BUS 521

COLLABORATIVE ACTIVITY
on
PREDICTING ONLINE CUSTOMER'S BEHAVIOR

By

GROUP-2

Mohammed Abuzar Syed
Rajyalakshmi Mabbu
Bhavana Yadav Krishnappa
Shreya Vishwas
Sohan Shylaja Mohan
Raviteja Reddy Gajjala
Abhishek Devdan Bhore
Sudharsan Elangovan

Under the esteemed Guidance of

Carl Peterson
(Professor)
&
Heather Sperlazzo
(Teaching Assistant)

TABLE OF CONTENT

1. Introduction

The concept of customer behavior analysis and buying is not new. Before the Internet, marketers would observe a customer's behavior to ascertain the goods they would be interested in purchasing.

Customer surveys were used, which were expensive and time-consuming. However, some companies are now able to predict customer behavior by looking at past purchases, search histories, or even social media profiles thanks to the power of data analytics.

1.1 Abstract:

Online shopping customer research often manages customer recognition and purchasing behavior. These studies check who, what, where, when, and how people make purchases. The examination of such customer behavior is helpful in determining the needs and requirements of the buyer for their future goals about the product.

E-commerce businesses can follow the usage and sentiments of their products through this review, and implement effective marketing techniques to provide their customers with a personalized purchasing experience, thereby growing their organizational advantage.

1.2 Problem Statement:

We are addressing the issue of predicting and understanding online shoppers' behavior because it is critical for the success and competitiveness of e-commerce businesses. Here's why we are focusing on this issue:

1. Competitive Advantage
2. Profitability
3. Customer Satisfaction
4. Data-Driven Decision-Making
5. Ethical Considerations

2. Overview

For this, data gathering and analysis have been fueling more and more marketing efforts. From social listening to feedback calls, AI-driven message boards, and more – companies are constantly on the lookout for cues and changes in behavior. If interpreted well, they can bring out key insights to scale campaigns and create content that "clicks" with your audience. Some behaviors that e-commerce brands are increasingly keeping tabs on are:

1. **Nature of the purchase:** When you analyze customer behavior, knowing the nature of the purchase – i.e. if it is habitual (buying often), impulse purchase, variety-driven, promotional, etc. can give you insights into purchasing patterns and how your customers would respond to different marketing campaigns.
2. **Average spending:** Knowing about the spending habits (high-spend, low-spend, etc) of your audience can help you understand their purchasing power and push relevant services/products to them.
3. **Demographics:** This is the most pressed upon factor. e Demographics include gender, age, income, location, and others.
4. **Churn rate** – Also known as customer attrition, it's when a customer discontinues purchasing from a brand. It's one of the easier metrics to calculate, predicting the probability that someone cancels or fails to renew.
5. **Frequency and Recency:** Learn more about how frequently customers engage with your brand and what is the average time interval between their purchases. This can help you target customers with one-time purchases or those with repeated purchases.
6. **Preferred channel of engagement:** Understanding what sources/ mediums/ channels your customers refer to for consuming information is essential to target them with the right message at the right time.

3. Descriptive Analysis:

Descriptive analytics is the process of using current and historical data to identify trends and relationships. It’s sometimes called the simplest form of data analysis because it describes trends and relationships but doesn’t dig deeper.
Descriptive analytics is relatively accessible and likely something your organization uses daily. Basic statistical software, such as Microsoft Excel, or data visualization tools, such as Google Charts and Tableau, can help parse data, identify trends and relationships between variables, and visually display information.
Descriptive analytics is especially useful for communicating change over time and uses trends as a springboard for further analysis to drive decision-making.
The goal of descriptive analytics is to uncover patterns, trends, and characteristics of a dataset by analyzing historical data. We can examine this data in the context of Amazon's behavioral dataset to better comprehend consumer behavior, product preferences, and other pertinent information.
In this project, we aim to look into

