UNIVERSITY OF ECONOMICS AND LAW
FACULTY OF INFORMATION SYSTEMS

FINAL PROJECT REPORT
INTERDISCIPLINARY RESEARCH METHOD COURSE

TOPIC: USER BEHAVIOR ANALYSIS UNDER THE INFLUENCE OF TIKTOK'S RECOMMENDATION ALGORITHM

Group 5:	**Nguyen Chi Duc**
	Nguyen Thi Bao Tran
	Nguyen Hoang Khanh Nhu
	Huynh Ngoc Phuong Nghi
Instructor:	**Assoc. Prof. Ho Trung Thanh, Ph.D.**

Ho Chi Minh City, March 2025

Members of Group 5

No.	*Full name*	*Student ID*	*Point / 10 (Individual contribution)*	*Signature*
1	Nguyễn Chí Đức	K234111429		
2	Nguyễn Thị Bảo Trân	K234111455		
3	Nguyễn Hoàng Khánh Như	K234111412		
4	Huỳnh Ngọc Phương Nghi	K234111403		

Acknowledgements

The research team would like to express our deepest gratitude to Assoc. Prof. Ho Trung Thanh, Ph.D., for his invaluable guidance, support, and encouragement throughout the implementation of this project. His expertise and constructive feedback have been instrumental in shaping the direction and quality of this study.

We also extend our sincere thanks to the Faculty of Information Systems, University of Economics and Law, for providing us with the resources and academic environment necessary to conduct this research.

Additionally, we are grateful to all survey participants and contributors who generously shared their insights and experiences, enabling us to gather meaningful data for analysis. Without their cooperation, this study would not have been possible.

We would like to extend our sincere and heartfelt thanks to all the scholars, authors, and research communities whose work has served as a vital source of inspiration and guidance throughout this project. Their dedication to exploring and expanding knowledge in the fields of technology, social media, and human behavior has not only enriched our understanding but also shaped the foundation upon which this study was built.

This project would not have been possible without the intellectual contributions and thoughtful perspectives shared through their publications. We are truly grateful for the opportunity to learn from their work, and we hold deep respect for their ongoing efforts to advance research that impacts both academia and real-world understanding.

Finally, we would like to thank our families and friends for their unwavering support and understanding during the course of this research. Their encouragement has been a source of strength and motivation for the entire team.

Group 5

Commitment

We, the members of Group 5, hereby declare that this research project, titled "User Behavior Analysis Under the Influence of TikTok's Recommendation Algorithm," is the result of our independent work, conducted under the supervision of Assoc. Prof. Ho Trung Thanh, Ph.D..

The study was carried out with academic integrity and in compliance with ethical research standards. All data, analyses, and findings presented in this report are original and derived from legitimate sources, which have been duly cited in the references section.

We affirm that this project was completed without any unauthorized assistance or plagiarism. Any external contributions have been explicitly acknowledged in the relevant sections of the report.

Ho Chi Minh City, March 2025

Group 5

Table of Contents

List of Tables

List of Figures

List of Fomulas

Gantt Chart

GROUP 5 | ACTION PLAN

HẠNG MỤC	TASK	DETAILS	PIC	STATUS	DEMO 1	DEMO 2	FINAL	NOTE
PROPOSAL (MID-TERM)	B1	Tổng quan tình hình trong và ngoài nước của TikTok và thuật toán gợi ý tìm kiếm	Trân	Final	10/01	12/01	26/01	
		Đánh giá sự khác biệt của NCKH trong và ngoài nước	Trân	Final	10/01	12/01	26/01	
		Thực trạng	Như	Final	10/01	12/01	26/01	
		Những vấn đề được giải quyết và những vấn đề tồn đọng	Như	Final	10/01	12/01	26/01	
	B2	Ý tưởng khoa học	Đức + Nghi	Final	10/01	12/01	26/01	
		Tính cấp thiết			11/01	12/01	26/01	
		Tính mới			12/01	13/01	26/01	
	B4	Tổng hợp tài liệu tham khảo	Trân	Final		26/1		trình bày APA
	B5	Mục tiêu + Câu hỏi nghiên cứu	Trân	Final	12/01	13/01	26/01	
		Đối tượng nghiên cứu	Như + Nghi	Final	12/01	13/01	26/01	
		Phương pháp	Đức	Final	12/01	13/01	26/01	
	Hình thức	Chỉnh hình thức tổng kết bài	All	Final		14/1		
FINAL-TERM	CHAP 1	Research Overview	Trân	Final	16/02	17/02	18/02	
		Fill nội dung phù hợp trong chap	Trân + Đức	Final	15/02	16/02	18/02	
	CHAP 2	Phân tích Dataset	Như + Đức	Final		18/02		
		Fill nội dung phù hợp trong chap	Trân + Nghi	Final	16/02	17/02	18/02	
	CHAP 3	Fill nội dung phù hợp trong chap	Như + Đức	Final	01/03	02/03	11/03	
	CHAP 4	Fill nội dung phù hợp trong chap	Như + Đức	Final	12/03	13/03	18/03	
	Source code		Đức	Final	20/02	21/02	18/03	
	PPT		Nghi	Final	14/03	15/03	18/03	
	Phần ngoài	Abstract, lời cam kết	Nghi	Final		18/02		
	Hình thức	Chỉnh hình thức tổng kết bài	All	Final		18/02		
EXCEL	Thu thập	Research dataset và chọn loại phù hợp	All	Final	16/02	17/02	18/02	
	Phân tích	Trình bày và giải nghĩa	All	Final	16/02	17/02	18/02	

Abstract

The rapid advancement of artificial intelligence (AI) and recommendation algorithms has reshaped how users interact with social media platforms. TikTok, a globally popular short-video platform, exemplifies this transformation through its ability to personalize content based on user behavior and preferences. However, the impact of this algorithm on user behavior, particularly in Vietnam, has not been thoroughly and comprehensively studied. This study seeks to answer the following questions: How does TikTok's recommendation algorithm influence user behavior? Specifically, what effects does it have on purchasing decisions? Additionally, the study examines ethical concerns related to algorithmic transparency and data privacy. TikTok is not just an entertainment platform but also a commercial and cultural tool with significant influence. However, excessive content personalization and the platform's addictive design may lead to negative consequences, including impulsive consumption, changes in social habits, and mental health issues. This research aims to better understand these impacts and propose solutions to balance commercial benefits with social responsibility. The study employs a mixed-method approach, including user surveys, behavioral data analysis, and machine learning techniques, to identify patterns in user engagement and decision-making. Data was collected and analyzed to clarify how TikTok's algorithm affects user behavior and psychological states. The findings reveal that TikTok's recommendation algorithm not only increases user engagement but also fosters consumption through personalized advertisements. The platform's addictive design creates a "ludic loop" that makes it difficult for users to disengage. Additionally, the algorithm shapes social norms by amplifying viral trends and challenges, influencing both individual preferences and societal behaviors. This research provides valuable insights into the influence of TikTok's algorithm on user behavior and contributes to broader discussions about ethical AI applications. The findings can support policymakers, businesses, and researchers in developing strategies to manage social media platforms effectively, ensuring a balance between user experience, commercial objectives, and social responsibility.

Tóm tắt đề tài

Sự phát triển mạnh mẽ của trí tuệ nhân tạo (AI) và các thuật toán gợi ý đã định hình cách người dùng tương tác với các nền tảng mạng xã hội. TikTok, một nền tảng video ngắn phổ biến toàn cầu, là một ví dụ điển hình với khả năng cá nhân hóa nội dung dựa trên hành vi và sở thích của người dùng. Tuy nhiên, sự ảnh hưởng của thuật toán này lên hành vi người dùng, đặc biệt tại Việt Nam, vẫn chưa được nghiên cứu đầy đủ và toàn diện. Nghiên cứu này tập trung vào việc trả lời các câu hỏi: Thuật toán gợi ý của TikTok ảnh hưởng như thế nào đến hành vi người dùng? Cụ thể, nó có tác động gì đến quyết định mua sắm của họ? Đồng thời, nghiên cứu xem xét các vấn đề đạo đức liên quan đến tính minh bạch của thuật toán và quyền riêng tư dữ liệu. TikTok không chỉ là một nền tảng giải trí mà còn là một công cụ thương mại và văn hóa có sức ảnh hưởng lớn. Tuy nhiên, sự cá nhân hóa nội dung quá mức và thiết kế gây nghiện của nền tảng này có thể dẫn đến những tác động tiêu cực, bao gồm hành vi tiêu dùng bốc đồng, thay đổi thói quen xã hội và các vấn đề sức khỏe tâm lý. Nghiên cứu này nhằm hiểu rõ hơn về các tác động đó, từ đó đề xuất các giải pháp cân bằng giữa lợi ích thương mại và trách nhiệm xã hội. Nghiên cứu sử dụng phương pháp kết hợp, bao gồm khảo sát người dùng, phân tích dữ liệu hành vi và áp dụng các kỹ thuật học máy (Machine Learning) để xác định các mô hình tương tác và ra quyết định của người dùng. Các dữ liệu được thu thập và phân tích nhằm làm rõ cách thuật toán của TikTok tác động đến hành vi và tâm lý người dùng. Kết quả nghiên cứu chỉ ra rằng thuật toán gợi ý của TikTok không chỉ tăng cường mức độ gắn kết của người dùng mà còn thúc đẩy hành vi tiêu dùng thông qua quảng cáo cá nhân hóa. Thiết kế gây nghiện của nền tảng tạo ra một "vòng lặp ludic" khiến người dùng khó rời khỏi ứng dụng. Đồng thời, thuật toán này cũng định hình các chuẩn mực xã hội, khuếch đại các xu hướng và thách thức lan truyền. Nghiên cứu không chỉ cung cấp cái nhìn sâu sắc về cách thuật toán TikTok ảnh hưởng đến người dùng mà còn đóng góp vào các thảo luận rộng hơn về đạo đức trong việc ứng dụng trí tuệ nhân tạo. Những phát hiện này có thể hỗ trợ các nhà hoạch định chính sách, doanh nghiệp, và nhà nghiên cứu trong việc phát triển các chiến lược quản lý nền tảng mạng xã hội, đảm bảo cân bằng giữa trải nghiệm người dùng, lợi ích thương mại và trách nhiệm xã hội.

