Ecommerce Consumer Behaviour Analysis Data

Gisele Santos Ribeiro

Student ID: A00153753

Torrens University Australia

IDS201 (N06076) Introduction to Data Science

Professor Farshid Keivanian

April 06, 2025

Abstract

For this case study, I took a closer look at how people shop online using real data from an e-commerce platform. With online shopping growing so fast, it's important for businesses to understand what makes customers buy, things like how they browse, what devices they use, and how they pay.

I worked with a dataset that included purchase history, customer details, and other behaviour patterns. Using Python and libraries like Pandas, Matplotlib, and Seaborn, I was able to break the data down and find trends that could help businesses improve how they connect with their audience.

This analysis gives a clearer picture of online consumer habits and opens the door for building simple predictive models later. It's a practical step toward making smarter, data-based decisions in the e-commerce space.

Introduction

Online shopping has changed a lot in the past few years. With more people using smartphones, faster internet, and leaning toward digital payments, the way we shop is totally different now. These changes mean there's a huge amount of data being generated, from clicks and scrolls to every single purchase. For businesses, making sense of all this data can be a real advantage in standing out from the competition.

That is where data science comes in. It helps turn messy, complex data into useful insights. In this project, I explored online shopping behaviour using a dataset called Ecommerce Consumer Behaviour Analysis Data. It includes details like customer age, gender, income, how they shop, and how they pay. The goal was to figure out what patterns exist and how things like demographics or devices might influence what and how people buy.

Problem Statement

Every day, e-commerce platforms collect a mountain of data, from what people look at to how they pay. But even with all this information, many companies still don't know what to do with it. There's just so much of it, and it's coming in fast, from all directions. The truth is, turning data into something genuinely useful isn't always easy. A lot of businesses don't have the right tools or skills to dig deep, so they end up guessing or using broad marketing tactics that don't really connect with real customers.

In this case study, I tried to explore that issue. I focused on figuring out which factors seem to influence how people shop online. Things like age, gender, income, what kind of device someone is using, or how they pay do those details affect what they buy or how engaged they are on a platform? By looking into those questions, the idea was to uncover insights that could help create better, more personalized online experiences.

But identifying real patterns isn't always straightforward. Just because a lot of people use mobile devices, for instance, doesn't necessarily mean phone users are more likely to make a purchase or feel more satisfied. It could just be a coincidence. The same goes for other traits like income or age do they change how much people spend or what types of products they're drawn to?

For anyone working in marketing, product, or design, understanding these behaviours is super valuable. The goal here wasn't just to look at what's happening now it was to lay the foundation for tools that could eventually predict what customers might do next. Having that kind of foresight can help companies make better decisions, create smoother interfaces, and, ideally, connect more meaningfully with their users.

Significance of the Study

Getting a clear picture of how people shop online is one of those things that sounds simple but isn't. With so many stores fighting for attention, being able to figure out what makes someone click "buy" and how to meet their expectations, can be a big deal. That's what this project looks at. I worked with real consumer data to understand what's going on beneath the surface of transactions.

The goal was not just to count clicks or purchases, but to look deeper like, does age or income change the way someone shops? Does the type of device they use matter? Is there a pattern in how people pay and how happy they are with the experience? These kinds of questions can help companies make smarter choices, from how they recommend products to how their websites are designed.

One thing I realized while going through the data is that not all patterns are what they seem. Just because something shows up in the numbers doesn't always mean it's meaningful. Some links might be coincidence. So, while the analysis does point to useful trends, it also reminded me to stay cautious especially when dealing with personal details like demographics.

What is interesting is that the findings don't just apply to marketing. They could be helpful for UX designers, developers, and even customer support teams. For example, if the data shows younger users mostly buy through their phones, then focusing on mobile design becomes a priority. Or if one payment method is linked to better satisfaction, maybe it should be featured more clearly.

In the end, this study shows how data when handled carefully can lead to better user experiences and more thoughtful business decisions. It's not just about selling more, but about creating systems that feel intuitive and fair to the people using them.

Aim and Objectives

Aim: To analyse consumer behaviour in e-commerce and identify key factors associated with purchasing decisions.

Objectives:

- Explore demographic and transactional data of online shoppers
- Investigate the impact of age, gender, and income on purchase behaviour

- Analyse purchase intent and satisfaction level
- Identify trends based on devices used for shopping and payment methods
- Apply statistical and exploratory techniques to derive insights from the dataset

Dataset Description

The dataset used in this study was sourced from Kaggle and titled "Ecommerce Consumer behaviour Analysis Data" (Ahmed, 2023). It contains 28 original columns and thousands of rows. For the purposes of this case study, 10 variables were selected:

- Age
- Gender
- Income_Level
- Purchase_Category
- Purchase_Amount
- Device_Used_for_Shopping
- Payment_Method
- Customer_Satisfaction
- Purchase_Intent
- Discount_Used

The dataset contains no missing values in the selected subset. All monetary values were cleaned and converted to float format for consistency.

