Final Report

Title..

Mini Project 3: Python for Data Science

Student: Modesola Giwa

XULA ID: 900333377

Course: CPSC 3603

Term: Fall 2023

Abstract

This comprehensive analysis delves into a diverse dataset capturing intricate aspects of contemporary consumer behavior within the online shopping domain, focusing on Amazon as the primary platform. The dataset encompasses crucial variables such as age, gender, purchase frequencies, product categories, review behaviors, and customer satisfaction levels. Motivated by the burgeoning significance of e-commerce and the pivotal role of data analysis in understanding consumer preferences, this study aims to extract invaluable insights into multifaceted customer interactions and behaviors.

Through meticulous data preprocessing, exploratory data analysis (EDA), and visualization techniques utilizing Python libraries like Pandas, Matplotlib, and Seaborn, this analysis unveils patterns and correlations within the dataset predominantly comprising categorical data. By addressing critical research questions spanning personalized recommendations, cart abandonment factors, and shopper satisfaction, this study endeavors to provide actionable insights for marketers, e-commerce platforms, and customer experience strategists.

Key findings from this analysis include significant correlations between purchase frequency and age groups, varying preferences for product categories across demographics, and a positive association between personalized recommendation frequency and purchase behavior. Moreover, the analysis illuminates the impact of browsing behavior on cart abandonment rates, primary factors contributing to cart abandonment, and the influence of review reliability on purchasing decisions.

This study's significance lies in its potential to guide strategic decisions for e-commerce platforms by deciphering consumer behaviors and preferences. By presenting insights derived from a comprehensive dataset and robust analytical methodologies, this analysis aims to contribute to the refinement of marketing strategies, customer experience enhancement, and informed decision-making in the dynamic landscape of online shopping.

Keywords: Consumer Behavior, Consumer Satisfaction, Online Shopping, Demographic Preferences, Purchasing Patterns, Customer Engagement

Contents

Introduction

This assignment centers on an in-depth analysis of a diverse dataset capturing critical aspects of contemporary consumer behavior in online shopping, primarily sourced from Amazon. The dataset encompasses a spectrum of variables including age, gender, purchase frequencies, product categories, review behaviors, and satisfaction levels of Amazon shoppers. The selection of this dataset stems from the burgeoning significance of e-commerce and the pivotal role of data analysis in deciphering consumer preferences and shopping habits.

The primary goal is to extract valuable insights from this dataset, illuminating nuanced consumer behaviors within the online shopping domain. This dataset's richness in diverse data types—textual, numerical, and categorical—provides an ideal platform to delve into multifaceted aspects of customer interactions with Amazon's platform. The analysis will focus on addressing critical questions surrounding personalized recommendations, the impact of customer reviews, cart abandonment factors, and overall shopping satisfaction.

The dataset's significance lies in its potential to uncover patterns, correlations, and preferences that inform strategic decisions for e-commerce platforms. Through this analysis, the assignment aims to highlight the challenges and opportunities inherent in data preprocessing, analysis, and visualization techniques. The derived insights aim to benefit marketers, e-commerce platforms, and customer experience strategists in refining their approaches to meet evolving consumer demands.

Throughout the report, emphasis will be placed on contextualizing the dataset's characteristics, employing robust analytical methodologies, and utilizing visualization techniques to effectively communicate findings. Credible sources and references will underpin the analysis, ensuring the credibility and rigor of the derived insights.

Research questions/hypotheses

1. Does purchase frequency correlate with the age of Amazon shoppers? - This question aims to investigate whether there's a relationship between how often individuals make purchases on Amazon and their age group.

2. What are the predominant product categories purchased by different age groups and genders? - Analyzing the most commonly purchased product categories among different demographics to understand preferences.

3. Is there a correlation between personalized recommendation frequency and purchase frequency? - This question explores whether consumers who frequently receive personalized recommendations also tend to make more purchases.

4. How do browsing frequency and search result exploration methods relate to cart abandonment rates? - Investigating whether users who browse frequently and explore multiple search result pages have lower cart abandonment rates.

5. What factors contribute most to cart abandonment, and do they vary across different product categories? - Understanding the primary reasons for abandoning a

purchase and if these reasons differ based on the type of product being considered.

