Aayush Khandekar

Individual Project Report

An Analysis of Consumer Behavior Patterns on E-Commerce Platform

Problem Statement:

In the rapidly expanding realm of e-commerce, understanding the intricate dynamics of consumer behavior patterns stands as a critical pursuit for both academic research and practical applications. This study aims to comprehensively analyze the consumer behavior patterns with the Amazon e-commerce platform.

While there is little existing information on consumer behavior in the broader context of e-commerce, there remains a clear necessity for a detailed investigation specifically centered on the dynamics within Amazon's platform. The platform's unparalleled dominance in the global marketplace presents a unique and complex environment for consumer interactions, necessitating a closer examination of the underlying factors that drive consumer decisions, preferences, and purchasing behaviors.

The evolution of e-commerce has brought about a significant shift in consumer behavior patterns by providing accessibility and convenience. Amazon is a leader among the e-commerce platforms, changing the way customers engage with items and shops. This study explores the complex aspects of consumer behavior on Amazon with the goal of analyzing the reasons and inclinations that influence customers in this online marketplace.

Amazon, a massive online retailer, has a unique perspective on consumer behavior because of its presence in a variety of markets. The research investigates the multitude of factors that impact customer decisions. In the highly competitive world of e-commerce, companies looking to improve their marketing tactics and client interaction must comprehend these behavioral patterns just as much as Amazon does.

By analyzing consumer behavior on Amazon, this project aims to study the fundamental factors that influence consumer behavior on e-commerce websites like Amazon. It also illuminates how consumer preferences are changing and the tactics that help Amazon maintain its enormous customer base in the face of a constantly shifting digital retail environment.

Dataset:

While searching and exploring for suitable datasets, I used Google Dataset Search and found the dataset linked below.

The Analysis was conducted on the following dataset (Dataset Link):

https://www.kaggle.com/datasets/swathiunnikrishnan/amazon-consumer-behaviour-dataset

The dataset comprises various attributes related to consumer behavior on Amazon's e-commerce platform. It encompasses demographic details such as age and gender, alongside crucial factors influencing purchasing decisions. These factors include purchase frequency, preferred purchase categories, reliance on personalized recommendations, browsing habits, search methods, and importance placed on customer reviews. Additionally, it explores behaviors like cart usage, abandonment factors, review tendencies, satisfaction levels, and opinions on Amazon's services. The dataset looks at how different types of people shop on Amazon and what provides satisfaction to the consumers. It tries to understand how age, gender, and the way people shop are connected.

Tool Choice and Justification:

R as a statistical programming language was selected for its statistical analysis capabilities, ideal for conducting in-depth exploratory data analysis (EDA). Its extensive library of packages and functions allowed for comprehensive data exploration, manipulation, and statistical modeling.

Tableau was preferred for its user-friendly interface and powerful visualization tools, making it possible to create visually appealing plots, interactive dashboards, and compelling storytelling. Its ease of use and flexibility in building visual narratives made it an excellent choice for presenting insights derived from the data analysis conducted in R.

Exploratory Data Analysis:

Online Shopping by Gender Statistics:

Each row in the dataset represents a unique transaction. Statistically, 58% of online purchases are done by females while 23% of online purchases are done by males.

```
            Female               Male            Others Prefer not to say
              352                 142                19                89
[1] "Female Percentage : 58.471761"
[1] "Male Percentage : 23.588040"
[1] "Others Percentage : 3.156146"
[1] "Prefer Not to Say Percentage : 14.784053"
```

Fig.1. Online Shopping Statistics by Gender

Visualizing Age:

The age range of 20 to 25 years old is the most common for online purchases, with a subsequent decrease in transactions observed as consumers age. While those in the 25–30 age range do make purchases, those in the 30–35 age range do so more frequently. Customers under the age of ten and those over the age of fifty made the majority of the lease purchases. Given that a customer between the ages of 0 and 5 made a transaction, there is also an outlier that can be observed.