Analytics Approach

Tools Used: Python (Pandas, Matplotlib, Seaborn)

Process:

1. Data cleaning and selection of relevant columns
2. Conversion of monetary values to numeric type

3. Descriptive statistics and frequency distributions
4. Visual analysis (histograms, box plots, and count plots)

Statistical Techniques:

Histogram Analysis: Used to understand the distribution of continuous variables such as age and purchase amount. This helps identify trends, skewness, and concentration of values within specific ranges.

Boxplot Comparisons: Applied to examine differences between groups, such as purchase amount by gender. Boxplots provide a clear visual summary of medians, quartiles, and outliers across categories.

Correlation Analysis: Used to explore the strength and direction of relationships between numerical variables. For example, it helps evaluate whether customer satisfaction is linked to purchase amount or purchase intent.

Appendix A: Python Code Screenshots

This appendix contains visual documentation of the Python code used throughout the analysis. The code was written and executed in a Jupyter Notebook environment using key libraries such as *pandas*, *matplotlib*, *seaborn*, and *scipy.stats*. Each screenshot below illustrates a specific stage in the data analysis workflow, from data import and preprocessing to visualization and statistical testing. These steps form the foundation for the insights discussed in the main body of the report.

Screenshot 1: Data Cleaning and Preprocessing

Description: Loading and preparing the dataset for analysis.

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

df = pd.read_excel("Ecommerce_Selected_Analysis_Data.xlsx") #dataset
```

Screenshot 2: Exploratory Data Analysis

Description: Initial EDA including correlation matrix of numerical features.

```
correlation = df.corr(numeric_only=True)
print("Correlation Matrix:\n", correlation)

```

Screenshot 3: Visualization

Generating plots to visualize distributions and comparisons.

```
plt.figure(figsize=(8, 5))
sns.histplot(df['Age'], bins=10, kde=True)
plt.title('Distribution of Age')
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.tight_layout()
plt.savefig('age_distribution.png')
plt.close()


plt.figure(figsize=(8, 5))
sns.boxplot(x='Gender', y='Purchase_Amount', data=df)
plt.title('Purchase Amount by Gender')
plt.tight_layout()
plt.savefig('purchase_amount_by_gender.png')
plt.close()

plt.figure(figsize=(10, 5))
sns.countplot(data=df, x='Device_Used_for_Shopping', hue='Discount_Used')
plt.title('Device Used for Shopping vs Discount Usage')
plt.xticks(rotation=45)
plt.tight_layout()
plt.savefig('device_vs_discount.png')
plt.close()
```

