**Sodowoodo Project 1**

*Outline and analyses*

Lisa Serrano

Ruby Ghabboun

Patricia Young

Aida Roman

**The unveiling of Amazon Consumer Behavior**

**Introduction:**

In this project we will analyze Amazon consumer behavior data, using more than two weeks of data we will look at a wide range of variables such as gender, age, user interactions from browsing all the way to cart completion, reviews, and purchase time patterns. Using this behavior data our goal is to gain a deeper understanding of the consumer decision making, habits and preferences to optimize marketing campaigns and the overall experience on the site.

**Transition:**

To be able to gain the best understanding we depicted 4 major questions we will be presenting.

First we will be analyzing what affect Age and Gender have on the data with Ruby.

**Question 1 Ruby:**

The average age is approximately 30.79. The age data has a standard deviation of around 10.19 indicates moderate variability (spread) from the mean.

The youngest age in the dataset is 3 years. It's unusual for a 3-year-old to make purchases directly on Amazon. It might be a data entry error, a misunderstanding, or a representation of family accounts used by individuals of different ages.It might be necessary to investigate further to understand how and why an entry with a 3-year-old's age is present in the dataset.

25% of the data has ages less than or equal to 23 years.

50% of the data has ages less than or equal to 26 years.

75% of the data has ages less than or equal to 36 years.

The oldest age in the dataset is 67 years.

The age distribution chart shows that the majority of Amazon customers are in the 21-30 age group, The next largest age group is 31-40 & Customers in the 51-80 and 11-20 age groups make up a smaller portion of the customer base.

This data suggests that Amazon is popular with young adults. This may be due to a number of factors, such as the convenience of online shopping and accessibility of technology.

Amazon should focus its marketing efforts on young adults, as they represent the majority of its customer base and may want to develop more products and services that are specifically targeted to young adults.

The box plot visualizes the distribution of ages across different genders. The plot provides a clear representation of central tendencies, variability, and potential outliers.

Females:

The median age is lower in females than in males Median Age: 28 years Interquartile Range (IQR): 20 to 36 years

Males:

Median Age: 30 years Interquartile Range (IQR): 22 to 37 years

Females:

Age Range: 3 to 67 years Whiskers span 1.5 times the IQR

Males:

Age Range: 5 to 65 years Whiskers span 1.5 times the IQR Outliers:

Females exhibit slightly more variability with a few outliers beyond the whiskers. Males tend to have a more compact spread with fewer outliers. Conclusion: The box plot indicates that the median ages between genders are quite similar, but females demonstrate a slightly broader range in age distribution and a few notable outliers compared to males.

This analysis provides insights into age distributions, central tendencies, and variability among different genders, offering useful insights into the age demographics within the dataset.

and the interquartile range (IQR) is wider in males than in females, suggesting that there is more variability in age among male customers.

The chart shows that the most common purchasing frequency for all age ranges is "Few times a month." However, there are some differences in purchasing frequency across age ranges.

Younger customers (ages 19-30) are split in habbits since the 2 most common purchasing frequency for this age range is either "few times a month" or "less than once a month", this suggests a divergence in purchasing behavior.

These differences might be due to varying income levels, personal priorities

Older customers (ages 41-50 & 51-80) even though they are a minority in this data set at only 20% combine the majority of them are more likely to purchase items "Few times a month". And thats most likely due to their finantial stability and probably family situation .