1. Data exploration
2. Sentiment Analysis of the Customer

3.1 Data Exploration

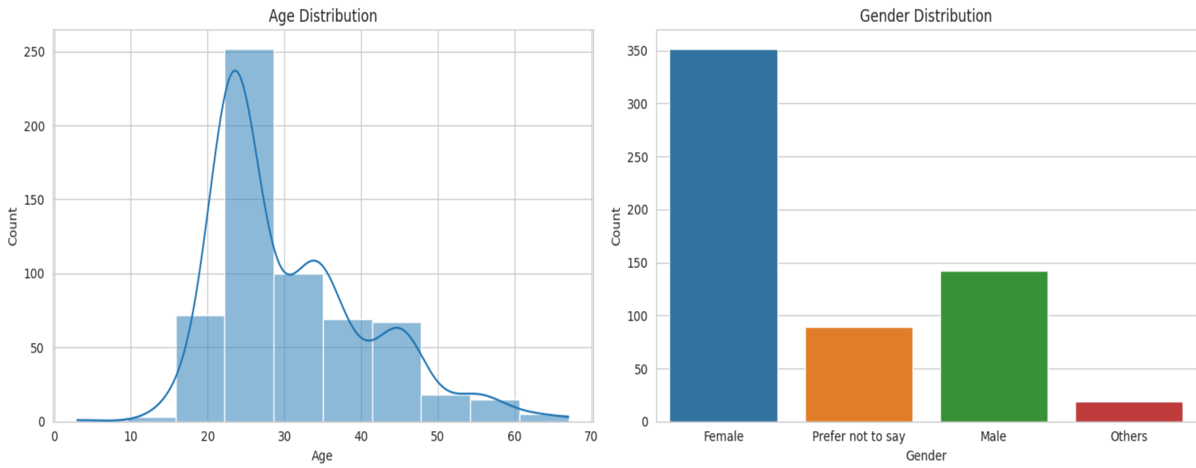
- **Dataset**: The dataset is sourced from kaggle.It consists of attributes like:
- **Timestamp**: The time the survey response was saved.
- **Age**: The age of the respondent.
- **Gender**: gender of the respondent.
- **Purchase_Frequency**: how often the respondent makes a purchase. **Browsing_Frequency**: How often the respondent browses products. **Product_Search_Method**: The method the respondent uses to search for products.
- **Customer_Reviews_Importance**: Importance of customer reviews to the respondent.
- **Shopping_Satisfaction**: The respondent's satisfaction with their shopping experience.
- **Service_Appreciation**: what the respondent appreciates about the service.

- **Improvement_Areas**: Areas where the respondent thinks improvements could be made.

3.2 Exploratory Data Analysis

Exploratory data analysis (EDA) is a critical step in understanding and gaining insight into a dataset, including Amazon's behavioral dataset. EDA involves summarizing the most important features of the data, often using visual methods.
By seeing the points mentioned below, you can understand the behavior patterns of potential new users,

1. **Customer Segmentation using demographic information of respondents (age, gender).**



From the renderings we can see the following:

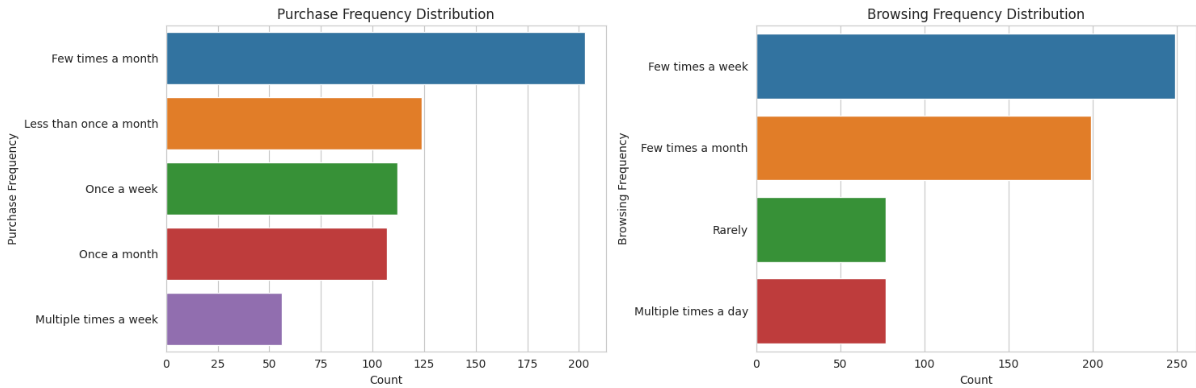
- **Age Range:**

Most of the respondents are in their 20s. The peak is about 23-24 years old.