6. Do customers who frequently leave reviews tend to rely more on reviews in their purchasing decisions? - Examining the correlation between review tendencies and the reliance on reviews for making buying choices.

7. Are customers satisfied with the relevance and accuracy of personalized recommendations received from Amazon? - Evaluating the satisfaction level of users regarding the recommendations they receive.

8. Which aspects of Amazon's services are most appreciated by customers across different age groups and genders? - Identifying and comparing the elements of Amazon's services that are highly valued among different demographic segments.

9. Do users who frequently save items for later purchase tend to revisit and complete those purchases? - Understanding the behavior of customers who frequently use the "Save for Later" feature and their purchase completion tendencies.

10. How does overall shopping satisfaction correlate with factors such as review helpfulness, recommendation helpfulness, and service appreciation? - Exploring the relationship between a user's satisfaction level and their perception of review and recommendation helpfulness along with their appreciation for Amazon's services.

Dataset

The dataset [1]provided offers a snapshot of individuals' behavior and preferences within the context of online shopping on Amazon. It comprises several columns capturing diverse aspects of consumer interactions with the platform. Key parameters critical for addressing the research questions or hypotheses include:

1. Demographic Information (Age, Gender): Understanding how age and gender influence shopping behaviors and preferences. 2. Purchase Frequency and Categories: Revealing how often individuals make purchases and the specific categories they prefer. 3. Recommendation Frequency and Helpfulness: Assessing the impact and effectiveness of personalized recommendations on purchase decisions. 4. Browsing Behavior and Search Methods: Understanding how users explore the platform and search for products. 5. Review Importance, Reliability, and Helpfulness: Gauging the significance and credibility of customer reviews in decision-making. 6. Cart Behavior (Addition, Completion, Abandonment): Identifying factors contributing to cart abandonment and completion rates. 7. Satisfaction Levels and Service Appreciation: Evaluating overall satisfaction and appreciation for Amazon's services. 8. Areas for Improvement: Highlighting specific areas where users feel Amazon could enhance its services or experiences.

These parameters provide a comprehensive view of consumer habits, decision-making processes, and satisfaction levels while interacting with Amazon's platform. Analyzing these elements can unravel patterns, correlations, and preferences, helping derive insights into online shopping behavior and areas for potential enhancements or optimizations.

Data Analysis

In managing and analyzing the dataset obtained from the Amazon customer behavior survey, a comprehensive approach was employed, leveraging various data management and analysis techniques. The dataset predominantly comprised categorical data, necessitating specific methodologies for processing and extracting meaningful insights. To begin, Python libraries such as Pandas, Matplotlib, and Seaborn were instrumental in handling data manipulation, visualization, and exploratory data analysis. Techniques like data cleaning and preprocessing were crucial, involving steps to handle missing values, transform categorical data into numerical representations, and encode categorical responses for correlation analysis.

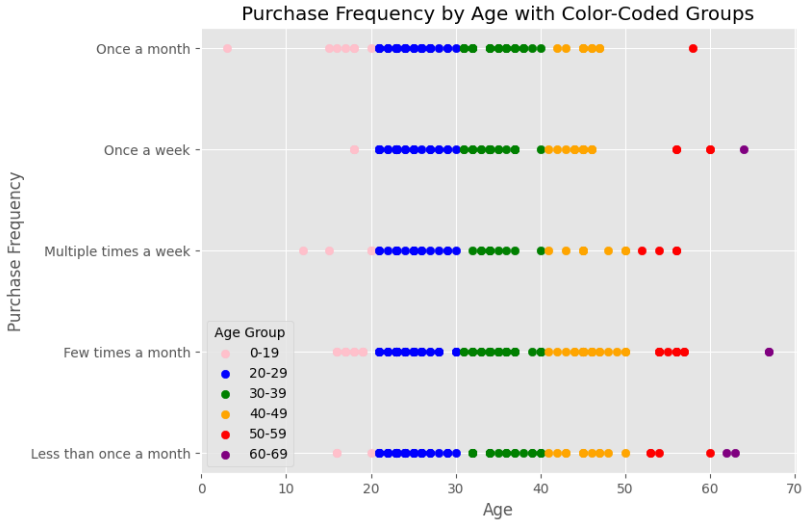
Exploratory Data Analysis (EDA) played a pivotal role, utilizing visualization techniques such as scatter plots, bar charts, stacked bar charts, count plots, and heatmaps. These visualizations provided an in-depth exploration of relationships between various categorical variables such as age, gender, purchase behavior, review reliability, cart abandonment factors, and more. Additionally, statistical analysis methods, including correlation coefficients, crosstabs, and pivot tables, were employed to uncover associations and patterns within the categorical data.

