Data Preprocessing: We performed data preprocessing using the pandas library in Python.

In the column of Timestamp, we converse date to datetime.

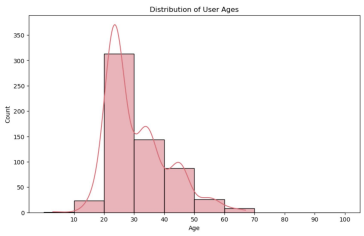
In the column of Purchase Categories, there are lots of categories in each cell, so we create new rows for each split purchase category and copy the original rows into the new rows.

(2) Project Background

Nowadays, online shopping is one of the primary propelling forces in the business sector, a position that it has secured owing to the evolution of the digital society. It should be noted that more consumers are opting to access goods and services through online channels due to the increased usage and ease of use of the Internet which results from advanced technology. This requires businesses to undertake a detailed study on consumer behavior and preferences to better comprehend what customers need.

2. Exploratory Data Analysis (EDA)

(1) Age distribution

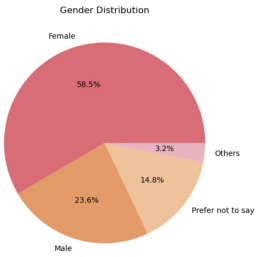


The bar chart illustrates the distribution of Amazon Customers' age. In 2023, the majority of users are between 20 to 30, with approximately 310. It is notable that fewer users using Amazon as the age get older. The ages of 40-50 roughly half the size of younger users (around 150). Although there is some representation from teenagers and the elders, it is relatively minimal, approximately 20 and 30 users respectively.

Key findings: The most important information to me of using this visualization is that the majority of Amazon's customer are youth and middle-age while the highest number falls around 20 to 30 years.

Compare to reality: Therefore, the comparison to reality is that the higher number of youth almost is realistic because modern youth is already more comfortable with online shopping, and Amazon has more suitable products and services for their usage and taste-consuming preferences, but the lower number of clients after the 40th year can also be realistic since people at the older age prefer more usual shopping places for the purchasing of goods.

(2) Gender distribution

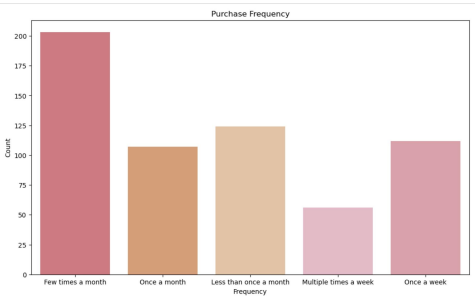


The pie chart shows the gender distribution among Amazon Customers. It is clear that women are more likely to utilize Amazon compared to men. Specifically, female users account for 58.5% of the total, while male users make up only 23.6%. Additionally, 18% of the users' gender remains unspecified.

Key findings: The pie chart depicting the gender distribution among Amazon customers reveals a significant gender imbalance. This suggests that Amazon's customer base is predominantly comprised of female users, highlighting the platform's appeal and accessibility to this demographic.

Compare to reality: This may well be due to the relatively high popularity of Amazon among female users and, on the contrary, the relatively low among males. Indeed, according to some studies and data, Amazon does have a high proportion of female users among its customers. This is most likely explained by the fact that while shopping is more of a female activity, they are more likely to look for various goods and services on the Internet and have more frequent and established consumption habits in online stores.

(3) Purchase frequency



The bar chart illustrates that how frequent the Amazon Customer shopping. The purchase frequency of most consumers is distributed between "less than once a week frequency", "once a week" and "once a month". A smaller number of consumers' purchasing frequency is distributed between "multiple times a week".

Key findings: Since some customers buy more is often revealed than others do, and most purchase infrequently, the data show variations for different intervals.

Compare to reality: This may actually be the customers' actual shopping habits on Amazon. For example, based on Amazon's historical data and market research, the majority of customers shop within the range of "less than once a week" to "once a month." This is because most people live life within these ranges, and shopping every day is too frequent.

(4) The relations between age, rating accuracy and shopping satisfaction

This Scatter plot shows that the users from 15 to 50 prefer to make a rating after shopping. And when their shopping satisfaction is not that good, which is less than 3, they would like to give a low rating (also less than 3). If they satisfy with their shopping, giving a satisfaction score greater than 3, they will also give a high

rating, generally above 3.

