

Analysis of Online Shopping Patterns and Consumer Behavior in E-commerce

Tesfamariam Tsegay. Ghezehey

Northwest Missouri State University, Maryville MO 64468, USA
s568695@nwmissouri.edu

Abstract. The growth of e-commerce has led to vast consumer data, offering opportunities to enhance customer experience and retention. This project, "Analysis of Online Shopping Patterns and Consumer Behavior in E-commerce," explores key purchasing trends, product preferences, and churn factors using a public dataset. [To be completed at the end]

Keywords: e-commerce · consumer behavior · retail analytics · shopping patterns · data mining

1 Introduction

1.1 Domain and Why?

E-commerce is a rapidly growing industry that generates vast amounts of data from customer interactions. Understanding consumer behavior helps businesses optimize marketing strategies, improve customer experience, and reduce churn.

1.2 Data Sources

Primary Dataset: - E-commerce Customer Behavior (Kaggle) - Contains customer demographics, purchase frequency and churn data

Added Dataset: - E-commerce Order and Sales (Kaggle) - Includes order details, customer location (State Code), product categories, brand, cost and sales revenue

1.3 Data Problem and Importance

Problem Statement: This project aims to analyze customer shopping patterns, churn behavior, and sales performance. By integrating order-level data, we can **identify key factors that influence customer retention, purchasing decisions, and revenue generation.**

Why is this important?

- Businesses can predict customer churn and improve engagement strategies.
- Helps in analyzing which **products and brands** drive the most revenue.
- Optimizes marketing efforts through customer segmentation and sales trend analysis.
- Helps in pricing strategies by analyzing ****cost vs. sales profitability****.

1.4 steps

Planning - Implementation Plan

- **Data Collection:** Import customer behavior and order details datasets from Kaggle.
- **Data Cleaning and Preprocessing:** Handle missing values, normalize category names, and merge datasets.
- **Exploratory Data Analysis (EDA):**
 - Analyze customer demographics and purchase behavior.
 - Identify the best-selling products, brands, and revenue trends.
 - Detect patterns in customer retention and churn.
- **Sales and Churn Prediction:** Apply machine learning models to identify high-value customers and predict churn risks.
- **Results and Insights:** Summarize findings and provide recommendations for customer retention and sales improvement.

1.5 key component and Limitations

- **key component:**
- **Customer Segmentation:** Identifying high-value customers.
- **Churn Analysis:** Analyzing why customers leave.
- **Sales and Product Analysis:** Understanding purchasing trends.
- **Limitations:**
- **Data Bias:** Dataset might not represent all global e-commerce businesses.
- **Feature Availability:** Limited variables may affect churn prediction accuracy.
- **Assumptions:** Missing customer feedback in the dataset.

2 Methodology

2.1 Data Collection

The datasets used in this project, were sourced from **Kaggle**, Specifically, two datasets were utilized:

- **E-commerce Customer Behavior Dataset** [1] - Contains customer demographics, purchase patterns, and churn indicators.
- **Online E-commerce Orders Dataset** [2] - Includes order details, product categories, brands, and sales revenue.

Both datasets were obtained in **CSV format**. No additional web scraping was required, as the data was readily available for download.

To prepare the datasets for analysis, I focused on the following major attributes:

- **Customer Demographics:** Age, location (state code), and anonymized customer name.

- **Purchase Behavior:** Order number, order date, status (completed, pending, canceled).
- **Product Information:** Product name, category, brand, cost, and total sales.
- **Churn Prediction Indicators:** Purchase frequency, average order value, and payment method.

The datasets were merged using **Customer ID** and **Order Number** as primary keys to create a unified dataset. Data preprocessing and cleaning were performed separately (in the upcoming Data Cleaning section) to ensure consistency and improve the quality of analysis.

Titles of Other Potential Sections

2.2 Data Cleaning

2.3 Feature Engineering

2.4 Exploratory Data Analysis (EDA)

Customer Demographics

Product Trends

Churn Insights

3 Results and Analysis

4 Limitations

5 Conclusion and Future Work

6 References

— Bibliography —

Two data sets were used in this study [1,?,?].

References

1. Kaggle: E-commerce customer behavior analysis dataset (2025), <https://www.kaggle.com/datasets/shriyashjagtap/e-commerce-customer-for-behavior-analysis/data>
2. Kaggle: Online e-commerce orders dataset (2025), <https://www.kaggle.com/datasets/ayushparwal2026/online-ecommerce>