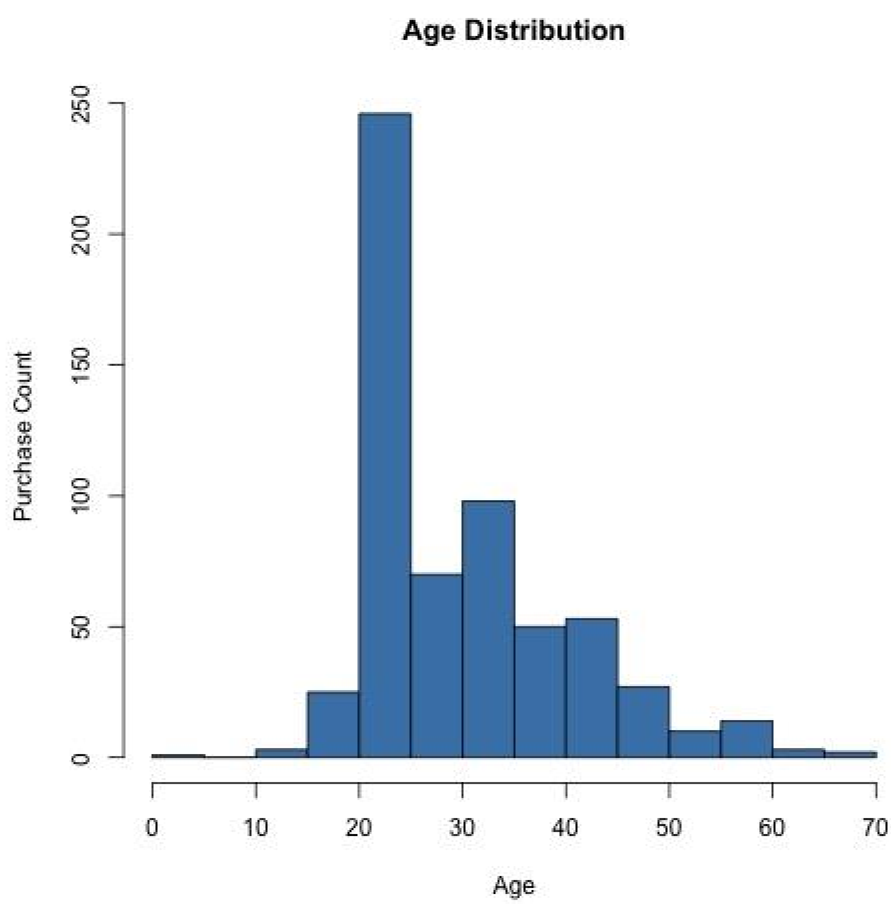

Fig.2. Purchase Count Across Different Age Groups

Descriptive Statistics:

The age range of online customers spans from 3 to 67 years. The first quartile (25th percentile) falls at 23 years, and the third quartile (75th percentile) falls at 36 years. The average age (mean) of customers is 30.79 years, while the median age (50th percentile) is 26 years.

Similarly, The average rating given by customers for the reviews is 2.48, with a median rating of 3. The first quartile (25th percentile) of ratings is 1, and the third quartile (75th percentile) is 3, as anticipated.

```
  Timestamp               Age            Gender           Purchase_Frequency Purchase_Categories
 Length:602         Min.   : 3.00   Length:602         Length:602         Length:602
 Class :character   1st Qu.:23.00   Class :character   Class :character   Class :character
 Mode  :character   Median :26.00   Mode  :character   Mode  :character   Mode  :character
                    Mean   :30.79
                    3rd Qu.:36.00
                    Max.   :67.00
 Personalized_Recommendation_Frequency_6 Browsing_Frequency Product_Search_Method
Search_Result_Exploration
 Length:602                              Length:602         Length:602            Length:602
 Class :character                        Class :character   Class :character      Class
:character
 Mode  :character                        Mode  :character   Mode  :character      Mode
:character



 Customer_Reviews_Importance Add_to_Cart_Browsing Cart_Completion_Frequency
Cart_Abandonment_Factors
 Min.   :1.00               Length:602           Length:602                Length:602
 1st Qu.:1.00               Class :character     Class :character          Class :character
 Median :3.00               Mode  :character     Mode  :character          Mode  :character
 Mean   :2.48
 3rd Qu.:3.00
 Max.   :5.00
 Saveforlater_Frequency Review_Left        Review_Reliability Review_Helpfulness
 Length:602             Length:602         Length:602         Length:602
 Class :character       Class :character   Class :character   Class :character
 Mode  :character       Mode  :character   Mode  :character   Mode  :character


 Personalized_Recommendation_Frequency_18 Recommendation_Helpfulness Rating_Accuracy
Shopping_Satisfaction
 Min.   :1.000                            Length:602                 Min.   :1.000   Min.
:1.000
 1st Qu.:2.000                            Class :character           1st Qu.:2.000   1st
Qu.:2.000
 Median :3.000                            Mode  :character           Median :3.000   Median
:2.000
 Mean   :2.699                                                       Mean   :2.673   Mean
:2.463
 3rd Qu.:3.000                                                       3rd Qu.:3.000   3rd
```

Fig. 3. Summary Statistics of Dataset

<u>Encoding the values</u>

```
{r}

encoded_values <- factor(data$Purchase_Frequency, levels = c("Less than once a month", "Once a month", "Few times a month", "Once a week", "Multiple times a week"))

data$Purchase_Frequency_Encode <- as.numeric(encoded_values)
```

Fig.4. Encoding Categorical Data

Since the dataset is not completely quantitative, encoding of the data was performed in order to convert the categorical data into numerical data for performing exploratory analysis and to generate visualizations.

The encoding generated a new column (“Purchase_Frequency_Encode”) with numerical values after converting the categorical data into suitable and appropriate numerical data.

<u>Visualizing Distributions of discrete variable: Purchase Frequency</u>

1. <u>Encoding Representation:</u>

   1 represents "Less than once a month" 2 represents "Once a month" 3 represents "Few times a month" 4 represents "Once a week" 5 represents "Multiple times a week"

2. <u>Data Distribution Analysis:</u>

   203 customers made purchases "Few times a month," followed by 124 customers who made purchases "Less than once a month." This count is close to the number of customers, which is equivalent to those who made purchases "Once a week." There are only 56 customers who make purchases "Multiple times a week."