Key findings: Therefore, shopping satisfaction is highly related to users' ratings. Hence, shopping satisfaction is a very influence factor of the accuracy of a user's rating. As illustrated in the scatter plot, user's satisfaction affects the reviews of customers and their overall ratings.

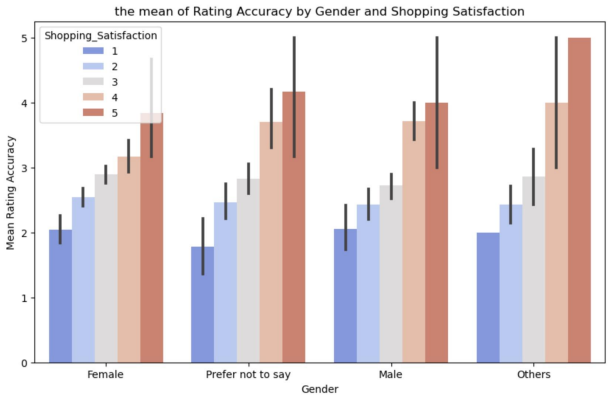
Compare to reality: In reality, many consumers evaluate their experience only after the shopping is finished. While such evaluation can assist other users in making better choices while doing the shopping or buying, and give merchants ideas on where and how they can be improved or optimized, usually such evaluation is directly connected to satisfaction of the shopping process. If the consumers are satisfied, they usually give high ratings. Therefore, this trend is reflected and reasonable to find in this analysis.

(5) Gender

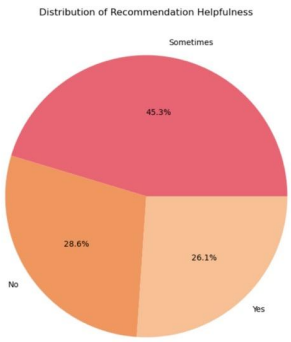
From this picture, we can see that each gender is presented from the highest satisfaction rating to the highest rating accuracy to the lowest satisfaction rating to the lowest rating accuracy. Among them, others' scoring accuracy is the highest among all gender groups when the satisfaction level is 5. Female is the lowest. Prefer not to say, when the satisfaction level is 1, the scoring accuracy is the lowest among all gender groups and Female is the highest.

Key finding: This provides good suggestions for the priority and reference of feedback opinions, which in turn of reflects the differences in experiences and expectations of different groups during the shopping process.

Compared to reality: In real situation, consumers of different genders and degrees of satisfaction may perceive and respond to the product or service experience differently. For instance, men or women might care more about the quality and reliability of certain products and hence produce a more accurate rating when they are satisfied. Conversely, less satisfied consumers may provide a more negative evaluation, resulting in ratings that are less accurate.



(6) The distribution of Recommendation Helpfulness

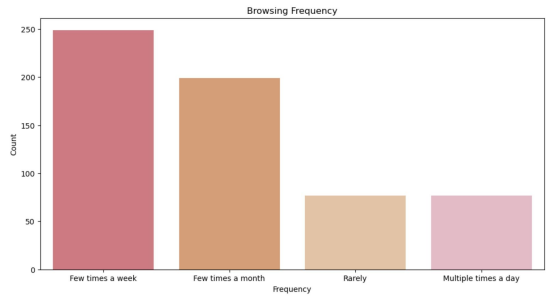


The pie chart shows that what customers' opinion of Amazon's Personalized Recommendation feature. Nearly half customers (45.3%) thinks that the Personalized Recommendation sometimes useful for them during their shopping experience. While those who answered yes or no accounted for half of the remaining respectively.

Key findings: the Amazon's Personalized Recommendation is not good enough to satisfy customers and improve their shopping experience.

Compared to reality: In fact, such a personal recommendation function has been commonly adopted in Amazon and other online shopping platforms and has always been regarded as a key means to enhance the user's shopping experience and promote sales. Personalized recommendation is helpful to many consumers, and through it, buyers can more easily find the commodities that meet their preferences and interest.

(7) The browsing frequency



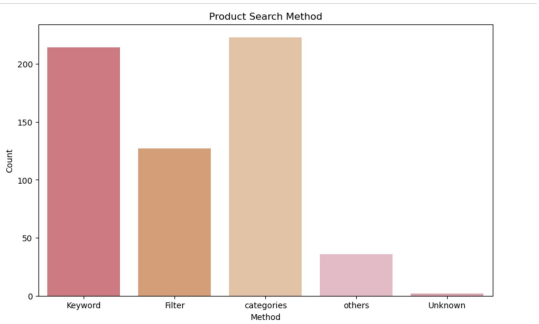
The bar chart illustrates the frequency of customer browsing on Amazon. It is clear that most people would like to browse Amazon a few times a week, with approximately 250 customers, which is 50 more than those who browse the Amazon few times a month. Additionally, the number of customers who prefer to browse the Amazon multiple times a day and rarely brows the Amazon are the same which are around 75.