The proportions of purchase categories by gender are following:

Beauty and Personal Care:

Female: 65.41% Male: 16.98% Others: 3.14% Prefer not to say: 14.47%

Clothing and Fashion:

Female: 58.31% Male: 23.03% Others: 2.62% Prefer not to say: 16.04%

Groceries and Gourmet Food:

Female: 51.35% Male: 29.73% Others: 6.31% Prefer not to say: 12.61%

Home and Kitchen:

Female: 55.45% Male: 28.18% Others: 2.73% Prefer not to say: 13.64%

Others:

Female: 45.78% Male: 38.55% Others: 3.61% Prefer not to say: 12.05%

The data provides a clear insight into the proportions of purchase categories across different genders. It's apparent that the preferences vary significantly among different groups. For instance, females tend to dominate purchases in categories like "Beauty and Personal Care" and "Clothing and Fashion" compared to males, whereas for the "Groceries and Gourmet Food" and "home and kitchen" category, the proportions are more evenly distributed between genders. The "Others" and "Prefer not to say" categories show a fair balance among different catigories. These observations highlight the diversity and distinct preferences across various product categories based on gender.

***Transition:*** *After looking in depth at age and gender lets now take a look if days of the week or time play a role in consumer behaviors with Lisa.*

**Question 2 Lisa:**

*What impact does day/ time have on consumer purchases?*

Slide 1 Analysis:

Chart 1- .**Understanding when consumers make purchases is crucial to help optimize sales and ensure ad spend strategy is being distributed at the most effective times.**

For this analysis we grabbed timestamps and separate completed purchases by days of the week.

In this graph We can observe for males the most purchases were done on Thursday, Tuesday and Wednesday (weekdays), as opposed to a huge dip on Friday and saturday.

For females the most purchases were done on Sunday and Wednesday, specially on wednesday.

For the weekend we saw a dip on Saturday for both gender therefore ads placed on this day would not be the most effective.

In another case I would love to explore more data that will show deals on the weekdays versus the weekends to see if this factor is having an effect on the data.

Chart 2- The next chart is an extension analysis of the first graph now showing us just purchases not divided into genders to give an outlook on what day the purchases were mostly made.

To no surprise we can see without bias that Thursday and wednesday are the most popular days to make purchases for Amazon.

Slide 2 Analysis:

Chart 1- In this chart we are going to look at purchases by time of day to better understand behavior and insights. I grouped the days in the 3 groups. Evening, Afternoon, and Morning.

The most sales were completed in the evening, this can maybe be the result of lighting deals and messages that are sent out at night which is another data set I would love to explore to make sure strategy plans are the most accurate and ensure more consumer purchases.

However without this data we can still assume most people are making purchases in the evening time after getting out of work.

Chart 2- In chart two we look at Gender again and contrast gender with time groups of where their purchases are being made. Surprisingly the data showed the same exact pattern for Females and Male when it comes to how much they shop at each time of the day with as you can see females doubling the amount of sales for each of the time groups.

***Transition:*** *Having gained valuable insights on customer behavior based on time and day of their purchases we now turn our attention to a different dimension of consumer engagement- consumer reviews with Trish.*

**Question 3 Trish:** *What impact does day/ time have on consumer purchases?*

Intro: As Lisa mentioned, I focused on the reviews from customers and recommendations given by Amazon.

**Slide 1 (Review & Rec Helpfulness by Purchase Frequency - bar graphs)**

**Chart 1:** First I started by looking at purchase frequency just to see where we started from. From there I wanted to see when customers were purchasing the most and from Ruby and Lisa we saw that females in general purchase more often and Thursdays and Wednesdays are the days that have the most purchases.

From there I wanted to compare reviews and recommendation helpfulness with purchase frequency. Although tough to tell, both the reviews and recommendations are important for the decisions of the customers. The ‘yes’ and ‘sometimes’ percentages are always greater than those who responded ‘no’ when grouped together.

**Slide 2 (Review & Rec Helpfulness by Purchase Frequency - pie charts)**

**Chart 2:** As we can see here, in review helpfulness both ‘yes’ and ‘sometimes’ percentages are larger than the ‘no’ percentage. But if we look at recommendation helpfulness, ‘no’ is slightly higher than the ‘yes’ category but ‘sometimes’ is still overwhelmingly larger.

This tells us that the reviews and recommendations are being used more often than not being used when it comes to making a purchase.