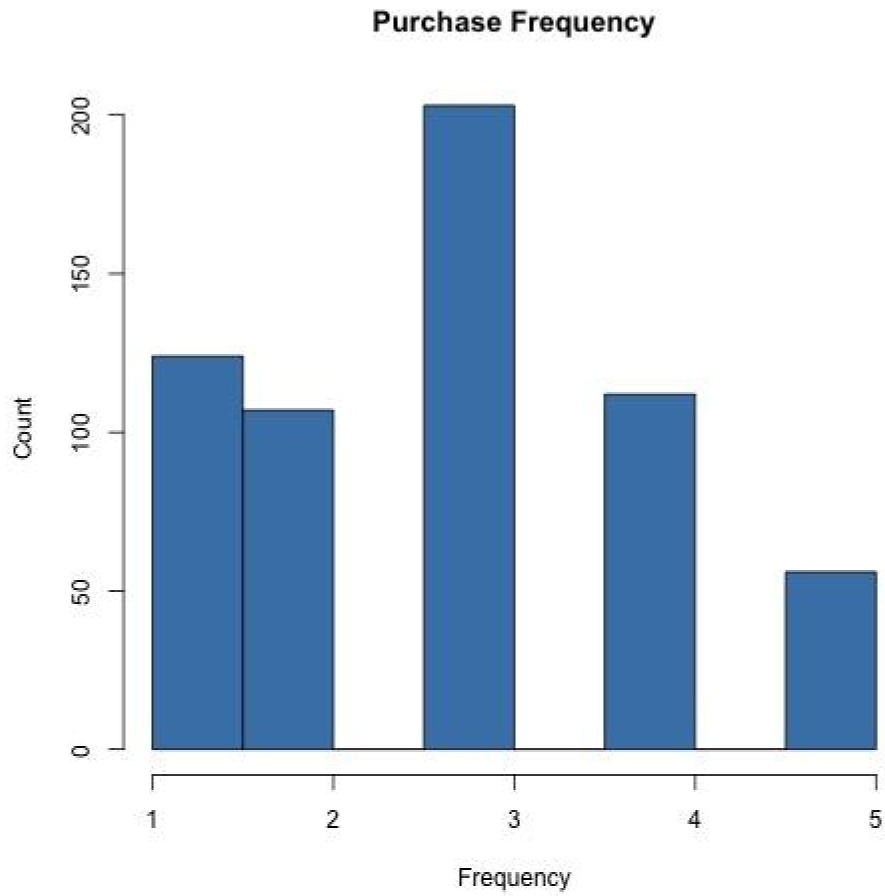

Fig.5. Visualizing Purchase Frequency

Generating Frequency Count:

1 represents "Less than once a month" 2 represents "Once a month" 3 represents "Few times a month" 4 represents "Once a week" 5 represents "Multiple times a week"