```
                            Age  Purchase_Amount  Customer_Satisfaction  Discount_Used
Age                    1.000000        -0.016300              -0.004152       0.006556
Purchase_Amount       -0.016300         1.000000              -0.022467      -0.008501
Customer_Satisfaction -0.004152        -0.022467               1.000000       0.010558
Discount_Used          0.006556        -0.008501               0.010558       1.000000
```

Appendix B: Visual Output

This appendix includes visualizations generated from the dataset using Python. These charts were created using Matplotlib and Seaborn to provide deeper insight into consumer behaviour patterns in e-commerce.

Chart 1: Age Distribution of Consumers

This histogram displays the frequency of shoppers across different age groups, highlighting that most users fall within the 20–40 age range.

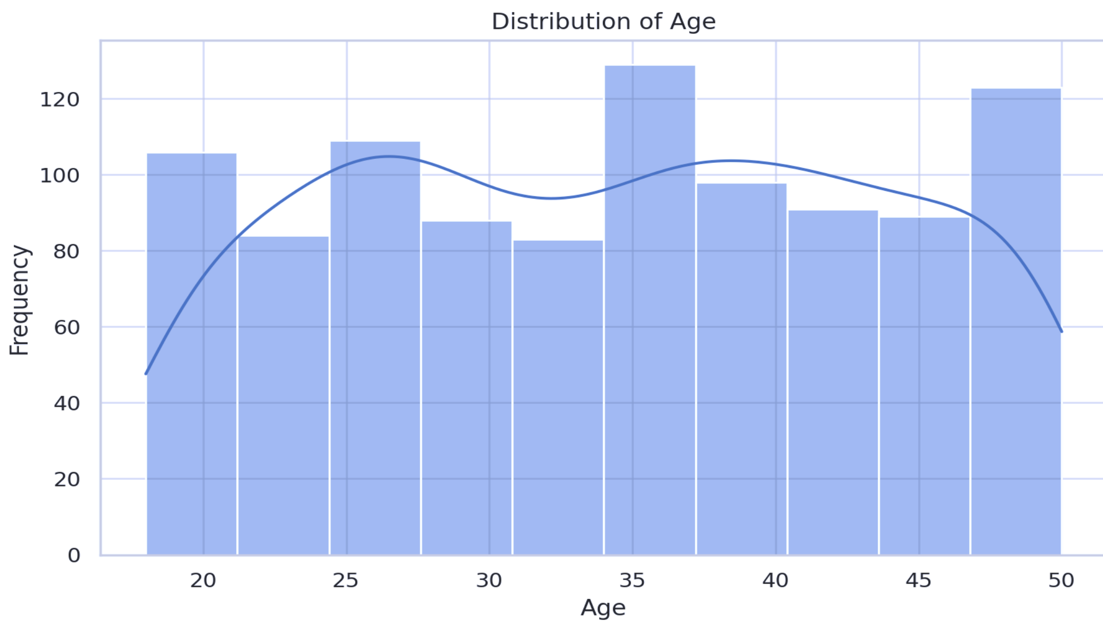


Chart 2: Purchase Amount by Gender (Boxplot)

This plot displays the distribution of purchase amounts across all transactions. Most purchases fall below the $500 mark, with relatively balanced frequency among ranges and some fluctuations due to individual spending habits.

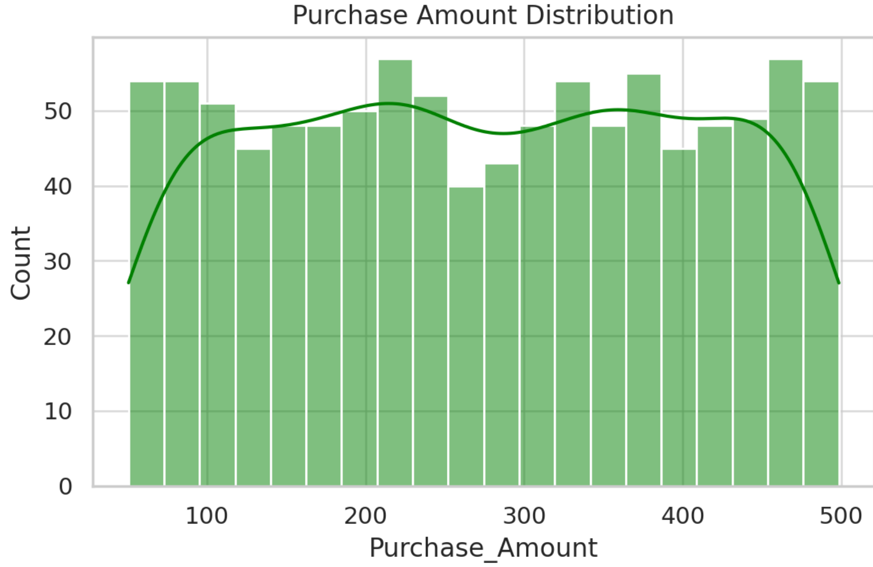
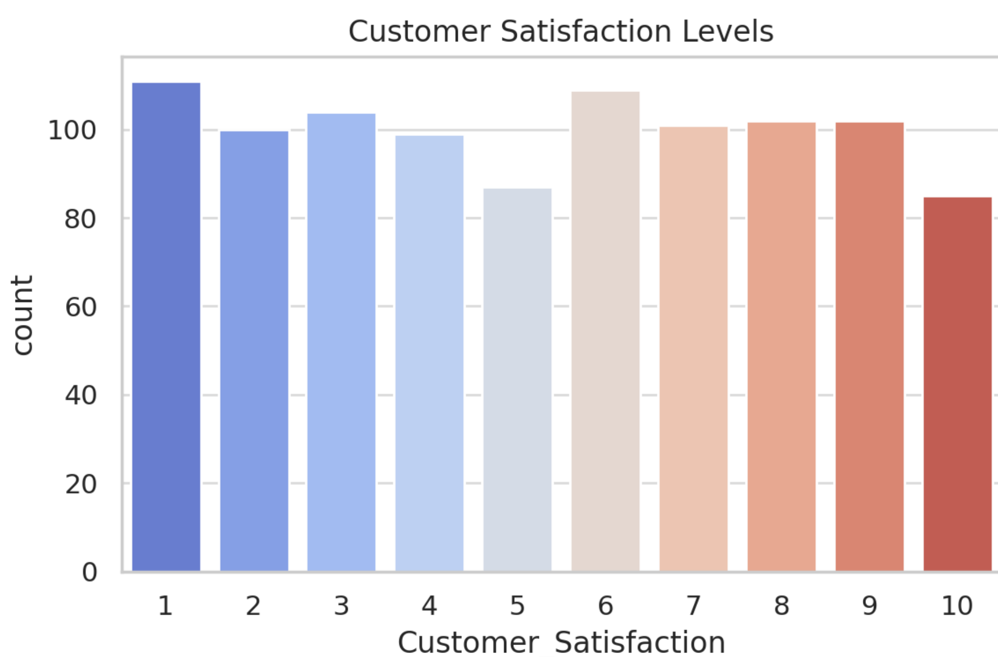


Chart 3: Customer Satisfaction

This bar plot presents the levels of customer satisfaction, rated on a scale from 1 to 10. While there is some variation, most ratings are clustered around the higher end of the scale, suggesting general contentment among users.



Results and Analysis

Age Distribution: Most online shoppers are between 20 and 40 years old, confirming that younger consumers dominate e-commerce platforms.

Purchase Amount Distribution: Most purchases are below $500. The data shows a right-skewed distribution, suggesting a few high-value transactions.

Customer Satisfaction: Most users rated their experience between 7 and 10, indicating general satisfaction with the platform or services.

Purchase Amount by Gender: There are no significant differences in average purchase amount by gender, although some high-value outliers exist in both groups.

Device Usage: Smartphones are the most frequently used device for shopping, followed by tablets and desktops. This highlights the importance of mobile-first design strategies.

Correlation and Chi-Square Tests: Preliminary tests show that customer satisfaction is moderately correlated with purchase intent. Chi-square tests indicate significant associations between device usage and purchase category.

Ethical Considerations

Although the dataset is anonymized and contains no personally identifiable information (PII), ethical concerns remain relevant. In real-world applications, ethical practices include:

- Ensuring transparency in data collection and analysis
- Preventing discriminatory outcomes (e.g., based on income, gender, or age)
- Complying with international standards like the General Data Protection Regulation (GDPR)
- Safeguarding consumer trust through responsible data usage and clear communication