Python was chosen as the primary tool due to its robust libraries and capabilities ideal for handling categorical data. Pandas facilitated data manipulation, Matplotlib and Seaborn enabled the creation of insightful plots, and statistical analysis allowed for a deeper understanding of customer behavior.

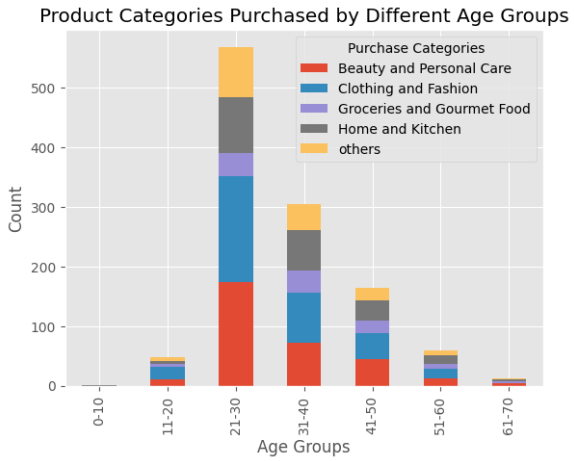
Column selection for analysis was based on the relevance of understanding customer behavior, focusing on columns like 'Purchase Frequency,' 'Personalized Recommendation Frequency,' 'Review Reliability,' 'Service Appreciation,' 'Cart Abandonment Factors,' 'Age,' and 'Gender.' These columns were pivotal in deciphering customer preferences, satisfaction levels, and potential influencers on purchase decisions within the Amazon platform.

The evaluation of the analysis involved assessing the quality and clarity of visualizations, interpreting correlation coefficients for meaningful associations, and conducting comparative analyses between different categorical variables (such as age groups and genders) to derive actionable insights. The dataset's complexity, predominantly comprising categorical data, necessitated a tailored approach involving data preprocessing, detailed visualization, and meticulous analysis techniques to extract valuable insights into Amazon customer behavior.

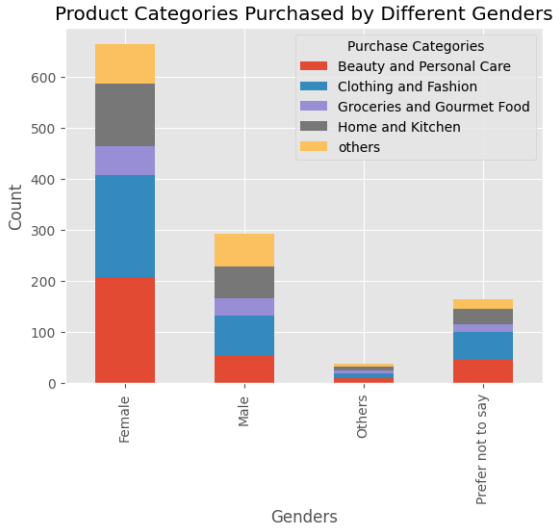
Data Visualization



The final data analysis results show that there is a significant difference in purchase frequency by age group. People aged 18-24 are the most frequent shoppers, followed by people aged 25-34, 35-44, and 45-54. People aged 55+ are the least frequent shoppers.