1	2	3	4	5
124	107	203	112	56

Fig. 6. Count for Purchase Frequency Observed

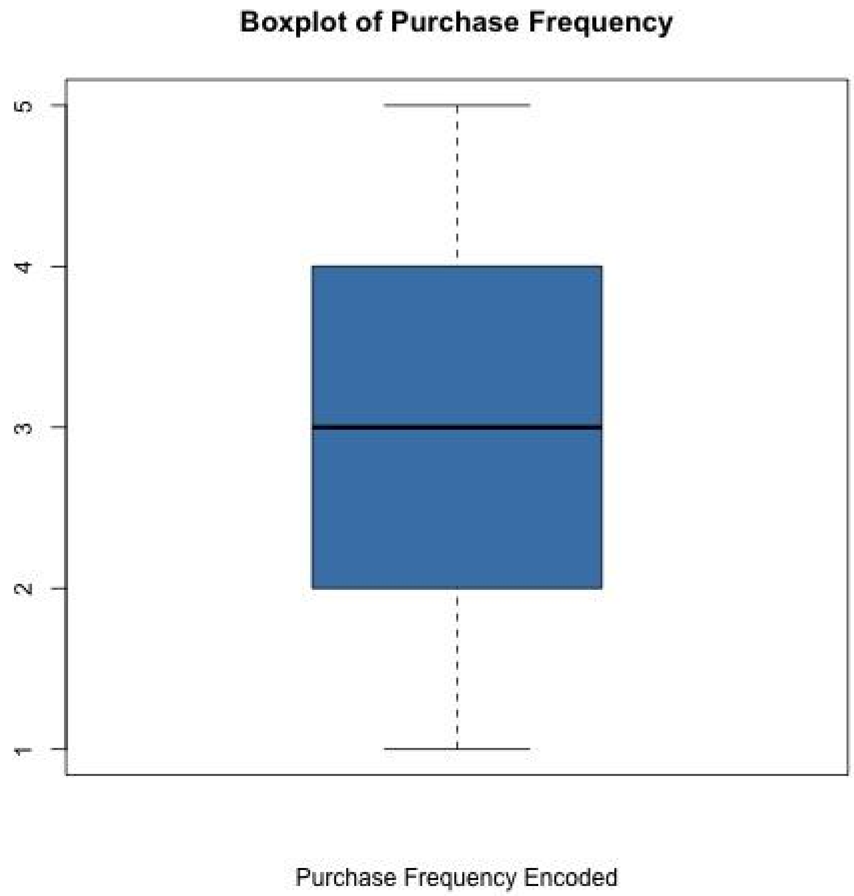

Fig.7. Boxplot of Purchase Frequency

<u>Correlation Matrix:</u>

1. <u>Age and Purchase_Frequency_Encode (0.0917):</u>

   There is a weak positive correlation (0.0917) between a customer's age and their purchase frequency encoding. This suggests a slight tendency for older customers to have a slightly higher frequency of purchases.

2. <u>Age and Shopping_Satisfaction (0.0039):</u>

   There appears to be an almost negligible positive correlation (0.0039) between a customer's age and their shopping satisfaction. This correlation is very close to zero, indicating almost no linear relationship between a customer's age and their satisfaction level while shopping.

3. <u>Age and Rating_Accuracy (-0.0102):</u>

The correlation between age and rating accuracy is negative but extremely close to zero (-0.0102), indicating almost no linear relationship between a customer's age and the accuracy of their ratings.

4. Purchase_Frequency_Encode and Shopping_Satisfaction (0.0343):

   There is a very weak positive correlation (0.0343) between a customer's purchase frequency encoding and their shopping satisfaction. This suggests a slight tendency for customers who make purchases more frequently (as encoded) to have slightly higher satisfaction levels while shopping.

5. Purchase_Frequency_Encode and Rating_Accuracy (-0.0569):

   A weak negative correlation (-0.0569) exists between a customer's purchase frequency encoding and their rating accuracy. This implies that customers with a higher purchase frequency tend to provide slightly less accurate ratings.

6. Shopping_Satisfaction and Rating_Accuracy (0.5140):

   There is a moderate positive correlation (0.5140) between shopping satisfaction and rating accuracy. This suggests that customers who are more satisfied with their shopping experience tend to provide more accurate ratings.

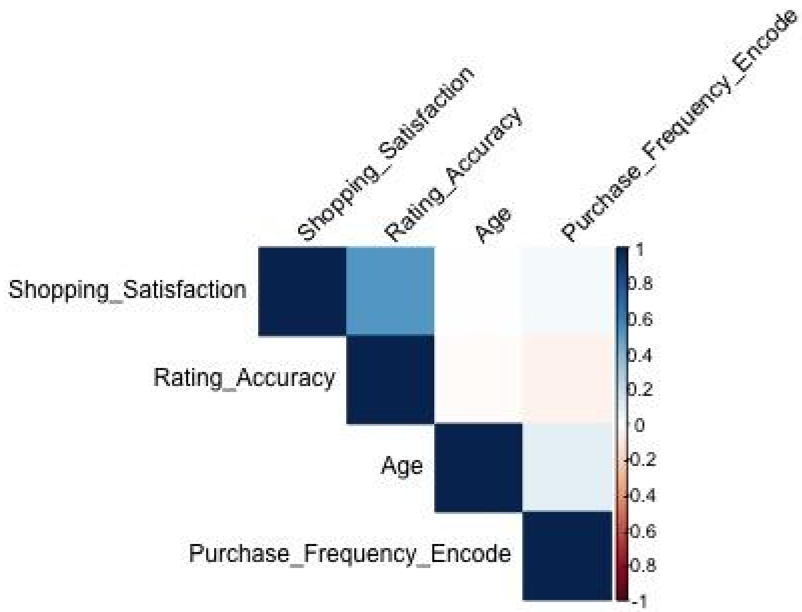

Fig.8. Correlation Matrix Plot

	Age	Purchase_Frequency_Encode	Shopping_Satisfaction	Rating_Accuracy
Age	1.000000000	0.09171786	0.003934349	-0.01020189
Purchase_Frequency_Encode	0.091717856	1.00000000	0.034341942	-0.05689663
Shopping_Satisfaction	0.003934349	0.03434194	1.000000000	0.51396239
Rating_Accuracy	-0.010201888	-0.05689663	0.513962386	1.00000000

Fig.9. Correlation Matrix by Numbers

Research Questions and Analysis:

Question 1: How does age differ among customers with different purchase frequencies?