RESEARCH OVERVIEW

1. Overview

To conduct research on "Analyzing User Behavior Influenced by TikTok's Search Recommendation Algorithm," our team will implement three main stages. The first phase focuses on collecting data through surveys and analyzing user data from TikTok, particularly from individuals who frequently shop on the platform. After collecting sufficient data, an in-depth examination and analysis will be conducted to assess both the effectiveness and potential risks of the AI algorithm, thereby revealing its relationship with user experience. Ultimately, comprehensive evaluations will guide the development of solutions and strategies for optimal algorithm utilization, benefiting stakeholders in marketing, including businesses, advertisers, consumers, content creators, brands, and researchers in algorithm development. This study combines the analysis of TikTok's AI algorithm with demographic aspects, such as changes in behavior, thoughts, and emotions of users after engaging with the application.

2. Research Objectives

The overarching objective of this study is to examine and evaluate the impact of factors in the search recommendation algorithm model on TikTok users' attitudes and behavioral intentions. Specifically, the study will measure the influence of these factors on user behavior, particularly attitudes and shopping intentions through TikTok Shop in Ho Chi Minh City. Additionally, it will develop a research model and test measurement scales to better understand the relationship between algorithmic factors and consumer behavior. Based on the research findings, the study will provide recommendations for managers to attract the youth in Ho Chi Minh City, enhance the appeal of TikTok Shop, and optimize search recommendation algorithms to promote users' shopping intentions.

3. Research Questions

The study was conducted with three objectives:

1) To test and measure the factors in the search recommendation algorithm model that influence the attitudes and behavioral intentions of TikTok users.
2) To develop a research model and test the measurement scales for the factors of search recommendation algorithms affecting user behavior in terms of attitudes and purchase intentions through TikTok Shop in Ho Chi Minh City.
3) To propose recommendations for managers to enhance the attractiveness of this market and attract young consumers in Ho Chi Minh City by focusing on algorithms that influence purchase intentions analyzed through the model.

Based on the issues presented in the study overview, the research team posed the following questions:

- **RQ1:** What factors affect the attitudes and behavioral intentions of TikTok consumers, and how are these factors interrelated?
- **RQ2:** Do TikTok users' attitudes and purchase intentions through TikTok Shop get influenced by these search recommendation algorithms?
- **RQ3:** Based on this, what managerial implications can be drawn for businesses operating in the TikTok Shop marketplace to attract users and increase revenue?

4. Research Objects

The study focuses on analyzing user behavior on the TikTok platform, emphasizing the impact of recommendation algorithms on usage habits and purchasing decisions. It explores various aspects such as content interaction (views, likes, comments, shares, and video watch duration), shopping behavior (purchase frequency, product/service types, and spending levels), and platform dependence (daily usage time and discomfort when not using TikTok). The research also investigates behavioral changes influenced by trending content and challenges, along with user awareness of how TikTok's algorithms operate. Additionally, it examines psychological and cognitive factors, including social connection, life satisfaction of excessive TikTok use. The study further highlights affects on platform usage and purchase decisions. This comprehensive analysis aims to provide insights into how TikTok's algorithms shape user behavior and decision-making processes.

5. Research Scopes

The total duration of the study is one month. Data research will gather data from a sample of 15,000 videos, along with metadata and user comments, aiming to highlight videos with high user interaction levels. Metrics such as likes, shares, comments, and views have been standardized (Scikit-learn, 2023) and serve as the main inputs for the algorithm. The research focuses on the TikTok platform and its recommendation algorithm, with the goal of evaluating the impact of content recommendation mechanisms on user behavior and purchasing intentions. The findings from this study aim to provide empirical evidence to better understand the effects of the algorithm on users, as well as offer suggestions for future research in the fields of digital media and e-commerce on the TikTok platform.

6. Research Methodology and Process

6.1. Applying Classification for Analyzing TikTok User Behavior

In order to collect and process TikTok user data while maintaining data consistency and integrity, the study makes use of quantitative approaches. Based on important behavioral indicators like the frequency of content interaction, the recentness of engagement, and overall activity levels, users are divided into groups using machine learning classification approaches. Understanding trends and preferences in user involvement is based on these metrics.

Utilizing feature engineering techniques, which convert unstructured engagement data into analytically useful properties, classification accuracy is improved. To divide people into relevant groups, the study uses clustering methods like K-Means. The consistency and applicability of these categories are also confirmed statistically, guaranteeing that the final user segments appropriately represent unique behavioral patterns.

6.2. Experiment Results

After the classification procedure, experimental tests are carried out to determine how effective the suggested strategy is. To find the ideal amount of user clusters, performance indicators such elbow method analysis and silhouette scores are used. The findings show considerable differences in content consumption patterns, number of interactions, and duration of engagement among user categories.

Through the discovery of user behavior patterns, this study offers important new information on how various audience segments interact with TikTok content. These results can be used to inform more tailored marketing campaigns, better content suggestions, and more effective user retention initiatives. Finally, by providing useful applications for both platform developers and content producers, the study advances our knowledge of TikTok user dynamics.

7. Tools and Programming Language

Python was the primary tool for data preprocessing, analysis, and modeling due to its versatility and extensive libraries. Key libraries used include:

- **Pandas:** For data manipulation, cleaning, and transformation.
- **NumPy:** For numerical computations and array operations.
- **Matplotlib and Seaborn:** For creating detailed and insightful visualizations.
- **Scikit-learn:** For implementing machine learning algorithms, feature selection, and model evaluation.
- **NLTK and SpaCy:** For natural language processing tasks, such as text analysis and tokenization.
- **TensorFlow and PyTorch:** For developing and training deep learning models, particularly in behavioral pattern recognition.

8. Scientific and Practical Significance

8.1. Urgency

Artificial Intelligence (AI) technology is revolutionizing how social media is experienced while fundamentally reshaping consumer engagement practices. AI-driven platforms like

TikTok play a crucial role in shaping purchasing decisions by curating personalized content and product recommendations based on user interactions. This study, utilizing in-depth interviews and user data analysis from TikTok, aims to explore how AI-driven recommendation systems influence consumer engagement and purchasing patterns.

According to a survey by Kang et al., preliminary results indicate that TikTok users exhibit a high level of adaptability to AI-powered personalization. However, their interaction with AI extends beyond passive content consumption; it actively shapes their shopping preferences and purchase intentions. As users engage with AI-driven recommendations, a reciprocal feedback loop emerges – where user behavior continuously refines the recommendation algorithm, ensuring that product suggestions align more precisely with their evolving preferences. This process significantly impacts impulse buying, brand exposure, and consumer trust in AI-driven marketing.

In the era of digital commerce and big data, social media platforms like TikTok have become powerful marketplaces where AI algorithms dictate product visibility, influence purchasing journeys, and drive sales conversions. This research will analyze the operational mechanisms of TikTok's recommendation algorithm and its effects on consumer behavior, including browsing habits, engagement with sponsored content, frequency of purchases, and psychological factors such as brand perception and purchase satisfaction. By examining these dynamics, the study seeks to assess the broader implications of AI-powered recommendation systems in shaping modern consumer habits, reinforcing brand loyalty, and transforming the digital shopping experience.

8.2. Novelty and Differences Compared to Other Social Media Platforms

While other platforms often prioritize content displayed from friends or pages that users follow, TikTok primarily relies on its algorithm to recommend content based on user behavior and preferences. This means that users can view videos from individuals they do not follow, as long as the algorithm predicts that they might find the content interesting. This is a significant departure from platforms like Facebook or Instagram, where the friend network plays a critical role in content distribution.

Analyzing consumer behavior through TikTok's recommendation algorithm introduces several novel aspects and distinctions compared to traditional methods, primarily due to its ability to automatically and continuously collect, process, and analyze large-scale data. Below are some key highlights:

8.2.1. Short-Form Video Format

TikTok primarily focuses on short-form videos, unlike YouTube, which emphasizes longer videos, or Facebook/Instagram, which offers various content formats. This creates a unique user experience, promoting rapid and continuous interaction. The short-form video format allows users to consume a large amount of content in a short period. Statistics reveal that as of September 2021, TikTok had welcomed 1 billion monthly active users and was the most downloaded app in 2020. In 2024, TikTok reached 1.04 billion monthly visits, with impressive figures such as an average usage time of 53.8 minutes per day and 137 million downloads in Q1 2024 alone. As a result, TikTok has become a major competitor to other social media and video platforms like Instagram and YouTube, pushing them to emulate TikTok's success by introducing similar features (e.g., Instagram Reels or YouTube Shorts – short videos with recommendation-based distribution systems). Even major e-commerce platforms in Vietnam, such as Shopee, have taken notice.

Social Media	Average Time Spent per Day
TikTok	**95 minutes**
Instagram	**62 minutes**
X (Twitter)	**30 minutes**
Snapchat	**19 minutes**

Figure 0.1: Average daily usage time across platforms. Source: Backlinko, 2024.

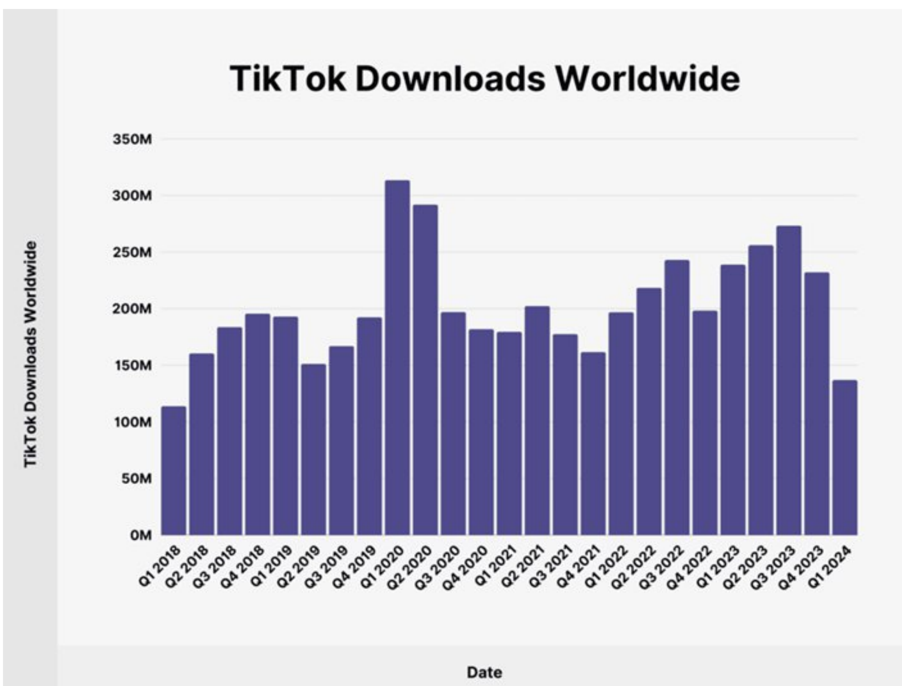

Figure 0.2: Global TikTok app download statistics. Source: Backlinko, 2024.