- **Gender distribution:**

Most respondents refer to themselves as "women". There are fewer respondents who identify themselves as "men" or "don't want to say".

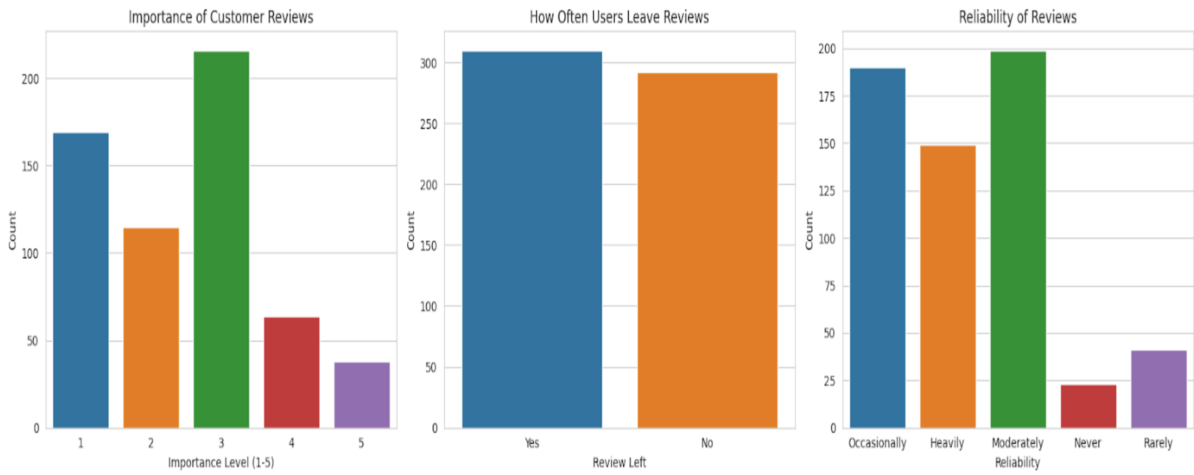
3.3 Behavioural Analysis using Shopping and browsing habits.



We can make the following observations from the visualizations:

a. **Purchase frequency:**
   i. A significant number of respondents in the store "A few times a month."
   ii. The next most common frequency is "Once a month", followed by "Less than once a month".
b. **Browsing frequency:**
   i. Many respondents browse products "a few times a month."
   ii. The next most common browsing frequency is "A few times a week".

3.4 Importance of customer reviews and other features.



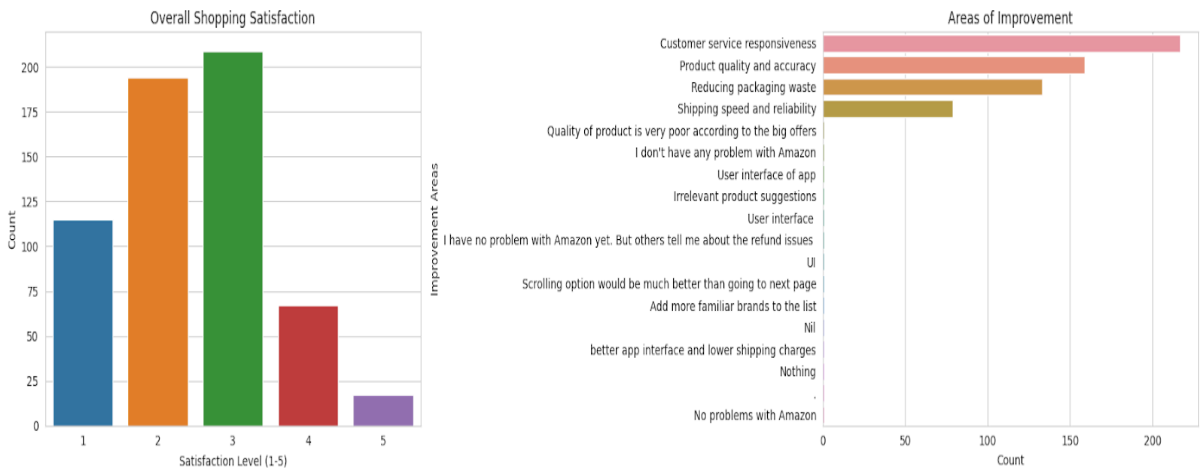
From the visualizations we can conclude the following:

a. **Importance of customer reviews:**
   i. A significant number of respondents rated the importance of customer reviews as a 3 (on a scale of 1-5).
   ii. Also, a large number of respondents give it a rating of 4, indicating that customer reviews play a vital role in their purchasing decisions. How often users leave reviews:

b. **How Often Users Leave Reviews:**
   i. The majority of respondents do not leave reviews.
   ii. A smaller proportion of respondents leave reviews, while very few are unsure.

c. **Credibility of reviews:**
   i. Many respondents find reviews "sometimes" trustworthy.
   ii. The next most common answer is "strongly", indicating that many users trust the reviews they read.

3.5 Levels of satisfaction and areas of development.



From the visualizations we can conclude the following:

a. **Overall satisfaction with the purchase:**
   i. A significant part of the respondents rate their satisfaction with the purchase as 3 (on a scale of 1-5).
   ii. There are also respondents who rate their satisfaction at 4, which indicates a generally positive shopping experience.

b. **Areas of development:**
   i. The most common area that users see in need of improvement is "Product quality and accuracy".
   ii. The next most mentioned area is the reduction of packaging waste.

4. Inferential Analysis:

Most of the time, it is too costly or difficult to get data from the entire population that's important to you, therefore the only data you can obtain are samples. Descriptive analysis is limited to summarizing the features of a sample; inferential analysis uses your sample to generate reasonable predictions about the population as a whole. It is crucial to employ unbiased and random sampling techniques when working with inferential statistics. It is impossible to generalize or draw reliable statistical conclusions from a sample that isn't characteristic of the population.

4.1. Sentiment Analysis of the Customer

Opinion analysis of the "Improvement_Areas" column of Amazon's customer behavior dataset can provide insight into the polarity of customer opinions regarding areas in need of improvement. We can use Natural Language Processing (NLP) techniques and sentiment analysis libraries to do this analysis.

Using the TextBlob library for sentiment analysis. We can calculate an opinion score for each text entry in the "Areas for improvement" column. These scores are usually classified as positive, negative, or neutral.

Code:

```
from text blob import TextBlob

# Function to classify sentiment using TextBlob
def classify_sentiment(text):
```

```
    analysis = TextBlob(text)
    if analysis.sentiment.polarity > 0:
        return 'positive'
    elif analysis.sentiment.polarity == 0:
        return 'neutral'
    else:
        return 'negative'

# Filter out predefined options and missing values
open_text_responses = data['Improvement_Areas'][~data['Improvement_Areas'].isin(['.', 'Reducing packaging waste', 'Product quality and accuracy', 'Shipping speed and reliability', 'Customer service responsiveness', 'Nothing', 'Nil'])]

# Apply sentiment analysis
sentiments = open_text_responses.apply(classify_sentiment)
sentiments.value_counts()
```

We have the following results based on sentiment analysis using TextBlob's plaintext responses in the Areas for Improvement column.

1. Neutral emotions: 6 answers
1. Positive emotions: 3 answers
1. Negative emotions: 2 answers

4.2. Sentiment Classification:

- Neutral Sentiments: There are 6 responses that were classified as neutral. These responses typically do not express strong positive or negative emotions and may contain constructive feedback or general suggestions without a strong emotional tone.
- Positive Sentiments: There are 3 responses classified as positive. Positive sentiments suggest that users appreciate certain aspects and have provided constructive suggestions for further improvement. This indicates a favorable tone in their feedback.

- Negative Sentiments: There are 2 responses classified as negative. Negative sentiments indicate that users expressed dissatisfaction or specific issues that they would like to see addressed. These responses highlight areas of concern or problems.