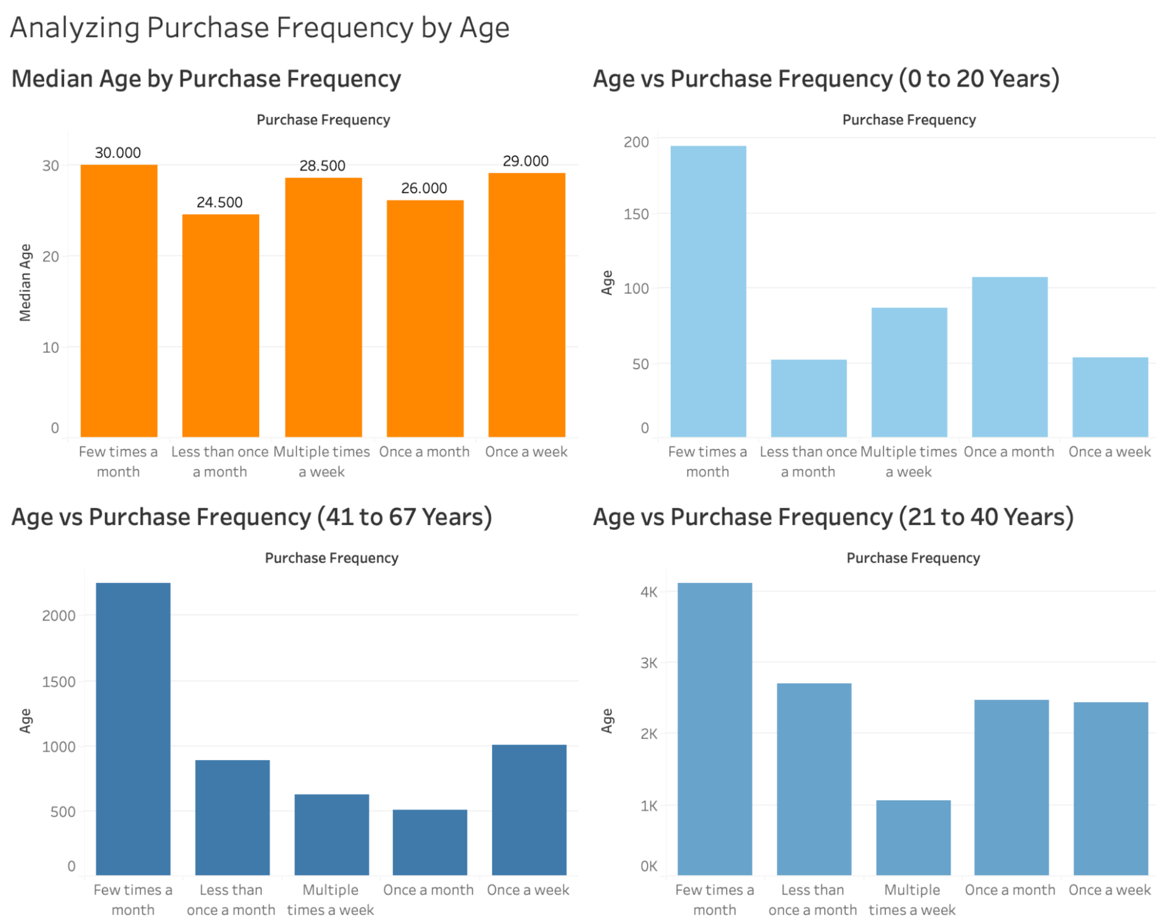

Fig.10. Analyzing Purchase Frequency by Age

Analyzing purchase frequency of products on an e-commerce platform by age is crucial for understanding varying consumer behaviors across different age groups. This analysis provides insights into age-specific shopping habits, preferences, and trends. It aids in optimizing inventory, enhancing customer engagement, and improving overall sales performance by aligning strategies with the specific demands of each age group.

As evident from the plot above, most of the purchases have been made by the age group of 21 to 40 years with the frequency of a few times a month. This age group is active on ecommerce platforms and has the tendency to shop online more than a few times a month. With an average purchase frequency of both once a month and once a week, this age group has all sorts of different customers. The Median Age by Purchase Frequency plot shows that the purchase distribution across all values of purchase frequency is somewhat evenly distributed. Hence, there are consumers of all types who make purchases few times a month, less than a month, multiple times a month, once a month and even once a week. For age groups 41 to 67 years old, the number of consumers are found to be less, as evident from the plot. Also, the consumers within that age group make purchases “few times a week”.

Question 2: What is the distribution of rating accuracy over the timespan of the dataset?