**Slide 3 (Review & Rec Helpfulness by Gender - bar graph)**

**Chart 3:** I then moved on to looking at gender; I wanted to see if a psecific gender was looking at or using them more often.

From these graphs we can see that females are in fact using the reviews and recommendations far more than any other gender.

**Slide 4 (Review & Rec Helpfulness by Age - bar graph)**

**Chart 4:** From there I looked at age; specifically if there was an age group that was using the reviews and recommendations more so.

From these graphs we can see that the age range of 21-30 year olds are using the reviews and recommendations more than any other age range. Next would be 31-40 and 41-50 year olds.

**Slide 5 (Purchase Category by Review & Rec Helpfulness - bar graph)**

**Chart 5:** Finally I wanted to see if customers used the reviews and recommendations more or less for certain categories of items. We were able to group the categories into these 5 that you see here. You can see that ‘Beauty and Pesonal Care’ and ‘Clothing and Fashion’ are the two categories that are purchased the most followed by ‘Home and Kitchen’ with ‘Groceriese and Gourmet Food’ with the lowest purchases.

But if you look at the reviews it looks like more people used them over saying no and sometimes as well. But compared to the recommendations, more people opped for the ‘sometimes’ response over ‘yes’. But overall the reviews and recommendations aree being used here as well over not being used at all.

*(Conclusion: Wether you are looking at review & recs compared with purchasee freq / gender / age / or purchase categories the reviews & recommendations are being used by the consumers. The percentages show that they are being used more often that not being used at all.)*

***Transition:*** *Having looked at wether customer reviews and personalized recommendations play a role in consumer behavior, lets now take a deeper look at the impacts of cart completion and abandonment from Aida.*

**Question 4 Aida:**

Percentage and impacts of abandonment/browsed and completed carts.

**1st Slide**

Key Findings

*Observation , analysis and Implications*

Browsing Behavior: (Findings)

* + A common behavior among customers is active browsing, with a significant number frequently engaging in browsing activities.
  + "Yes" is the second most frequent response, indicating high engagement with the browsing feature.
  + A substantial number of customers also browse "Maybe," suggesting moderate browsing activity.
  + A smaller group of customers do not engage in browsing, as indicated by the "No" category.

Browsing Optimization: (Implications)

* + Given the high engagement with browsing, Amazon could consider enhancing the browsing experience, offering personalized content, and optimizing search result exploration to drive conversions.

Added to the cart behavior by cart completion: (Findings)

* + As we can observe in the pie chart, only the 3rd part of the customers do actually complete their purchases after adding them to the cart.

Enhancing Cart Completion:

* + To increase the conversion rate, businesses may consider strategies to encourage customers who add items to the cart to complete their purchases. This could include providing incentives, or improving the shopping experience.
  + The weak correlation indicates that while there is a positive relationship between adding items to the cart and completing a purchase, there are other factors influencing purchase behavior.
  + Further analysis and experimentation may be needed to identify these factors and optimize the conversion process effectively.

**2nd Slide**

Save-for-Later Behavior: (Findings)

* + "Sometimes" is the most common behavior among customers, indicating that a significant number occasionally save items for later.
  + "Often" is the next most frequent behavior, suggesting that a substantial portion of customers frequently use the "Save for Later" feature.
  + Fewer customers use "Rarely," "Never," or "Always," indicating these behaviors are less common.

Saved for Later after added to cart

* + The majority of customers did not save items for later after adding them to the cart.
  + The Pearson correlation coefficient of 0.06 suggests a weak positive correlation between the frequency of adding items to the cart and the frequency of saving items for later.

Saved for Later Feature: (Implications)

* + While most customers do not use this feature, Amazon could explore ways to promote the "Save for Later" function and enhance its utility. It might be particularly appealing to those who frequently engage with this feature.

**3rd Slide**

Cart Behavior (abandoment and completion)

* + A significant portion (40.70%) of shopping carts are abandoned, indicating that many customers initiate the shopping process but do not complete their purchases.
  + On the positive side, 59.30% of shopping carts are completed, showing that the majority of customers do complete their purchases.