Key findings: This points out that a large share of the users frequently interact with Amazon compared with others. But this includes the different types of users who maybe interacting with Amazon at some lower frequency than others. The users may be shopping on Amazon less than these users because of their different shopping behavior.

Compare to reality: In real world, in order to better evaluate the frequency of Amazon user browsing, it is necessary to get some other data sources, such as the activity log of the users, the analysis of the traffic structure of the website, and other data, so that we can really understand the user's browsing behavior, and make a more accurate judgment.

(8) The method of product search by customer

This bar chart illustrates that customer prefer to use keyword and categories method to search the product that they want, with the number more than 200 respectively. In addition, 125 customers like to use filter to find their demands. What's more, there are also a small number of people who use other methods to find what they want to buy.

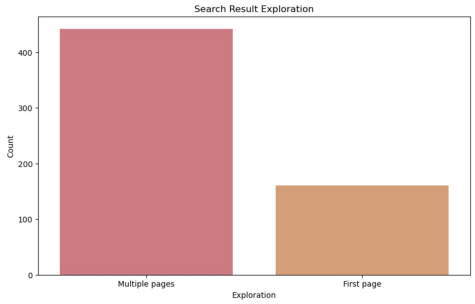


Overall, it clear that customers predominantly rely on traditional search methods such as keywords and categories when browsing for products. However, the inclusion of filter usage suggests a willingness to refine search results further. This highlights the importance of providing intuitive search functionalities that cater to diverse user preferences, ultimately enhancing the overall shopping experience.

Key findings: While most customers base their browsing on traditional search results, which is demonstrated by their reliance on both keywords and categories, the implementation of filter usage indicates that they are willing to narrow down the search results. As a result, search should be sufficiently intuitive for ensuring that users can manipulate the search results in multiple ways, promoting a better overall shopping experience.

Compare to reality: In fact, many customers may be inclined to search for the right products using keyword searches and category filters. Such an approach is typically more versatile and inclusive as customers can search by specific keywords according to their demands and desires or scroll through different product categories to investigate different items. Moreover, the search by filter function is also one of the typical for Amazon patterns. This option allows the users to narrow their search results in a more precise manner to locate the products most closely matching their needs more quickly.

(9) The search result exploration



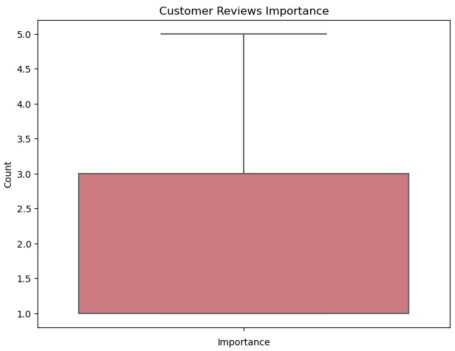
This bar chart illustrates that large number people tend to search through multiple pages when they browse search results, with more than 400 customers doing so. In contrast, only 150 customers is able to find their ideal product through the first page of search results.

Overall, the data suggests that customers generally engage in extensive browsing when searching for products on the platform, indicating a thorough exploration of available options.

Key findings: This underscores the importance of optimizing search algorithms and providing relevant recommendations to enhance the user experience and facilitate efficient product discovery.

Compare to reality: In reality, most users will instead come to Amazon at the moment which is right for them to do some shopping. That might be once a week for casual browsers, or two or three times a week for those who browse for items frequently, look up prices and deals, and so on. The individual frequency will vary wildly depending on of the shopper, their needs, and their habits, but it will also always fall within a certain range.

(10) The importance of Customer Reviews



The boxplot illustrates that the maximum is 5 and minimum is 1. And the Q1 is 3, Q3 is 1 and median is also 3.

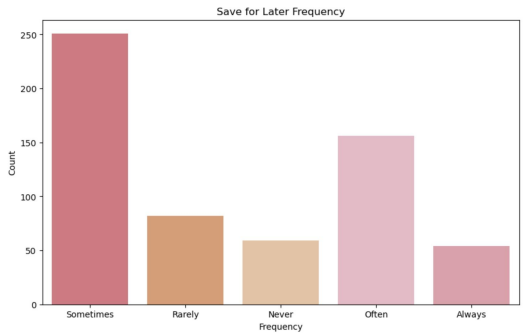
Overall, the boxplot shows the distribution of the importance of customer reviews, with most customers rating it at a moderate to low level.