8.2.2. High-Level Content Personalization

TikTok's algorithm combines collaborative filtering and content-based filtering to provide highly personalized content recommendations for each user, a stark contrast to traditional analysis methods that often rely on audience groups or market segmentation. Specifically, TikTok records users' behavioral history, and its algorithm not only considers likes, shares, or comments but also factors in video watch time, hashtags, and even video image and audio characteristics. When new users join TikTok, they receive content from a traffic pool, typically featuring highly liked and diverse videos. Swiping up and down to skip videos or staying briefly on a page indicates disinterest in the content. Conversely, prolonged engagement or a series of interactions, such as following, liking, and commenting, increases the likelihood of the video being recommended again.

TikTok's algorithm, which makes use of big data research and real-time tracking, can swiftly identify new trends. This allows marketers and content producers (such as UGC and PGC) to immediately identify trends and create material that appeals to users. This promotes sales, improves brand awareness, and even makes it easier for TikTok to create its own e-commerce site, TikTok Shop. Commerce integration can attract users, sometimes causing the FOMO effect, which can result in impulsive buying. Because of this, TikTok now poses a serious threat to big rivals like Shopee and Lazada.

8.2.3. The Principle of "Escapism"

According to research by co-authors Marc K. Peter and Markus Rach comparing TikTok's algorithm to Facebook's, TikTok's algorithm also operates on the principle of "escapism," focusing on delivering entertaining, engaging, and surprising content to captivate users. This highlights TikTok's distinction from other platforms that typically rely on friend networks. These factors provide researchers with deeper insights into the motivations and emotions of users when engaging with social media.

Table 1. Correlation of motivators for the use of TikTok (r) (n = 217)

	Sample (n)	Escapism	Archiving	Self-expression	Communication	Voyeurism
Germany	165	0.561	0.681	0.648	0.701	0.658
Austria	31	0.582	0.647	0.614	0.648	0.678
Switzerland	21	0.527	0.683	0.598	0.694	0.681

Table 2. Correlation of motivators for the use of Facebook (r) (n = 217)

	Sample (n)	Escapism	Archiving	Self-expression	Communication	Voyeurism
Germany	165	0.731	0.467	0.692	0.652	0.399
Austria	31	0.649	0.482	0.638	0.619	0.426
Switzerland	21	0.663	0.501	0.546	0.605	0.417

Figure 0.3: TikTok user behavior data. Source: How TikTok's Algorithm Beats Facebook & Co. for Attention Under the Theory of Escapism, 2022.

Instead of relying on traditional demographic data (age, gender, location), TikTok's algorithm primarily analyzes user behavior to make recommendations. This enables researchers and businesses to access more detailed and accurate insights into user preferences and habits, minimizing potential biases that may arise from demographic data.

CHAPTER 1: THEORETICAL BACKGROUND AND RELATED WORK

1.1. Overview of Domestic and International Research

1.1.1. Origins of TikTok

TikTok launched as a social media platform in both China and the United States. Initially, the platform was designed with educational purposes in mind. However, Musical.ly, a precursor to TikTok, captured the attention of young users in the U.S., who used it to create short, trend-setting videos synced to popular songs. In September 2016, the Chinese tech company ByteDance launched Douyin in China – a social media platform enabling users to create and share short videos. Douyin stood out with its impressive recommendation algorithm, which could suggest videos tailored to user preferences.

To expand its user base beyond China, ByteDance acquired Musical.ly and its user community in November 2017, rebranded Musical.ly as TikTok, and integrated Douyin's advanced algorithm into TikTok. This acquisition and rebranding explain why media reports from 2017 to 2022 often referred to TikTok as a "singing and dancing app." As of July 2024, Douyin and TikTok operate in parallel but are geographically restricted – Douyin is only available to users in China, while TikTok is accessible to users outside China. Although Douyin (sometimes called "Chinese TikTok") and TikTok essentially share the same social media framework, this study focuses solely on TikTok.

TikTok is headquartered in Singapore and, according to the company, operates independently of its parent company, ByteDance, which is based in China – a business structure that has attracted political scrutiny, particularly from the U.S. TikTok's financial performance during this period saw significant growth: revenue of $150 million in 2018 skyrocketed over 100-fold to $16 billion in 2023, with profits reaching $6 billion (according to Business of Apps, 2014), primarily driven by advertising sales.

1.1.2. Overview of TikTok Internationally

In the context of a market economy and globalization, social media platforms such as Facebook, Instagram, and TikTok have experienced significant growth, exerting a profound influence on consumer behavior worldwide. According to United Nations data on global population, the world's population currently stands at approximately 8.16 billion people – an increase of 70 million (0.86%) compared to the previous year. In other words, nearly two-thirds of the global population now use the internet and social media platforms.

Data from various reports show that the number of regular social media users has increased by 5.6%, with users spending an average of 2 hours and 23 minutes per day on these platforms. Social media is also increasingly integrating itself into users' online habits and daily routines, influencing activities such as searching for locations, shopping, entertainment, and dining.

Prabhakar Raghavan, Senior Vice President at Google, once shared that "an estimated 40% of Gen Z users turn to TikTok or Instagram to search for restaurants instead of using Google." For this demographic, TikTok and Instagram are the top two choices for discovering entertainment content, with survey results showing that 62% of respondents preferred TikTok, while 66% chose Instagram.

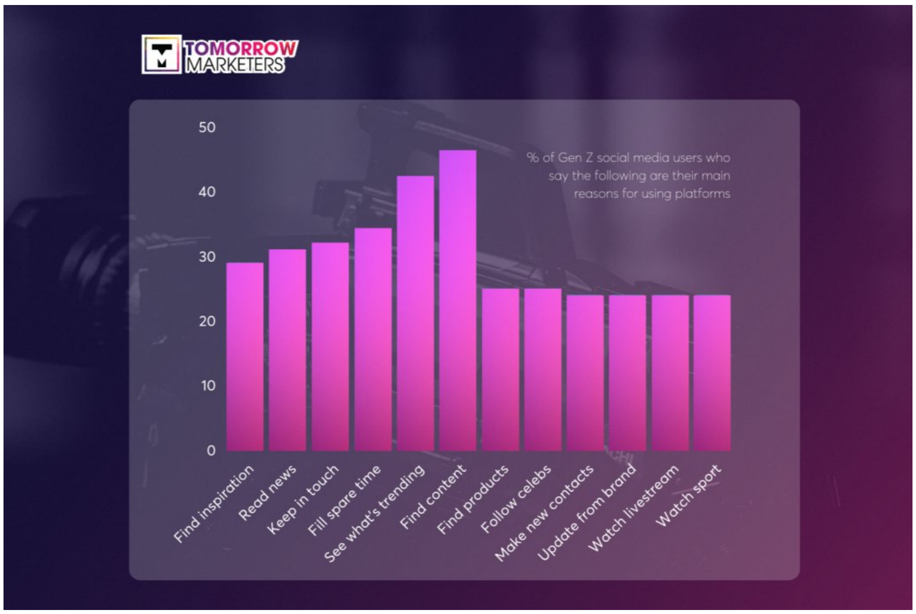

Figure 1.1: Survey on the Purpose of Gen Z Users Utilizing Social Media Platforms. Source: TOMORROW MARKETERS, 2023.

TikTok, one of the most popular apps today, has surpassed 5.04 billion users worldwide, according to the WeAreSocial Digital 2024 report. Notably, TikTok is the platform with the highest average usage time on Android, with each Android user spending an average of 34 hours per month – equivalent to over one hour per day. Globally, TikTok has become a key subject of research in the fields of e-commerce and consumer behavior. According to a GlobalWebIndex (2023) study, the time spent on TikTok is significantly higher compared to other platforms such as Instagram (11 hours) and Facebook (19 hours). Some researchers have cautioned that TikTok's highly personalized content could lead to shopping addiction behavior. TikTok is widely considered more effective at influencing user behavior compared to competitors like Facebook and Instagram, thanks to its short-video format and real-time interaction features, such as live streaming and live product trials. According to an eMarketer (2023) report, TikTok accounted for 18% of global social commerce revenue. With over 5 billion users, TikTok not only impacts individual consumer behavior but also plays a significant role in shaping consumption trends on a large scale.

Figure 1.2: The Influence of #TikTokMadeMeBuyIt on Purchasing Behavior. Source: Advertising VietNam, 2022.

Companies have leveraged TikTok to promote their products, leading to a shift in how they engage with consumers. A prime example is the #TikTokMadeMeBuyIt campaign, which garnered over 30 billion views and directly contributed to increased sales for major brands. TikTok supports not only large corporations but also small businesses and individual sellers in promoting their products. According to a Shopify (2023) report, more than 50% of small businesses using TikTok Shop saw significant revenue growth in their first year on the platform. As a result, TikTok has become one of the top applications globally, delivering substantial benefits to users, businesses, and the economy.

1.1.3. Overview of TikTok Domestically

TikTok's personalized content suggestions and continuous promotional campaigns have encouraged impulsive buying behavior, particularly among first-year university students and female users. A survey by the University of Education – Vietnam National University, Hanoi (2024) found that over 76.92% of students admitted feeling regret over unnecessary purchases influenced by TikTok content. Several studies have proposed solutions to enhance personal financial management. Students are encouraged to create budgets and avoid emotional shopping. A report by the Journal of Psychology and Education (2024) noted that 47.12% of students found budgeting helpful in controlling their spending more effectively.