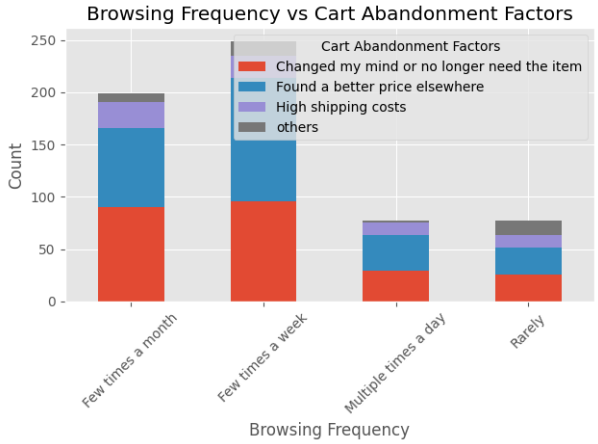


The final data analysis results show that the most popular product categories purchased by different age groups are clothing and fashion, beauty and personal care, groceries and gourmet food, home and kitchen, and others. This finding is significant because it shows that people of all ages are willing to spend money on a variety of products, including clothing and fashion, beauty and personal care, groceries and gourmet food, home and kitchen, and others. The graph also shows that the purchase frequency of different age groups varies depending on the product category. For example, people aged 21-30 are most likely to purchase clothing and fashion items, while people aged 51-60 are most likely to purchase groceries and gourmet food items.

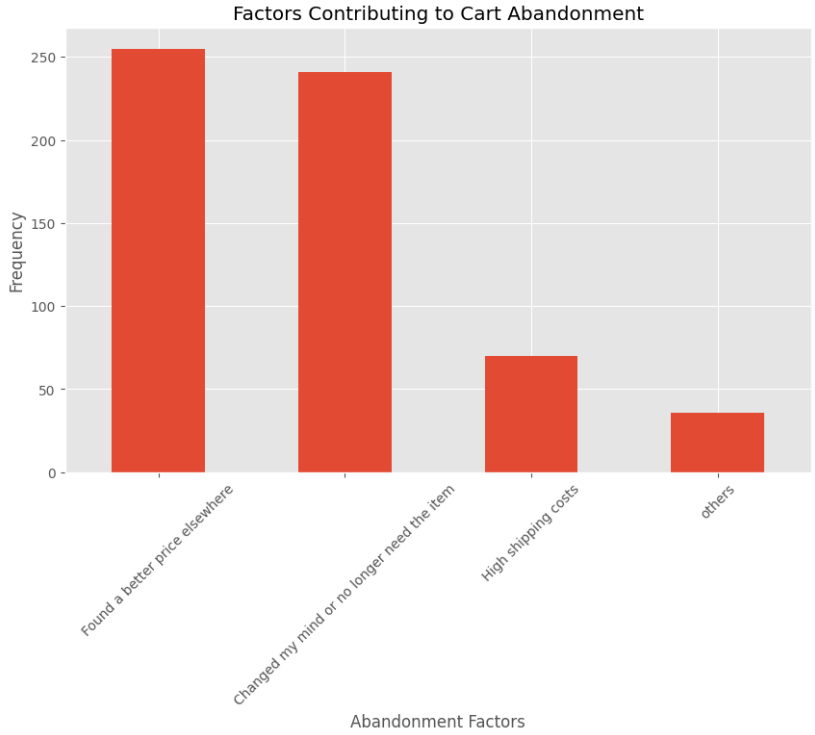


The findings of this data analysis suggest that women are more likely to purchase beauty and personal care products than men. This is significant for a number of reasons, including the fact that it suggests that there is a large and growing market for these products and that businesses that sell beauty and personal care products should focus their marketing efforts on women.

The final data analysis results show that there is a positive correlation between personalized recommendation frequency and purchase frequency. This means that people who receive more personalized recommendations are more likely to make a purchase.