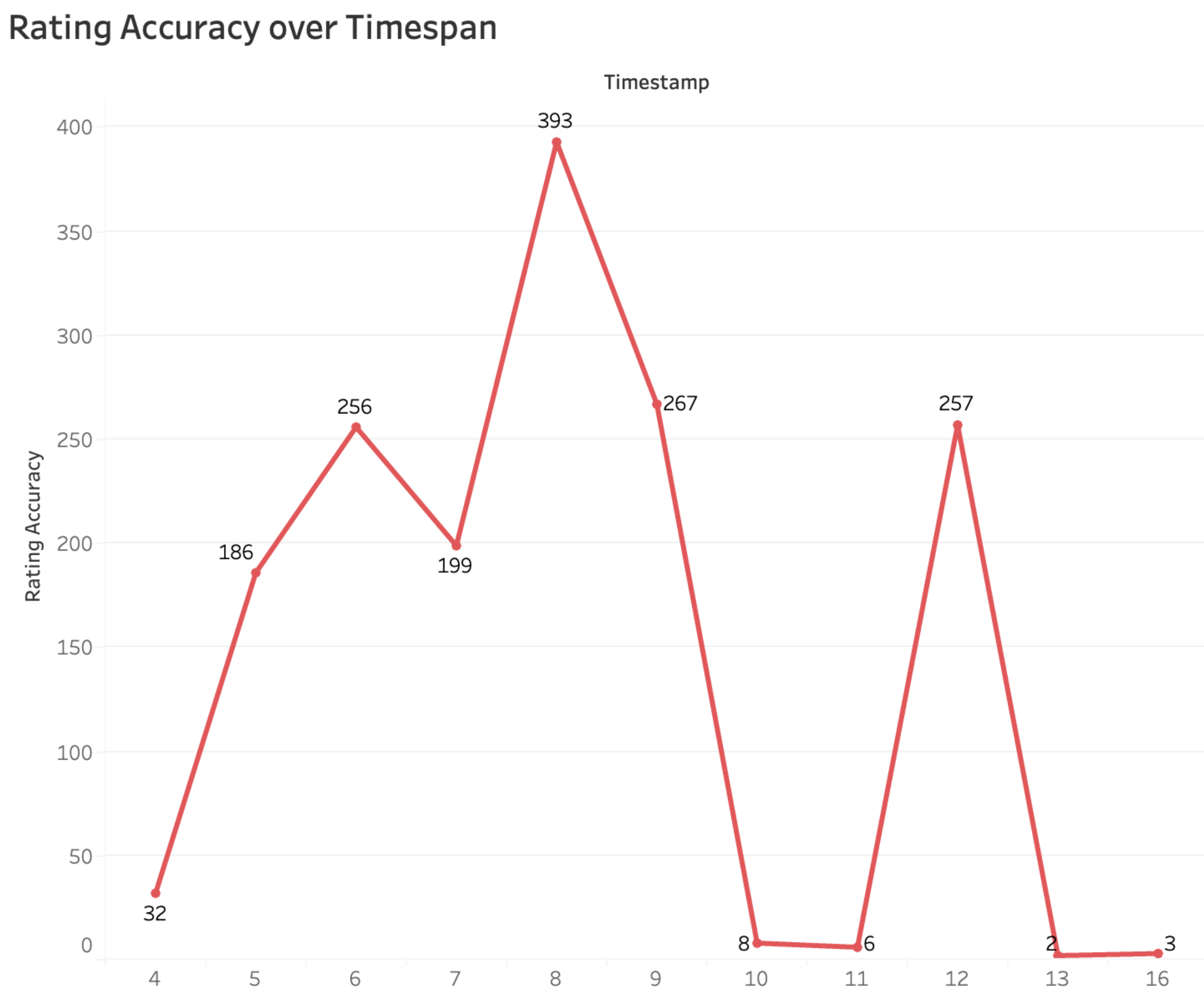

Fig.11. Rating Accuracy vs Timespan

Rating accuracy holds significant importance on e-commerce platforms as it directly influences consumer trust and purchase decisions. When customers rely on product ratings and reviews to make informed buying choices, the accuracy of these ratings greatly impacts their perception of product quality and reliability. In the above plot, it is evident that the rating accuracy follows an approximate normal distribution with highest accuracy presented on the 8th. Post and prior to the said date, we can observe that that accuracy is on a steady decline and incline respectively except on 10th and 11th of the particular month. High rating accuracy ensures that the right product has been recommended to the customer based on their viewing history on the ecommerce website. With an exception on 10th, 11th, 13th and 16th, the rating accuracy has been correct enough to provide right recommendations to the consumers.

Question 3: How does browsing frequency relate to the use of personalized product recommendations?