Beyond influencing shopping behavior, TikTok has also impacted how students interact with their communities. Recently, shopping through live streams or relying on reviews from Key Opinion Leaders (KOLs) has become a popular trend. According to reports, 60.33% of students use TikTok to make purchases through livestream sessions and highly-rated community content.

TikTok has become one of the most popular social media platforms globally, with billions of downloads and millions of active users. In Vietnam, TikTok continues to grow rapidly, attracting a large number of users, especially students and young people. Data from early 2024 indicated that TikTok had 67.72 million users aged 18 and above in Vietnam, representing 92.6% of the adult population. According to a study, 93.39% of students use TikTok, and 90.91% use it for shopping purposes. This demonstrates that TikTok is not just a place for entertainment but also an essential channel for online shopping. The platform provides a diverse space where users can share short videos, create content, and connect with the community.

1.2. Evaluation of Differences in Scientific and Technological Advancements Domestically and Internationally

1.2.1. Algorithms and Artificial Intelligence (AI)

TikTok stands out with its powerful AI algorithm capable of precisely personalizing content. This algorithm analyzes user behavior (such as views, interactions, and watch time) to recommend videos tailored to individual preferences, creating a seamless and addictive user experience. TikTok's user-focused recommendation strategy is more prominent than the friend-network-based strategies employed by other social media platforms, making TikTok an ideal subject for studying the interplay between AI algorithmic power and user influence, along with its associated consequences.

TikTok's algorithm continually learns from user data, improving its content suggestions over time. The platform also offers AI-powered video editing tools, including filters, effects, and smart audio editing features. TikTok's highly-rated personalization capabilities allow users to quickly find videos of interest, creating an endless viewing loop.

1.2.2. Design and User Experience (UX)

TikTok's simple, intuitive interface allows users of all ages to quickly familiarize themselves with the platform. One of TikTok's strengths is its short-video format, initially limited to 60 seconds and later expanded to 3 minutes, making it easy for users to consume content rapidly. The platform's high-entertainment videos are a key factor in attracting users, with easily shareable and viral content driving significant engagement. TikTok encourages user interaction through features like likes, comments, shares, and video saving, enhancing community engagement and connection. The algorithm helps users discover new content, fostering a continuously exciting experience that keeps users hooked.

When users first access the app, they register using a phone number, email, or third-party service like Google or Facebook ("social sign-in" supported by TikTok's OAuth Authentication protocol). Users then provide a birth date (minimum global age requirement of 13), create a username and password, and are encouraged to upload a profile picture and write a short bio. Users choose their preferences (e.g., "comedy," "gaming," "beauty and fashion," "health and fitness"), which the algorithm uses to generate an initial "For You" page (FYP). TikTok then delivers full-screen videos for users to watch and interact with. Scrolling up provides new videos, creating a continuous stream of algorithm-generated content. Users can engage with videos in various ways, such as liking, commenting, sharing, and saving. Videos are often tagged with hashtags that are searchable by other users. Additionally, users can create their own videos using a variety of music and visual and sound effects.

1.2.3. Business Model

TikTok's business model centers on creating a short-video ecosystem that attracts both users and advertisers through easily consumable content. The platform has integrated e-commerce features, allowing users to shop directly on the app, blending entertainment and shopping into a "shoppertainment" model. TikTok frequently offers promotions with affordable prices to trigger FOMO (fear of missing out).

TikTok plays a crucial role in influencer marketing, where KOLs (Key Opinion Leaders) and KOCs (Key Opinion Consumers) create and promote content that helps brands connect with a broad consumer base. As such, TikTok is not only a platform for creative content sharing and life moments but also a hub for product showcasing and business sales interactions. This model is further strengthened by ad revenue, as TikTok leverages its large user base and ad personalization capabilities to optimize marketing campaign effectiveness. According to TikTok's 2024 GIGAN statistics, influencers with follower counts between 500,000 and 1 million charge an average of $150 to $3,500 per promotional post. The influencer tiers on TikTok are as follows:

- **Nano Influencers** (1,000 to 10,000 followers): $20 to $100 per post
- **Micro Influencers** (10,000 to 100,000 followers): $30 to $400 per post
- **Macro Influencers** (100,000 to 1 million followers): $150 to $3,500 per post
- **Mega Influencers** (more than 1 million followers): Over $3,500 per post

1.3. Overview of TikTok's Recommendation Algorithm

1.3.1. Concept

The recommendation algorithm on TikTok, also known as the "For You" algorithm, is a sophisticated system that shows users films that are most likely to be of interest to them. This algorithm incorporates elements like user-generated material, particular user preferences, and previously liked and seen movies. To provide individualized recommendations, it uses content-based recommendation techniques and collaborative filtering. TikTok's lack of a homepage or start button, which allows users to view just content chosen by the recommendation algorithm and has videos that play immediately as the app launches, is one of its distinctive content distribution and discovery strategies. TikTok's extraordinary popularity can be attributed to these technologies.

The algorithm's AI-based recommendation system plays a more dominant role in shaping user experiences than traditional search mechanisms, as TikTok autonomously presents suggested content to users. This strong control over what users see limits personal choice and largely defines the TikTok experience. Research identifies two key factors related to the recommendation algorithm: perceived recommendation accuracy and perceived randomness of suggestions.

1.3.2. The Function of TikTok's Recommendation Algorithm

TikTok integrates AI-driven features into key social networking aspects, including content consumption, content creation, and social networking connections. The "For You" feed is personalized for each user based on their previous interactions and preferences, including watch time, likes, comments, shares, and follows (Bandy & Diakopoulos, 2020).

To curate and recommend videos to users, TikTok employs natural language processing to identify textual and audio elements in favored videos, computer vision to classify video visual components, and analysis of hashtags and captions linked to those videos. TikTok's algorithm is so powerful that it can learn user preferences and vulnerabilities within just 40 minutes (Lovejoy, 2021).

TikTok also supports content creators by helping them produce viral videos using trending sounds, hashtags, filters, and optimal posting times (Davis, 2019). Ma and Hu (2021) describe TikTok's AI-based algorithm as its core technology, explaining that it leverages computer vision to infer user preferences and interaction patterns, thereby creating personalized and unique user experiences.

Schellewald (2022) suggests that TikTok's algorithm can be configured and adjusted as part of the "everyday meaning-making processes." Overall, the dynamic power relationship between users and TikTok's AI-driven algorithm is flexible, interactive, and mutually defined.

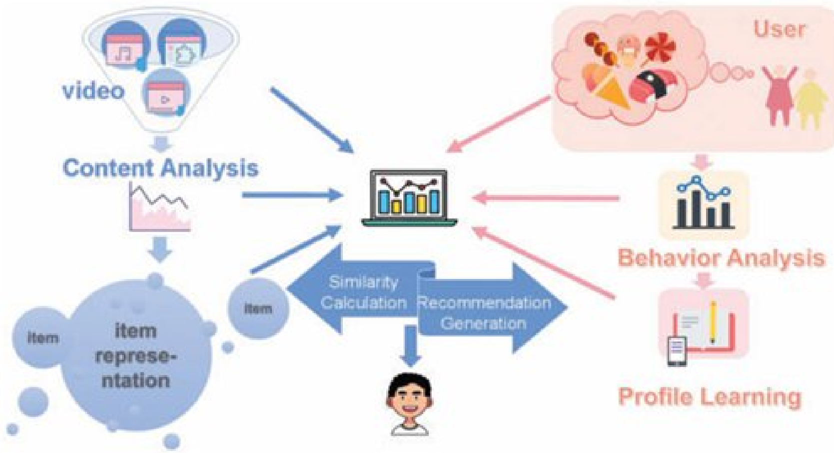

Figure 1.3: The main process of content-based recommendation. Source: Short Video Recommendation Algorithm Incorporating Temporal Contextual Information and User Context, 2022.

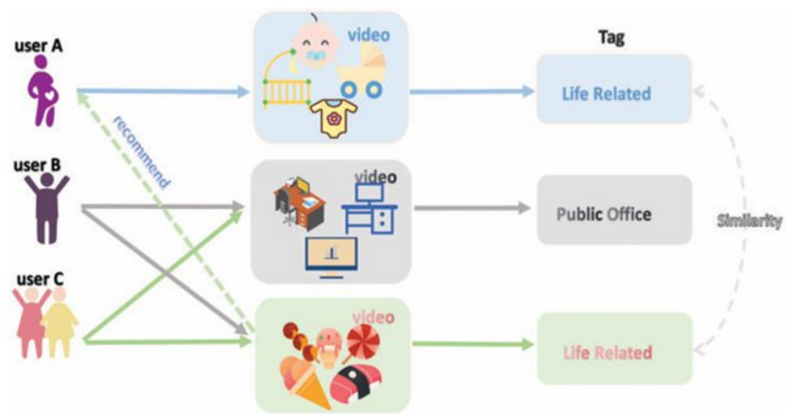

Figure 1.4: Content-based recommendation. Source: Short Video Recommendation Algorithm Incorporating Temporal Contextual Information and User Context, 2022.

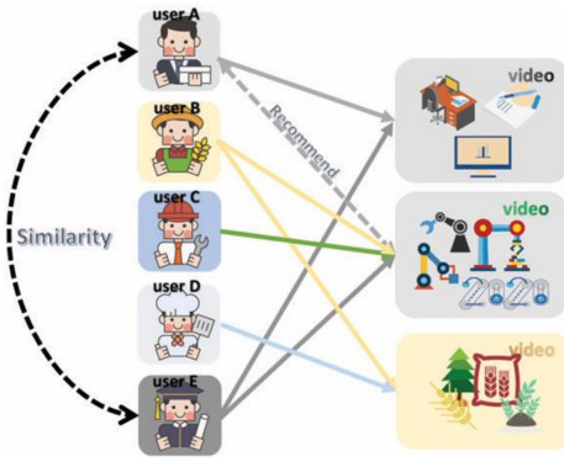

Figure 1.5: User-based collaborative filtering. Source: Short Video Recommendation Algorithm Incorporating Temporal Contextual Information and User Context, 2022.