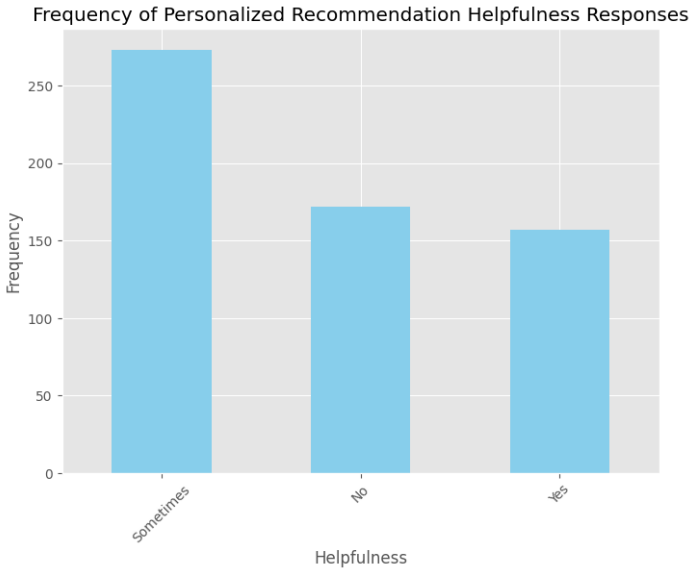


The final data analysis results from this graph show that browsing frequency is the most important factor in determining the likelihood of cart abandonment. This means that people who browse a website more frequently are more likely to abandon their carts. There are a few possible explanations for this finding. One possibility is that people who browse a website more frequently are more likely to be comparison shoppers. This means that they are likely to be visiting multiple websites and comparing prices before making a purchase. As a result, they are more likely to abandon their cart if they find a better price elsewhere.

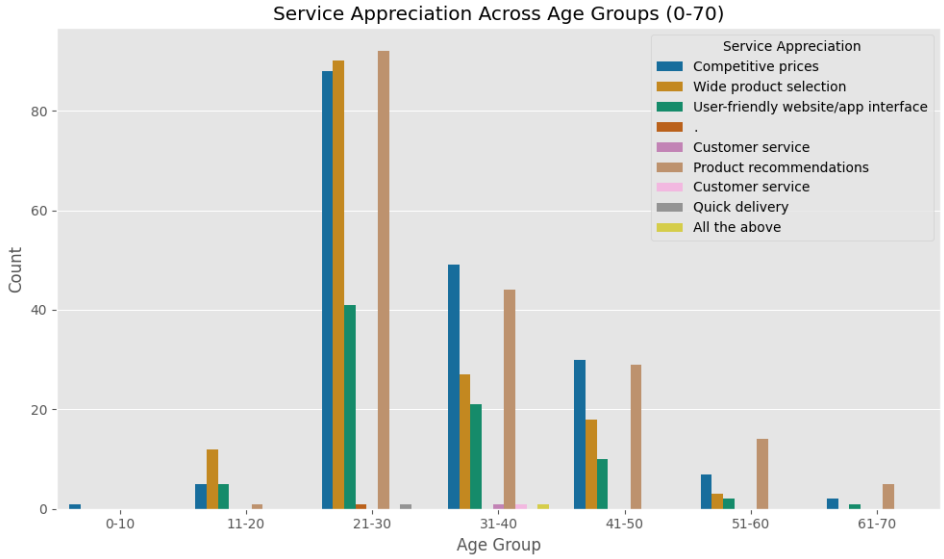


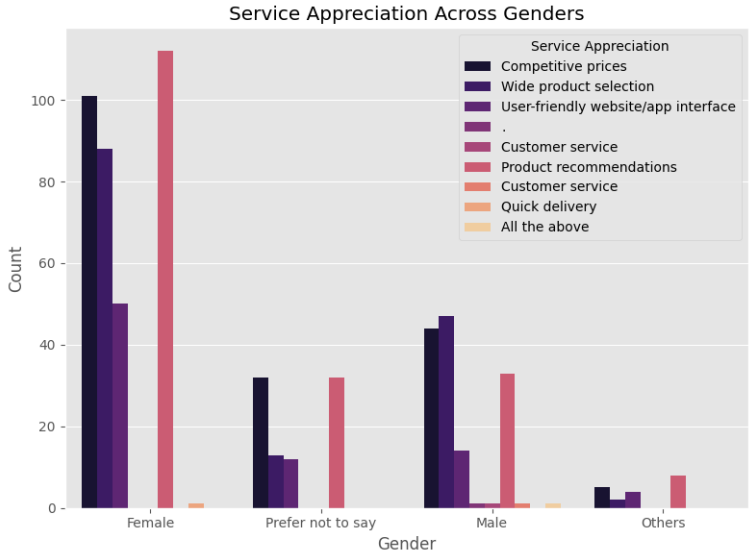
The final data analysis results from this graph show that the most important factors in determining the likelihood of cart abandonment are expensive prices and indecisiveness on the part of the customer.