Cart Abandonment Factors (Findings)

* + The most frequent reasons for cart abandonment are "Found a better price elsewhere" and "Changed my mind or no longer need the item."
  + "High shipping costs" are also a factor in some cart abandonments, highlighting the impact of shipping costs on purchase decisions.

Cart Abandonment: (Implications)

* + Understanding the common abandonment factors, such as price sensitivity and shipping costs, provides insights for targeted strategies to reduce cart abandonment rates.
  + Amazon can offer competitive prices, transparent shipping costs, and effective price comparison tools to mitigate these factors.

**4th Slide**

Cart Abandonment by Age Group (Findings)

* + Change my mind or no longer need the item" has the highest frequency in the '30-40' and '50+' age categories but still is a common abandonment factor across all age groups.
  + "Found a better price elsewhere" is a significant abandonment factor in the younger and middle-aged customer groups. This may indicate that younger and middle-aged customers are more price-sensitive and actively search for better deals.
  + "High shipping costs" as an abandonment factor is relatively less frequent but still notable in the '30-40' and '50+' age groups. This suggests that shipping costs might be a more significant concern for these age groups.

Correlation with Age

* + The correlation Coefficient in 0 means that there is virtually no linear correlation between age and the numerical representation of cart abandonment factors indicating that age does not predict or explain changes in the numerical representation of cart abandonment factors.

Age-Specific Strategies: (Implications)

* + Tailoring strategies to age groups can be effective. For example, offering deals and promotions to price-sensitive younger customers, and focusing on clear value propositions for middle-aged customers.

**5th Slide**

Abandonment and Completion Rates by Gender (Findings)

* + Females have the highest abandonment rate (41.48%), while "Others" have the lowest (34.21%).
  + Males have the highest completion rate (61.97%), while "Prefer not to say" have the lowest (56.74%).

Abandonment and Completion Rates by Gender (Implications)

* + The analysis by gender provides valuable insights into the shopping behavior of different gender groups. This information can be used to tailor marketing strategies and improve the shopping experience for specific customer segments.
  + While females have the highest number of both abandoned and completed carts, the completion rate for males is higher. This suggests that males may be more likely to complete their purchases once they add items to their carts.
  + The "Others" category has a relatively lower abandonment rate and a higher completion rate, indicating that this group tends to complete their purchases more often and abandon carts less frequently.
  + To improve cart completion rates and reduce abandonment rates, businesses can target specific gender groups with tailored strategies. For example, understanding the reasons behind cart abandonment in each group can help address specific pain points and increase conversion rates.

Conclusion

In conclusion, the analysis of Amazon's customer behavior dataset has revealed valuable insights into cart behavior and abandonment rates. These findings provide a foundation for Amazon to refine its strategies, enhance the customer experience, and increase conversion rates, ultimately boosting its overall success.

***Transition: This way we conclude the analysis of our project . I leave you with Trish who will give us a wrap up of everything we have covered***

**Wrap up:**

**Ruby:** *This data provides valuable insights into gender AND AGE-specific purchasing habits. Tailoring marketing and product strategies based on these insights might lead to better sales and customer engagement.*

**Lisa:** *Understanding when consumers make purchases is crucial and we will unveil patterns in the timing and frequency of the purchase to help optimize sales and ensure ad spent strategy is being distributed at the most effective times.*

**Trish:** *Wether you are looking at reviews & recommendations compared with purchase freq / gender / age / or purchase categories the reviews & recommendations are being used by the consumers. The percentages show that they are being used more often that not being used at all.*

**Aida:** *Understanding the customer’s common behaviors, from the moment they start searching for a product until they complete or not the purchase, can help to create strategies to improve and optimize the experience and this way, achieve the desired results.*

**Hypothesis:**

**(Question and results )**