TikTok's recommendation algorithm considers a variety of factors to decide which videos to display to users:

- **User Interaction:** The algorithm monitors user engagement, such as liking, commenting, sharing, or watching videos in full. Higher interaction rates with certain content increase the likelihood of similar videos appearing in the user's feed.
- **Video Information:** Specific metadata associated with a video plays a role in recommendation decisions. This includes hashtags, background music, and effects used in the video, which help TikTok classify and match content with user preferences.
- **Account and Device Settings:** Information such as country, language preferences, and device type is factored in to localize and optimize the user's experience on the platform.
- **Similar Content:** The algorithm recommends videos that are similar to those the user has previously engaged with. If a user interacts with certain themes or trends, TikTok will continue displaying related content to maintain user interest and engagement.

1.4. Issues

1.4.1. Impact of Personalized Search Recommendations on User Behavior

TikTok's personalized search recommendations have a significant impact on users' shopping behavior, particularly among younger audiences. These suggestions are tailored based on users' previous shopping habits, preferences, and online interactions. The platform's AI-driven personalization enables users to discover products they might not have considered before, creating a seamless integration between content consumption and purchasing decisions.

Key statistics highlight TikTok's growing role in shaping shopping behaviors:

- **TikTok Shop Usage:** Among daily TikTok users, 30% have used TikTok Shop, and 33% have shown interest in using it. Weekly TikTok users also demonstrate high engagement, with 26% using TikTok Shop and 29% expressing curiosity.
- **Impulse Purchases:** 71.2% of consumers reported making purchases after stumbling upon products through TikTok's Stories or news feed.
- **App Utilization for Shopping:** 58.2% of users have used TikTok as a shopping platform.
- **Influencers' Role:** 45% of users have purchased items based on paid recommendations from influencers, and 39.1% actively search for products in TikTok stores.

These statistics reflect how TikTok's personalized algorithms not only drive consumer engagement but also cultivate an evolving trend of social commerce, where entertainment and shopping are seamlessly blended to influence user behavior.

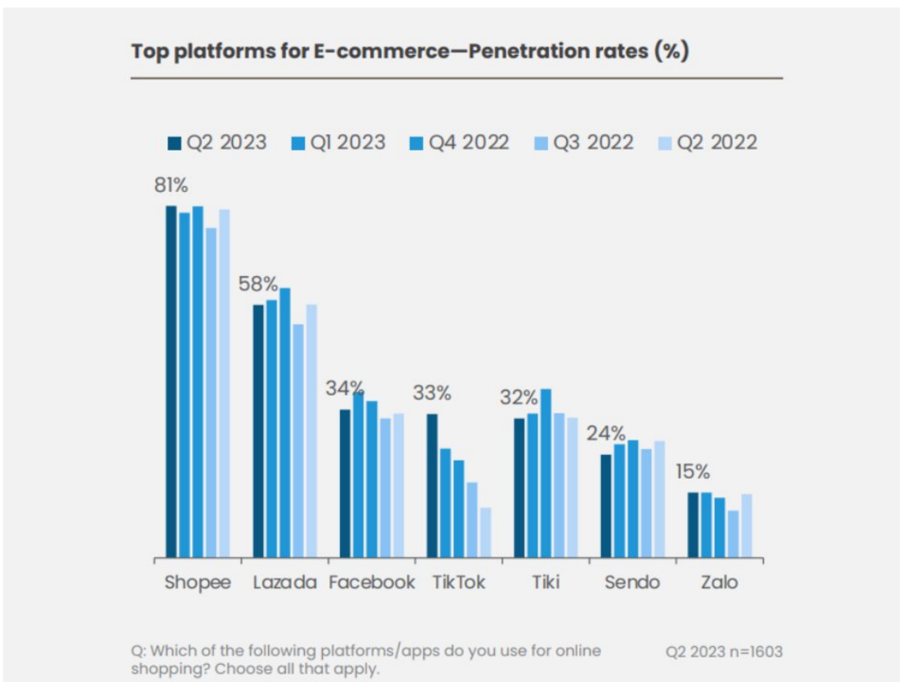

Figure 1.6: Top platforms for E-commerce – Percentration rate (%). Source: TOMORROW MARKETERS, 2023.

The accuracy of TikTok's personalized recommendations plays a crucial role in shaping users' purchasing decisions. When suggested content aligns with users' interests, engagement levels increase, leading to more frequent purchases. Conversely, irrelevant or inaccurate recommendations may lead to frustration, loss of trust in the platform, and decreased shopping frequency. Personalized search suggestions can also stimulate impulsive buying behavior, facilitated by AI-driven browsing analysis, simple payment processes, and attractive promotional campaigns. These factors contribute to a higher conversion rate from purchase intention to actual buying behavior, even for unplanned purchases.

TikTok's algorithm reinforces user preferences by continuously suggesting content based on prior interactions. This creates a "closed loop," where users are gradually confined to a specific type of content. For example, frequent engagement with cooking videos leads to similar content recommendations, reinforcing habits and interests over time. Studies by Samuel Hardman Taylor and Y. Anthony Chen suggest that regular exposure to positive content that resonates with a user's identity fosters a sense of social connection. This perception of being understood by TikTok's algorithm enhances users' feelings of validation and connection with others. TikTok trends can also shape social behaviors beyond the platform. Viral challenges may encourage users to adopt specific behaviors, some of which carry risks. Additionally, fashion styles, communication methods, and language trends are heavily influenced by TikTok's popular content.

1.4.2. Addictive Nature of TikTok

TikTok's user-friendly interface and never-ending supply of brief films are made to keep users interested for hours on end. Diverse and imaginative content attracts users, and it's simple to lose track of time when swiping to see the next video. TikTok offers a private platform for amusement and self-expression for a large number of users. TikTok is "a small private space, even more private than other social networks," according to Linda, a 21-year-old. "TikTok is very imaginative, and once it starts, it's completely customized for you, so it never feels weird or dull," she said. In agreement, Karen (age 23) pointed out that "the For You page quickly adapts to what you really want to see." TikTok also creates a "ludic loop," where providing more data to the algorithm enhances its suggestions, keeping users engaged longer. Carol (26 years old) shared that "all the videos I watch now are perfectly tailored to what I like." Robert (22 years old) commented that scrolling through TikTok For You Page (FYP) "feels intuitive and requires little thought, just swipe down and watch the next video.

Research indicates that excessive TikTok usage may contribute to functional neurological disorder (FND) symptoms among teenagers, who may mimic the symptoms displayed in videos. The constant exposure to comparison-driven content can negatively impact self-esteem and mental health. TikTok's algorithm-driven content delivery, combined with filters and visual effects, poses unique challenges by reinforcing dependent usage patterns and negatively affecting users' mental health. Additionally, TikTok's personalized content can create "information bubbles" or "echo chambers" that limit users' exposure to diverse perspectives. This reinforcement of personal viewpoints may lead to confirmation bias, reduced critical thinking, and impaired social awareness.

1.4.3. Commercial Objectives

Advertising is TikTok's primary revenue source. The platform's algorithm targets ads based on users' personal information and behaviors to maximize ad effectiveness. Personalized ad suggestions help consumers discover products and services that meet their needs, enhancing the shopping experience. Users often feel understood and valued, leading to increased purchase frequency. These consumer-friendly product suggestions contribute to TikTok's unique social commerce experience.

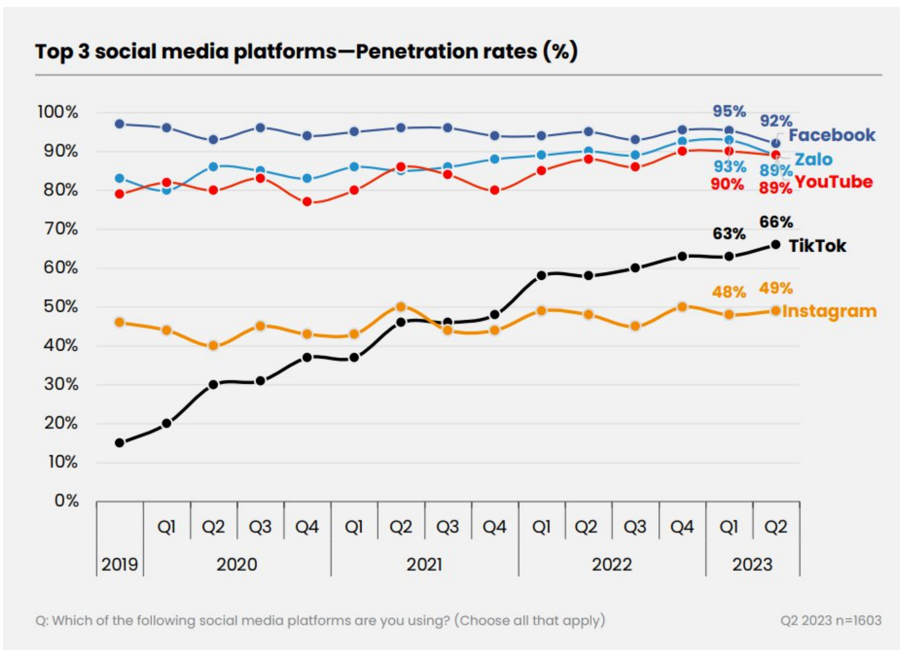

Figure 1.7: Top 3 social media platforms. Source: TOMORROW MARKETERS, 2023.

Some researchers argue that TikTok operates under the model of "surveillance capitalism," where user data is leveraged for profit generation. The ultimate goal of the platform is not only to provide entertainment but also to manipulate user behavior on a large scale. TikTok collects a vast amount of data on users, including personal information, preferences, habits, and interactions on the platform. This data is then used to enhance the recommendation algorithm, personalize user experiences, and target advertisements.

The ads on TikTok are designed to foster a sense of belonging and homogeneity among users, making them feel like they are part of a community. This feeling can drive purchasing behavior, as users are more likely to buy products they feel aligned with. Online consumers, particularly students, are influenced by social media platforms and word-ofmouth marketing, including reviews from previous customers and influencers. This allows them to easily compare products based on real-life experiences.

In previous years, price was the most important factor in purchase decisions. However, reviews have now emerged as the most influential factor on student purchasing behavior, surpassing other considerations such as price, free shipping, brand, and recommendations from friends and family.

1.5. Resolved Matters

1.5.1. Issues Related to User Behavior Influence

Studies have shown that TikTok users may not fully understand how the algorithm works or its influence on their behavior. Many users perceive the algorithm as a tool to enhance their experience without realizing its potential to shape and control their attention. To address this, some studies suggest increasing algorithm transparency and educating users about how it functions. Research has explored how TikTok's personalized recommendations can lead to impulsive shopping behaviors, especially among students. This raises questions about the platform's responsibility in protecting users from the negative effects of advertising and promoting healthier shopping habits. Studies also recommend that businesses focus on the quality and creativity of their ads rather than solely aiming to drive sales.