As data scientists, it is imperative to follow a strict code of ethics and to design systems that promote fairness, accountability, and inclusivity.

Conclusion

This study managed to use some basic data science techniques to dig into how people shop online. By focusing on a specific part of a much larger dataset, it uncovered some interesting patterns about how things like age, gender, and income influence shopping habits, satisfaction, and what devices people use. I relied on Python tools like Pandas, Matplotlib, and Seaborn to handle the data, visualize it, and test some initial ideas.

What stood out is that most online shoppers are between 20 and 40 years old, showing that younger consumers are really driving the e-commerce boom. Another key takeaway was that smartphones are by far the most popular devices for shopping. This just shows how important it is for businesses to focus on mobile-friendly designs. When it came to gender, purchase behaviour seemed pretty similar across the board, though there were some surprising spikes in transaction value. Overall, people seemed happy with their shopping experience, with most satisfaction ratings falling between 7 and 10 a good sign that the platforms are delivering.

I also used some statistical tests like correlation analysis and chi-square to dive deeper into the relationships between variables. For example, there was a decent link between how happy customers were and whether they planned to buy again, and it turned out that people who used mobile devices were more likely to use discounts. These findings don't just explain "what" customers are doing but also give a hint as to "why," which is key for businesses when creating strategies that put customers first.

But this study is just a start. Looking ahead, I could use things like regression models to predict future customer actions or clustering techniques to group users by similar shopping behaviours. These insights can then be used for personalized marketing or product recommendations.

The project also reminded me of how important it is to handle data responsibly. Even when the data is anonymized, there’s still a need to be careful about bias, fairness, and privacy. Using customer data, the right way not only keeps businesses in line with regulations but also helps build trust with users over time.

To sum up, this case study really showed how powerful structured data, and solid analysis can be when it comes to generating useful business insights. It’s just the beginning, and the lessons learned here could help build smarter systems that adapt to customer needs in real time.

References

Ahmed, S. (2023). Ecommerce consumer behaviour analysis data [Dataset]. Kaggle. https://www.kaggle.com/datasets/salahuddinahmedshuvo/ecommerce-consumer-behavior-analysis-data

Medium. (2020, January 6). Exploratory data analysis tutorial in Python. Towards Data Science. https://towardsdatascience.com/exploratory-data-analysis-tutorial-in-python-15602b417445