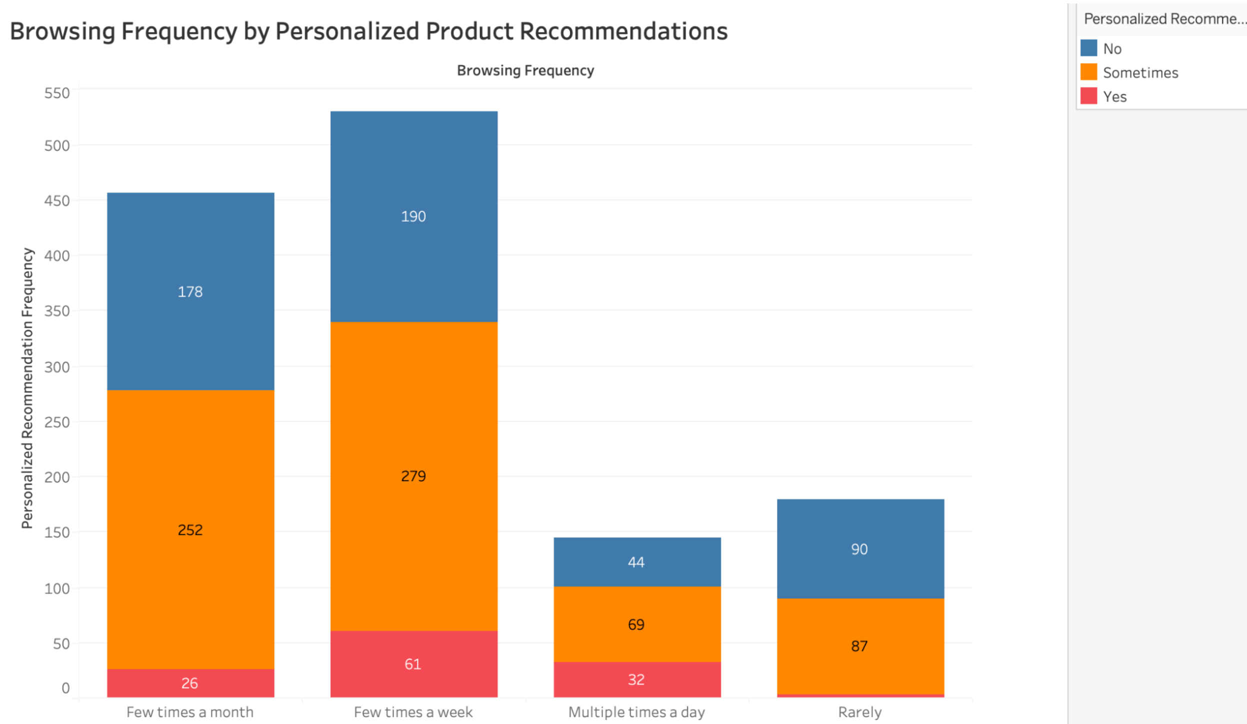

Fig.12. Browsing Frequency by Personalized Product Recommendations

Browsing frequency often correlates with personalized product recommendations on e-commerce platforms. Higher browsing frequency indicates increased engagement, allowing the platform's recommendation algorithms to gather more data points. This data enables the ecommerce company to generate more accurate and personalized recommendations. Users who frequently browse tend to receive more refined and relevant recommendations, as the system adapts and refines suggestions based on their browsing behavior and preferences, creating a more personalized shopping experience. The above plot illustrates that for browsing frequency of “Few times a month” and “Few times a week” the personalized recommendations have been the highest. The personalized recommendations done “Sometimes” have the count and frequency observed in the dataset followed by when the personalized recommendations were not provided i.e. personalized recommendations were “No”. When the browsing frequency is “Multiple times a day” and “Rarely”, we can see that the total count of recommendations were less. This can be an observed error in the personalized recommendation’s algorithm and this error could lead to potential losses in the conversion of a product from “viewing” to “purchase” by the customer. Increasing the personalized recommendations for browsing frequency “Multiple Times a Day” can lead to numerous customers being recommended the products they were looking for, leading to potential increase in new orders.

Question 4: How important are customer reviews to customers with different purchase frequencies?