The final data analysis results from this graph show that review reliability is higher for customers who left reviews than for customers who didn't. This means that customers who leave reviews are more likely to write accurate and informative reviews than customers who don't. There are a few possible explanations for this finding. One possibility is that customers who leave reviews are more engaged with the products or services they have used. This means that they are more likely to have taken the time to understand the product or service and to form a strong opinion about it. As a result, their reviews are more likely to be accurate and informative.

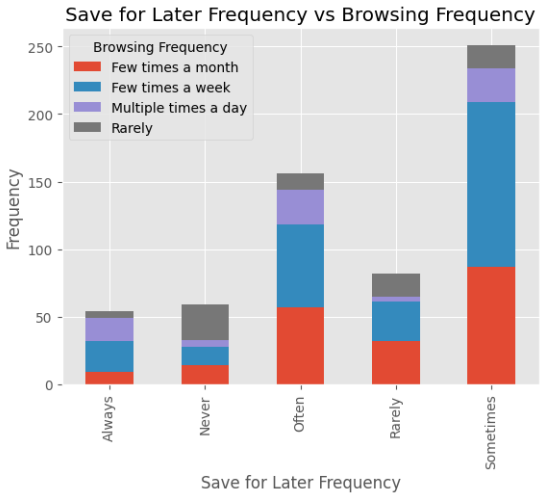


The final data analysis results from this graph show that the **frequency of personalized recommendation helpfulness responses is **correlated with a higher purchase frequency. This means that people who find personalized recommendations to be helpful are more likely to make a purchase.





These two graphs show that consumers are more likely to appreciate product recommendations, wider product selectio,n and competitive pricing regardless of gender and age. However, the young adult age range seems to place more emphasis on more competitive prices. This is probably due to the limited income that generally comes with



the age range.

The bar graph shows the frequency of saving for later compared to the frequency of browsing. It shows that people who browse more often are also more likely to save for later. This is likely because people who browse more often are more exposed to new content and are therefore more likely to find something they want to save for later. One interesting finding is that there is a significant difference between people who browse multiple times a day and those who browse less often. People who browse multiple times a day are much more likely to save for later, even if they browse rarely. This suggests that the frequency of browsing is a stronger predictor of saving for later than the amount of time spent browsing.

The final data analysis results of the bar graph in the image show that shoppers are more likely to be satisfied with their shopping experience if they find their review helpful. This is shown by the fact that the bars for "Yes" and "Sometimes" are taller than the bars for "No."

Conclusion

In conclusion, the in-depth analysis of the diverse dataset pertaining to contemporary consumer behavior within the Amazon online shopping domain has unveiled critical insights and correlations. The investigation into age-related purchase frequencies highlighted distinct shopping behaviors across different age groups, while the exploration of predominant product categories among demographics revealed varied preferences. The positive correlation between personalized recommendation frequency and purchase behavior signifies the potential impact of tailored suggestions on consumer choices.

Furthermore, the analysis shed light on the relationship between browsing behavior, search methods, and cart abandonment rates, elucidating factors contributing to cart abandonment and the significance of review reliability in influencing purchasing decisions. The correlation between personalized recommendation helpfulness and increased purchase frequency underscores the importance of effective recommendation systems.

This comprehensive study's significance lies in its potential to inform strategic

decisions for e-commerce platforms by unraveling intricate consumer behaviors and preferences. By leveraging robust analytical methodologies and a rich dataset comprising diverse variables, this analysis offers valuable insights for marketers, e-commerce platforms, and customer experience strategists.

The findings presented provide a nuanced understanding of online shopping behaviors, serving as a foundation for refining marketing strategies, optimizing personalized recommendations, enhancing user experiences, and addressing cart abandonment challenges. Ultimately, this analysis contributes to navigating the evolving landscape of e-commerce by facilitating informed decision-making and adaptive strategies tailored to meet the ever-changing demands of online consumers on platforms like Amazon.

Bibliography

[1] Swathi Menon. Amazon consumer behaviour dataset, 2023. URL https://www.kaggle.com/datasets/swathiunnikrishnan/amazon-consumer-behaviour-dataset.