Key findings: Customers do not fully trust the existing review system.

Compare to reality: Customers could have various opinions and assessments about how important reviews matter. It's clear that customer reviews are at the heart of selling products on an e-commerce platform. Reviews can convey valuable information to other potential buyers, enabling them to make informed buying decisions.

Therefore, customer reviews might be very important to many customers, especially the ones who are more inclined to rely on other people's experiences and knowledge. There are also customers who think reviews are not that important at all. They can judge the quality and value of the product themselves and rely less on the experiences and knowledge of other people. These customers believe the importance of customer reviews is overstated. However, some customers might simply question the authenticity and credibility of reviews and deny the importance of reviews.

(11) The save for later frequency



The bar chart shows that 250 customers will sometimes save for later if they find something they like, which is 150 more than the number of customers often save for later. By contract, the number of customers who rarely save items for later, as well as those who never or always do so, remains relatively consistent.

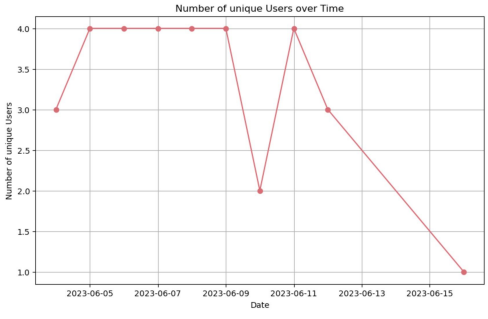
Key findings: This suggests varying preferences among customers regarding the practice of saving items for future consideration. And it also shows that it is reallly

essential to set the function of “save for later”.

Compare to reality: In fact, to determine how customers are actually using Save to Buy Later, it might be necessary to use more data sources and dig deeper in your research to validate whether the results reflect your customers’ real behavior and needs. But we think it depends on personal behaviour and whether the customer want to buy immediately.

(12) The number of unique users over time

*unique user: the number of unique users represents the number of unique users during a specific time period, each user is counted only once, even if they were active multiple times during that period



The line graph illustrates the number of unique users in June 2023. Initially, the number stands at 3 on May 6, 2023. Then it increases to 4, and it remain for 5 days. However, on the following day, it drops to 2 before rebounding to 4 on June 11, 2023. Following this peak, the number gradually declines.

Key findings: This consistency can only mean that the actual engagement of the website with users is changing on a day-by-day basis . The line seems to wobble up and down according to the number of unique users, but the baseline is relatively fixed, and only every now and again do you get a big jolt in either direction. There are so many contributing factors to these "jolts", from marketing and promotional efforts to the seasonal nature of the adult industry, as well as stuff that happens in the outside world that we can't possibly predict or control.

Compare to reality: In fact, the number of users of an Internet platform is usually affected by a variety of factors, including but not limited to marketing activities, product promotions, seasonal demand, user experience and so on.

3. Conclusion

These analyses above show various behaviors and preferences of Amazon customers. We can find that most users falling within the 20 to 30 age range, the number of females is more than males,purchase frequency mostly distributed below once a week, a close correlation between shopping satisfaction and rating accuracy, personalized recommendations is not very useful for most users, more users like to use keyword and category to search, most users browse multiple pages when browsing search results, customers generally believe that customer reviews of their shopping are important, and there are certain differences in the behavior of saving products for future purchases and the number of users changes over time and is affected by a variety of factors. The results provide valuable insights into user behavior and provide direction for improving products and services.

The limitations of this analysis maybe include potential biases in the data collection process, such as sampling bias or self-reporting bias, which may affect the generalizability of the findings.

In the future, firstly, we expect a continued rise in mobile shopping, driven by the increasing prevalence of smartphones and enhanced mobile shopping experiences. Additionally, personalized and AI-driven recommendations will become more sophisticated, offering tailored product suggestions based on individual preferences and browsing history. Furthermore, we predict a growing emphasis on sustainability and ethical consumption, with consumers prioritizing eco-friendly products and socially responsible brands.

In summary, the future of online retail will be shaped by mobile-led, personalized recommendations, continuity and immersive technologies, requiring businesses to be adaptable to thrive in this dynamic environment.