To address symptoms of functional neurocognitive disorder (FTLB) in teens, TikTok has suggested measures to reduce exposure to harmful content, alongside raising mental health awareness. These measures have significantly reduced symptoms. TikTok's algorithm can create "information bubbles" or "echo chambers" where users are only exposed to content and viewpoints similar to their own. This can reinforce biases and limit exposure to diverse perspectives. Some studies propose that platforms should optimize algorithms to provide varied content and encourage exposure to different viewpoints.

1.5.2. Issues Related to Addiction

Studies have analyzed the factors that make TikTok addictive, including its simple interface design, reward mechanisms, "ludic loop," and the lack of user control. These studies highlight how TikTok exploits psychological mechanisms to keep users engaged. TikTok developers have faced pressure from critics and the public regarding the platform's addictive nature. As a result, they have introduced "anti-addiction" features such as time-limit notifications and reminders to take breaks. However, the effectiveness of these measures is still under investigation. Studies have found that many TikTok users are not fully aware of the platform's addictive nature. While they may feel they are using TikTok in a controlled manner, the algorithm may be manipulating their behavior. To tackle this, education about the signs of addiction and strategies to manage social media use is needed.

1.5.3. Issues Related to Commercial Objectives

Research has analyzed TikTok's business model, where advertising is the primary source of revenue. The algorithm is used to collect user data and target ads effectively. These studies also highlight the tension between TikTok's profit goals and its responsibility to users. Some studies argue that TikTok operates under the model of "surveillance capitalism" where user data is used to generate profit. This model raises questions about user privacy and the transparency of tech companies. Studies emphasize the importance of transparency in TikTok's operations. Users need to understand how the algorithm works, how their data is used, and how platform decisions affect their experience. This transparency can help users make more informed decisions and have better control over their TikTok experience.

1.6. Addressing and Resolving Outstanding Challenge

1.6.1. Issues Related to User Behavior Influence

Although studies have highlighted how TikTok's algorithm works, users still struggle to fully understand how the algorithm shapes their experience. Many have misconceptions about the algorithm's capabilities and purpose, often seeing it as just a tool without recognizing its potential impacts. Current research mainly focuses on the short-term effects of TikTok's algorithm. There is still limited research on the long-term impact of continuous exposure to personalized content, particularly regarding the cognitive, emotional, and social development of young users. A complex issue is determining the boundary between user autonomy and algorithmic influence. Are users truly choosing content freely, or are they being manipulated by algorithms designed to maximize engagement and time spent? More research is needed to clarify this issue. TikTok's algorithms may shape social norms by amplifying popular content and reducing the diversity of viewpoints. This could lead to negative outcomes such as increased mimicry of unhealthy trends. More research is needed to better understand TikTok's impact on social norms. While TikTok provides a platform for accessing information, its algorithm may create "information bubbles" where users are only exposed to similar viewpoints and pre-selected information. This can increase social polarization and reduce exposure to diverse perspectives, causing division and limiting multifaceted understanding.

1.6.2. Issues Related to Addiction

Although TikTok has introduced several "anti-addiction" features, it remains unclear whether these measures are truly effective in helping users control their time on the platform. More research is needed to assess their effectiveness and find better solutions.

Studies have shown a link between excessive TikTok use and mental health problems (e.g., anxiety, depression) and physical health issues (e.g., sleep disorders, reduced physical activity). However, many questions remain about the mechanisms and extent of this relationship. Current research mainly focuses on younger users. More studies are needed to examine whether TikTok's addictive nature differs across various user groups (e.g., older adults, those with mental health issues). While there are similarities across social media platforms, each platform has unique characteristics. Comparing user experiences, especially in terms of addiction, between TikTok and other platforms (e.g., Facebook, X) is essential for a more comprehensive understanding of technology's influence on behavior.

1.6.3. Issues Related to Commercial Objectives

Although some studies have analyzed TikTok's algorithm, the platform still lacks full transparency about its algorithms and decision-making processes. This makes it difficult for researchers, policymakers, and users to understand and control the platform's influence. TikTok is a commercial entity, and profit goals are a key factor in its operations. However, many questions remain about how TikTok balances its profit objectives with social responsibility towards users, particularly in areas like advertising, child protection, and combating misinformation. While personalized ads can benefit both businesses and users, they may also have negative effects, such as data abuse, impulsive shopping, and creating undesirable user experiences. More research is needed to understand these impacts and develop strategies to mitigate risks. TikTok is not just an entertainment platform; it's also a powerful marketing tool. TikTok's growth has significantly impacted other industries, such as music, fashion, and retail. Further research is needed to better understand these impacts and how industries can adapt to these changes. TikTok has content moderation systems to remove content that violates platform policies. However, the effectiveness of these systems is still limited, and concerns persist regarding bias and lack of transparency in the moderation process.

CHAPTER 2: DATA PREPARATION

2.1. Chapter Overview

This chapter provides a structured approach to preparing the TikTok user dataset for analysis. It outlines the data collection process, dataset structure, key variables, and necessary preprocessing steps to ensure data quality. These steps are essential for conducting **exploratory data analysis (EDA)** and developing **machine learning models** that can generate meaningful insights into user behavior on TikTok.

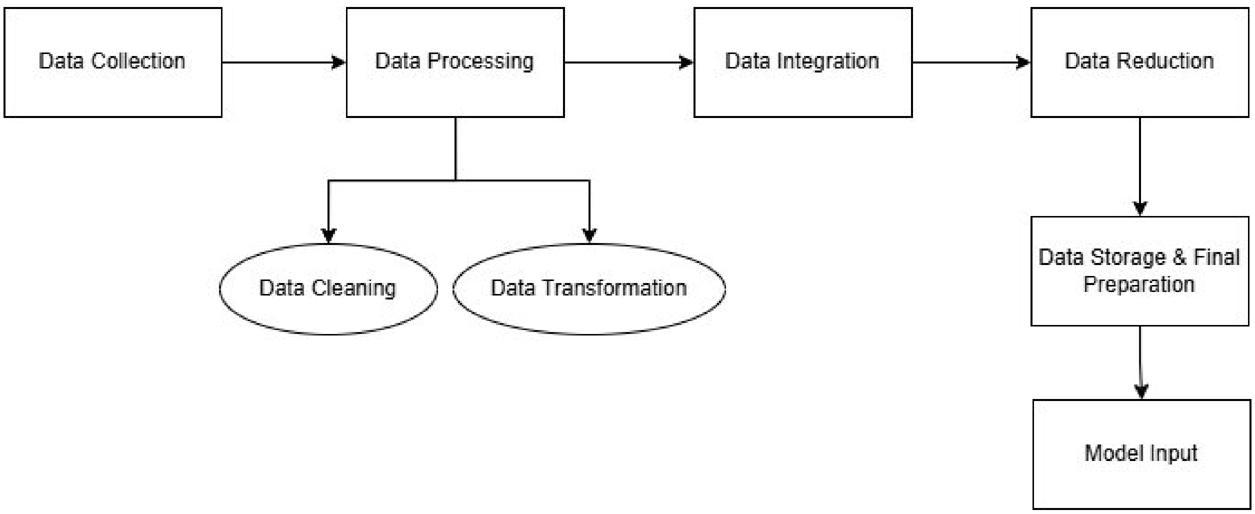

Figure 2.1: Data Preparation Process.

2.2. Data Understanding and Summary

The dataset is a **comprehensive collection** of TikTok user responses, including both **demographic** and **behavioral attributes** related to engagement, ad interactions, and shopping behavior. This data was gathered and enriched with additional user activity metrics extracted through **TikTok's public API**. The dataset contains a total of **1,200 responses**, representing a diverse range of TikTok users. Each row in the dataset corresponds to a unique respondent, while each column represents a specific variable related to their TikTok usage and online shopping habits.

2.2.1. Demographic Information

Understanding user demographics is crucial for segmenting the dataset and analyzing behavioral patterns across different age groups and employment statuses. These variables provide key insights into how demographic factors influence TikTok engagement and purchasing behavior.

The **user_id** is a main reference for each entry in the dataset, uniquely identifying each user. An analysis of behavioral patterns across age groups is made possible by the age variable, which records the user's numerical age. For consistency in analytical comparisons, the gender characteristic is encoded as **1 = Male and 2 = Female**, guaranteeing a standardized classification. Furthermore, the employment_status variable allows for an examination of the influence of employment on TikTok engagement, ad interaction, and consumer expenditure by classifying users as **1 = Employed and 2 = Unemployed**. These demographic characteristics aid in the identification of patterns pertaining to economic behavior and social influence on the platform.

2.2.2. TikTok Usage and Engagement

User engagement metrics provide a deeper understanding of how individuals interact with TikTok's platform. These variables help measure user activity levels, content consumption habits, and social influence within the TikTok ecosystem.

The average amount of time a person spends on TikTok each day is tracked by the **daily_usage_time** variable, which represents overall platform engagement. The dynamics of social networks are recorded by **tiktok_following_count**, which shows how many accounts a user follows overall, and **tiktok_follower_count**, which shows how many followers a user has. This measure is especially helpful for evaluating a user's reach and social influence within the TikTok community. The average amount of time (in seconds) that a user spends watching videos is finally measured by **avg_watch_time**. This variable is essential for figuring out how long people spend on a piece of content before scrolling, which provides information about viewer behavior and content retention.

2.2.3. Ad Exposure and Interaction

Advertising exposure and user interaction levels are key indicators of how users engage with TikTok's monetization ecosystem. Understanding the extent to which users interact with advertisements and livestream content provides valuable insights into their responsiveness to promotional materials and real-time engagement behavior.

While using TikTok, a user's daily exposure to adverts is indicated by the **ad_exposure_per_day** variable. Ad saturation in a user's browsing experience and its possible effect on engagement are evaluated with the use of this metric. **Livestream_exposure** also records how many livestreams a user watches each day, indicating their interest in influencer-driven promotions and real-time content. An important measure of user response and advertising efficacy is the **ad_click_rate**, which is the proportion of shown ads that a user clicks on. Greater interest in the promoted content is indicated by a higher click-through rate, which could result in more conversions on TikTok Shop.