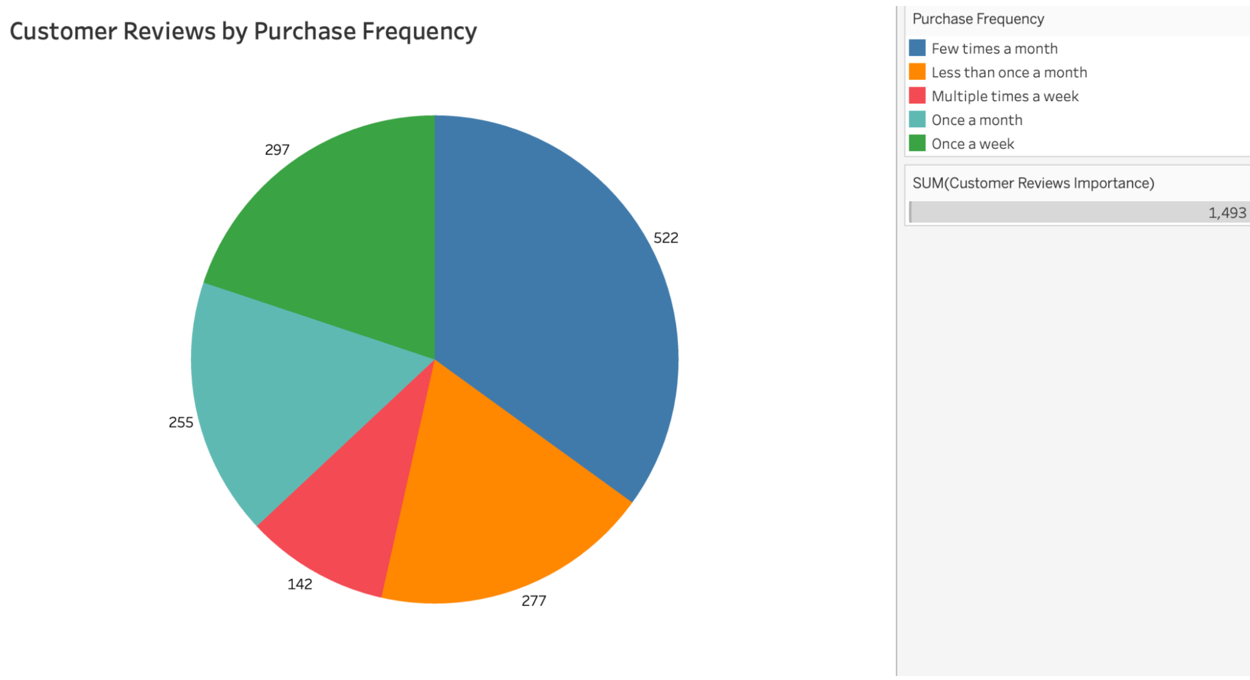

Fig.13. Customer Reviews by Purchase Frequency

Analyzing the above pie chart, it is evident that most customer reviews were observed for customers with a purchasing frequency of “Few times a month”, while the least number of customer reviews were observed for customers with a purchasing frequency of “Multiple Times a Week”. The order of customer reviews by purchasing frequency is “Few Times a month”, “Once a week”, “Less than once a month”, “Once a month”, “Multiple times a week”. From the above data, we can conclude that the customers who shop less are more likely to leave product reviews while on the contrary, the customers that shop more are more likely to not leave reviews.

Question 5: What aspects of Amazon's services are most appreciated by customers with high overall satisfaction?

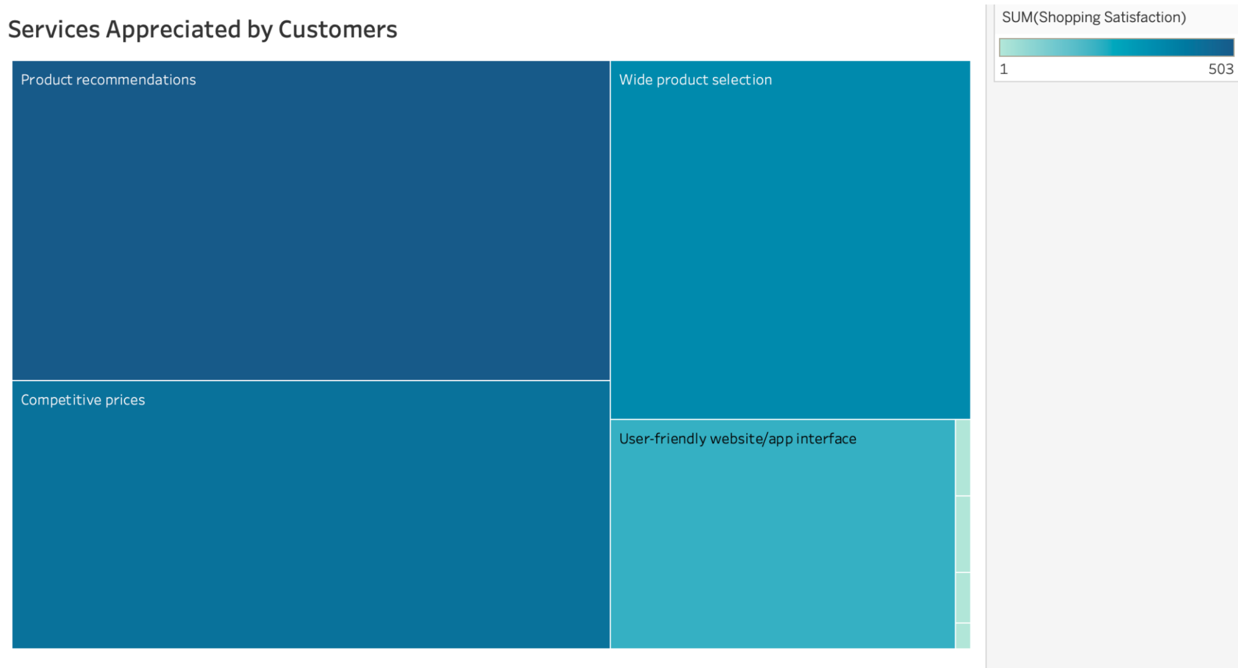

Fig. 14. Services Appreciated by Customers

Among the different services offered by Amazon as an ecommerce website, the services most appreciated by customers are “Product Recommendations”, “Competitive Pricing”, “Wide Product Selection”, “User-Friendly Interface”, etc. As it can be predicted, the customers appreciate the ability to have a wide range of products to choose from while also appreciating the competitive prices offered by the ecommerce website. Surprisingly, these services do not come before “Product Recommendations” service, but after. The user interface of Amazon also plays an important role in guiding customers through the website for searching their desired products. As a result, the customers value this service since it reduces their time in finding their desired products.

Project Timeline:

Timeline	**Task**
11/15/2023	Initial Proposal
11/22/2023	Data Cleaning Perform Data Encoding (Categorical to Numerical) Exploratory Data Analysis
11/29/2023	Generate Visualizations in Tableau

	Review and Make Modifications (If Required)
12/06/2023	Generate Final Draft and Presentation
12/13/2023	Project Due

Conclusions:

This analysis on online shopping has revealed the complexity and diversity of consumer behavior. Improving online purchasing requires an understanding of consumer preferences, search patterns, and decision-making factors. This knowledge enables websites to enhance user experience and develop more intelligent online product advertising strategies. It's obvious that improving the online buying experience for everyone depends on staying up to date with consumer behavior.