  Actionable Insights:
- The majority of the feedback in the "Improvement_Areas" column is neutral, which provides insights into areas where users believe there could be improvements. Analyzing these neutral responses can help identify common themes and trends in user suggestions.
- Positive feedback can be used to reinforce and build upon aspects that users appreciate. Acknowledging and implementing these suggestions can further enhance user satisfaction.
- Negative feedback should be closely examined and addressed to improve user satisfaction. Identifying the specific issues mentioned in negative responses is crucial for making necessary improvements.

Most of the feedback in the Improvement_Areas column appears to be neutral, with some positive and negative sentiments. This suggests that while some users may have specific areas they would like to see improved, others may provide more general suggestions or feedback without a strong emotional tone.

7. Conclusion:

In the e-commerce landscape, the customer is the king. So, there’s a constant battle between brands to not just catch the eye, but retain themselves in customers’ minds too. And the only way to do it is to work in their best interest- understanding the “why” of every purchase and bringing the “when” and “who” in your favor.
Thanks to behavioral analytics tools, marketers can now track, record, and influence customer behaviors, all to improve user experiences. But how successful they get depends on the data they capture, as it should be contextual and structured. Furthermore, choosing a tool custom-made for e-commerce business would be a smart choice. Tools offer real-time insights that can help you gain an edge when it comes to optimizing strategies for ever-changing brand-customer dynamics!

In conclusion, the analysis of customer behavior and sentiment within the context of Amazon's dataset provides valuable insights into the factors that influence online shopping and satisfaction. The dataset allows us to explore customer demographics, shopping patterns, and preferences, shedding light on critical aspects of e-commerce businesses. Notably, age and gender distributions reveal the majority of respondents are in their 20s, with a higher representation of women. Purchase and browsing frequencies suggest that many customers engage with online shopping frequently, with varying degrees of reliance on customer reviews. Overall, customer satisfaction ratings indicate a generally positive shopping experience, while areas for improvement center around product quality and the reduction of packaging waste. Additionally, sentiment analysis of open-text feedback shows that users express a range of sentiments, including neutral, positive, and negative, highlighting the importance of addressing specific issues while continuing to enhance the overall online shopping experience. This analysis serves as a valuable resource for e-commerce businesses aiming to meet customer needs, gain a competitive edge, and make data-driven decisions for sustainable growth.

8. Reference:

1. They demonstrated that the mined opinion has a solid insightful incentive for coming business sector bearings. Sentiment Analysis was utilized to estimate the end record of Tata Services and an exactness of 85.99% was found simultaneously. Sentiment investigation is frequently used to construct a social conduct diagram on a human's web-based conduct to observe the relationship between exchange and volume costs of stocks (Jiajia Li). Sentimental Analysis was additionally performed on the information removed from Senti-WordNet utilizing a crossbreed choice model to show what market patterns mean for an item prominence and rate (MICHAEL M. TADESSE, 2020; A.Ashraf, 2021; W.H.Bangyal, 2021;)[24-25].

2. (I. M. W, Dr Rizwana Bashir) The way of behaving consumers shopping over the web has been broken down into five variables. These variables are time, protection, trust, comfort, and item assortment. The exploration led for this design was a survey and the outcomes were examined measurably. Trust was asserted as such a human quality that influences their purchasing propensities (Martin, 2020).

3. The shopping conduct of individuals in Pakistan gets impacted by their mental and passionate mentalities. Protection is additionally considered as the most noticeable variable in this online shopping pattern in the way that individuals may now and then not have a real sense of reassurance sharing their data over the web. Then again, entrancing costs of different stuff may likewise draw in individual consideration and help to ask online shopping. Thus, trust over the source is the issue that influences individuals purchasing ways of behaving (H. u. R. I. J. Sajjad Nazi) (Kapoor, 2020; W.H.Bangyal, 2021; A.Ashraf, 2021; W.H.Bangyal, 2022)[22,23,26]
4. https://www.kaggle.com/datasets/cynthiarempel/amazon-us-customer-reviews-dataset/data
5. https://www.kaggle.com/code/swathiunnikrishnan/consumer-behaviour-analysis-of-amazon-a-study/input?select=Amazon+Customer+Behavior+Survey.csv