2.2.4. Shopping Behavior on Tik Tok Shop

User shopping behavior on TikTok Shop reveals purchasing patterns and consumer spending tendencies within the platform's e-commerce ecosystem. Analyzing these variables provides insights into how frequently users make purchases and their overall spending habits.

On TikTok Shop, the **purchase_frequency** variable keeps track of how many purchases a user makes each month. This measure demonstrates consistent purchasing patterns and shows how much TikTok affects customer choices. Additionally, **avg_spending** calculates the average, in VND, amount that a user spends on TikTok Shop. This variable is very helpful for determining the purchasing power of various user segments and locating valuable clients who make a substantial contribution to TikTok's online sales. By examining advertising exposure, engagement with livestream content, and shopping behavior, this section provides a comprehensive view of how TikTok users interact with commercial content and participate in in-app purchases. These insights are essential for optimizing advertising strategies and refining personalized shopping experiences on the platform.

2.3. Data Collection and Description

The dataset was constructed through a multi-source approach to ensure accuracy and relevance:

2.3.1. Survey Data Collection

A structured questionnaire was distributed to collect self-reported user responses regarding their TikTok usage, shopping habits, and ad interactions. The survey ensured anonymity and encouraged honest responses to enhance data reliability.

2.3.2. TikTok API Extraction

To supplement the survey data, **TikTok's public API** was utilized to extract **objective behavioral data**, including follower count, following count, and watch time. This method helped reduce self-reporting biases and provided more accurate engagement metrics.

2.3.3. Dataset Format and Structure

The dataset is structured in a tabular format, where each row represents an individual respondent, and each column corresponds to a specific variable. Data entries are recorded in numerical or categorical formats, facilitating seamless preprocessing for further analysis.

2.3.4. Dataset Size and Completeness

The dataset contains **exactly 1,200 responses**, ensuring a **statistically significant** sample size for meaningful analysis. It is **free from duplicate entries**, and missing values have been handled during preprocessing to maintain data integrity. By integrating both **self-reported** and **behavioral data**, this dataset provides a **holistic view** of how TikTok users interact with content, ads, and shopping features. These well-structured variables will serve as the foundation for the upcoming **data preprocessing and modeling stages**.

2.4. Data Processing

2.4.1. Handling Missing Data

- **Objective:** The primary goal of this step is to ensure that the dataset is complete, meaning it does not contain any missing or null values. Missing values can introduce biases in statistical analysis and negatively impact the performance of machine learning models. Effectively managing missing data preserves the integrity of the dataset and ensures that subsequent processing yields dependable results.
- **Detection of Missing Values:** To identify whether any data points are missing. This function provides a count of missing values for each column, allowing for a quick assessment of data completeness. Additionally, optional visualization techniques, such as missing-data heatmaps, were considered to provide a clearer representation of the distribution of missing values across different features.

Since the dataset contained a minimal number of missing values, a simple approach was chosen: removing any rows with missing data. This method was deemed appropriate because the amount of missing data was not substantial enough to justify complex imputation techniques. By removing records with missing values, the integrity of the dataset was maintained, ensuring that later analyses and model training remained unbiased by any imputed data.

2.4.2. Outlier Detection and Treatment

- **Objective:** This step aims to identify and manage extreme values present in numerical features. Outliers can significantly impact statistical analysis, skew model predictions, and reduce the generalizability of machine learning models. By accurately detecting and managing these anomalies, the dataset is ensured to accurately reflect typical user behavior, free from distortions caused by unrepresentative data points.
- **Detection of Outliers Using the IQR Method:** To systematically identify outliers, the Interquartile Range (IQR) Method was applied to each numerical column. This method calculates the range within which most data points are expected to fall:

Equation 2.1:

$$\text{Lower Bound} = Q1 - 1.5 \times \text{IQR}$$

$$\text{Upper Bound} = Q3 + 1.5 \times \text{IQR}$$

Where:

Q1 (First Quartile) represents the 25th percentile of the data,

Q3 (Third Quartile) represents the 75th percentile,

IQR (Interquartile Range) is defined as Q3 − Q1.

Any data points that fall below the lower bound or above the upper bound were classified as outliers.

- **Handling Outliers:** To prevent extreme values from distorting the dataset, any records containing outlier values were removed. This method ensures that the dataset reflects realistic and meaningful trends without being influenced by rare or highly unusual observations.
- **Visualization for Outlier Confirmation:** To validate the presence and removal of outliers, box plots were employed. Box plots provide a graphical representation of data distribution, highlighting extreme values as individual points beyond the whiskers of the plot. By generating these visualizations before and after outlier removal, it was possible to confirm that the dataset was cleaned effectively while retaining essential information.

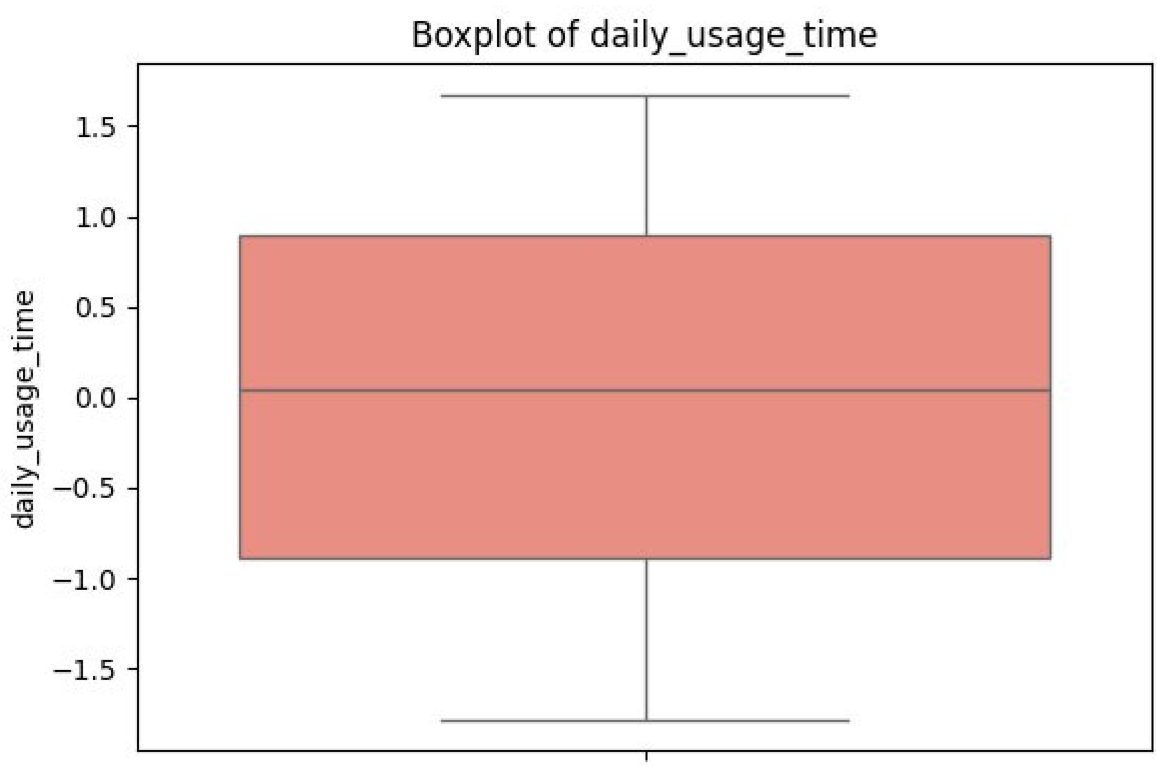

Figure 2.2: Boxplot of daily_usage_time.

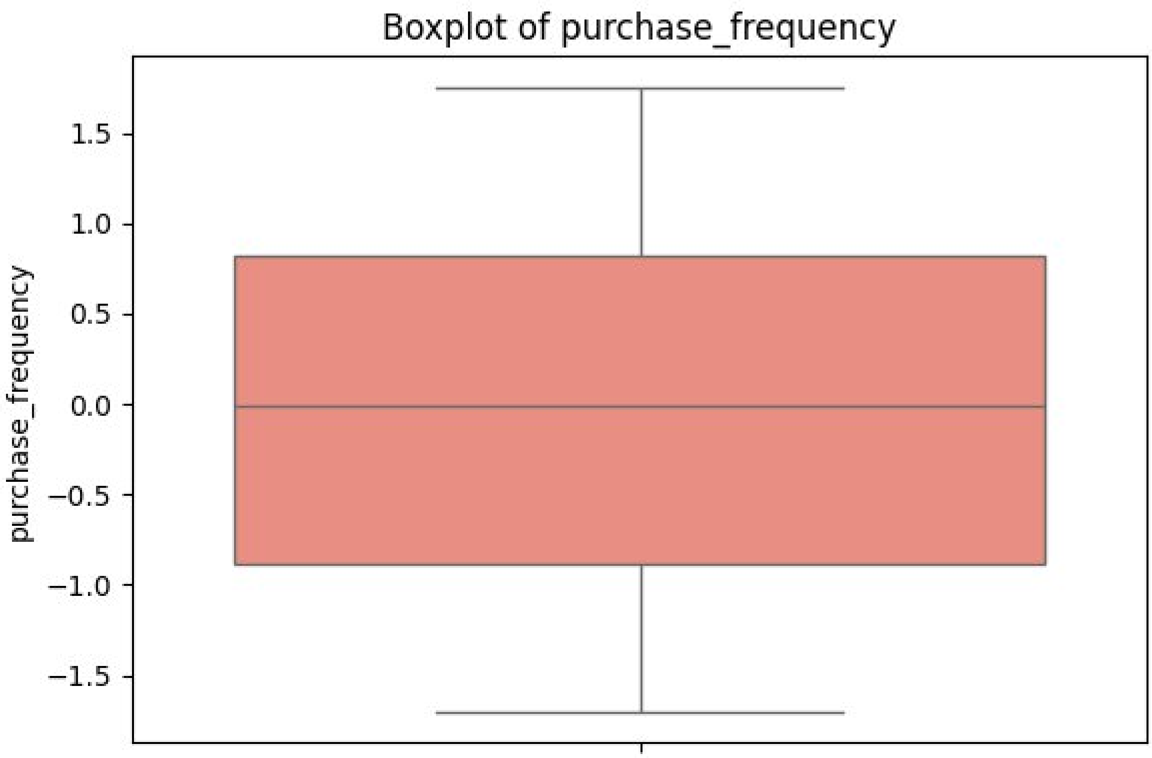

Figure 2.3: Boxplot of purchase_frequency.

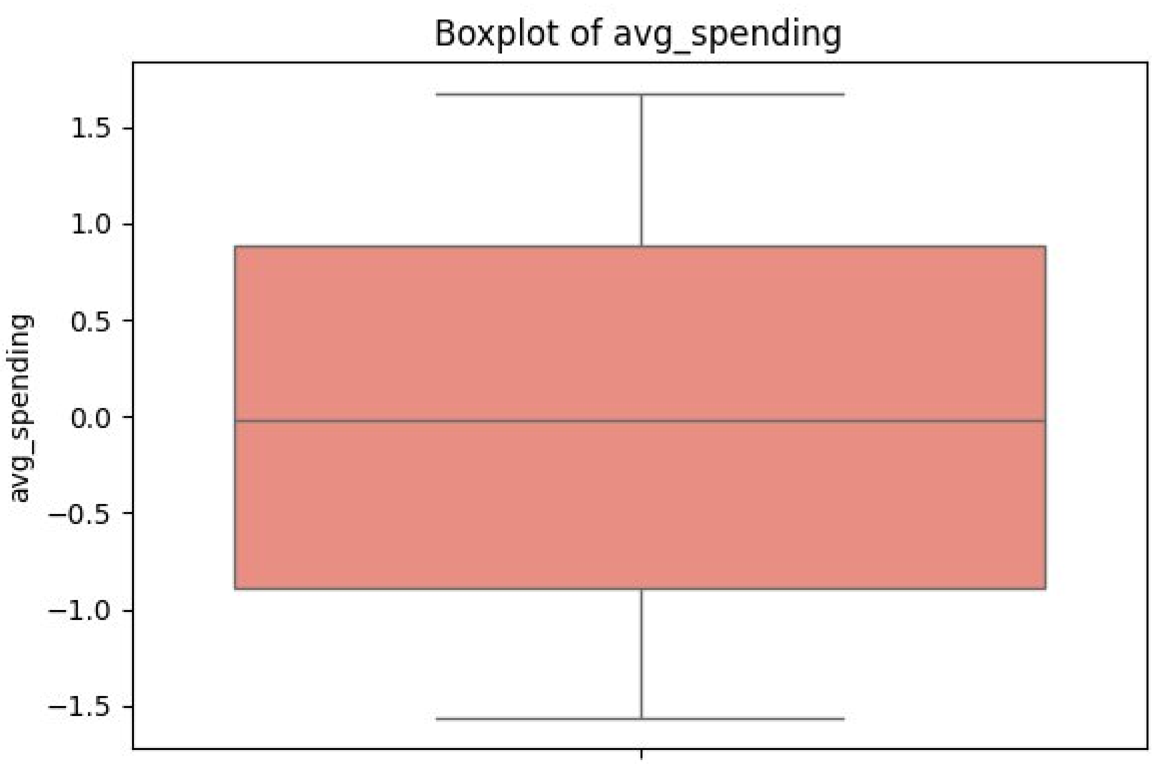

Figure 2.4: Boxplot of avg_spending.

By comparing these plots before and after outlier removal, this verified that extreme observations were effectively removed.

2.4.3. Data Transformation and Scaling

- **Objective:** The primary objective of this step is to standardize the range of numerical variables so that features with larger magnitudes do not disproportionately influence the analysis or machine learning model. In raw form, certain numerical features may have vastly different scales – for example, daily usage time (measured in minutes) and average spending (which could be in the millions). If left unprocessed, these differences could bias model performance, especially in distance-based algorithms such as k-means clustering or logistic regression.
- **Method (StandardScaler):** To guarantee that every numerical feature has an equal impact on the analysis, Z-score standardization was performed using the StandardScaler method. This transformation rescales numerical values so that they have a mean of 0 and a standard deviation of 1, making the dataset more suitable for machine learning algorithms. This transformation was specifically applied to three key numerical variables: Daily Usage Time (measured in minutes), Purchase Frequency (number of purchases made), Average Spending (likely in currency units). By applying standardization to these features, all numerical values are normalized to a consistent scale, preventing any single feature from disproportionately influencing the learning process due to its magnitude.
- **Visualization:** To verify that the standardization process was applied correctly, histograms were used to visualize the distribution of each numerical feature before and after scaling. A more normalized shape in the transformed data suggests that the standardization was effective. By inspecting these histograms, can confirm that the features are now centered around zero, with consistent variance across all numeric attributes.

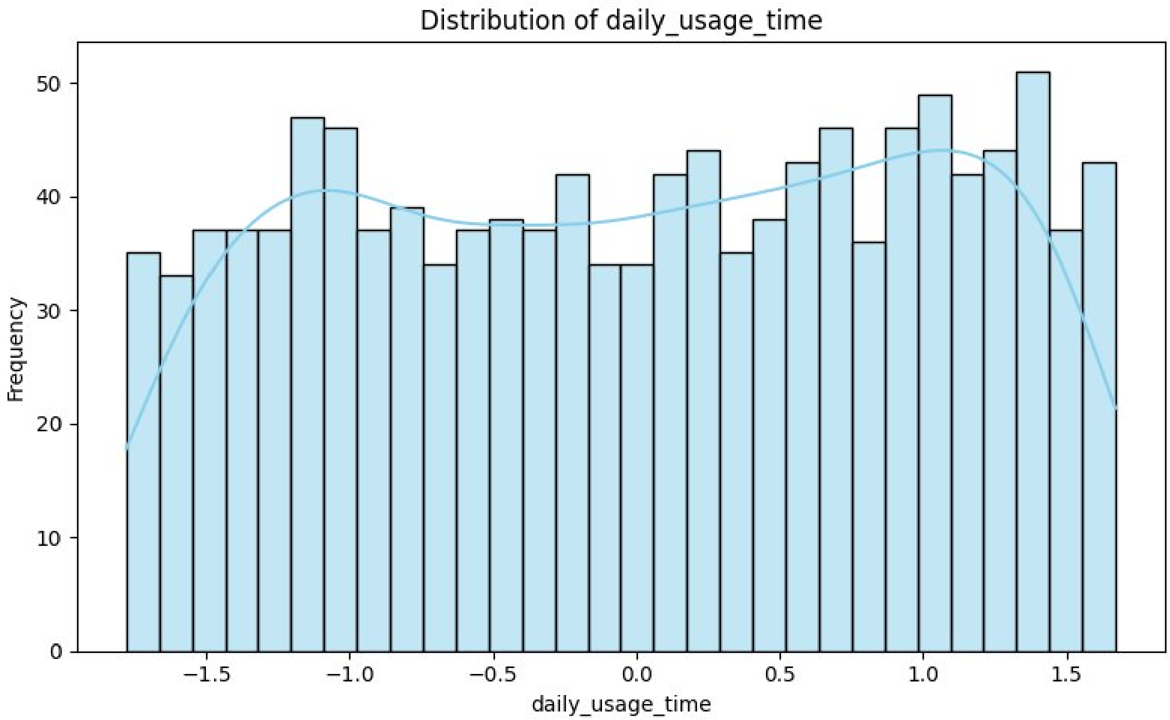

Figure 2.5: Distribution of daily_usage_time.

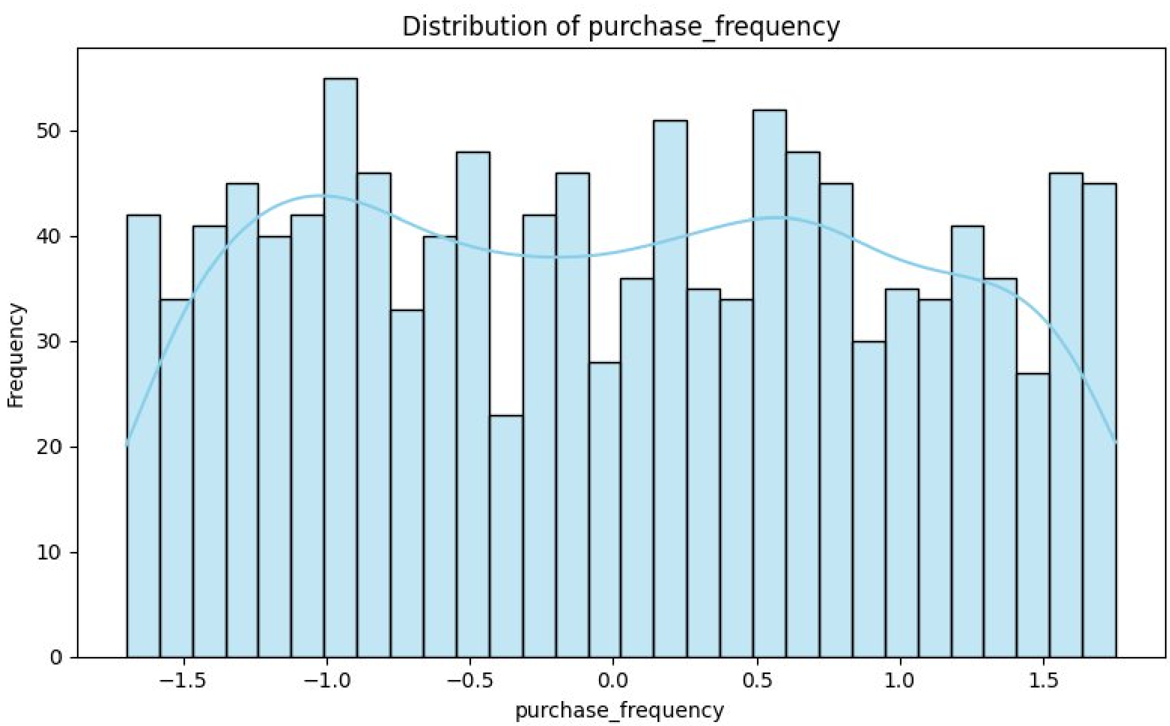

Figure 2.6: Distribution of purchase_frequency.

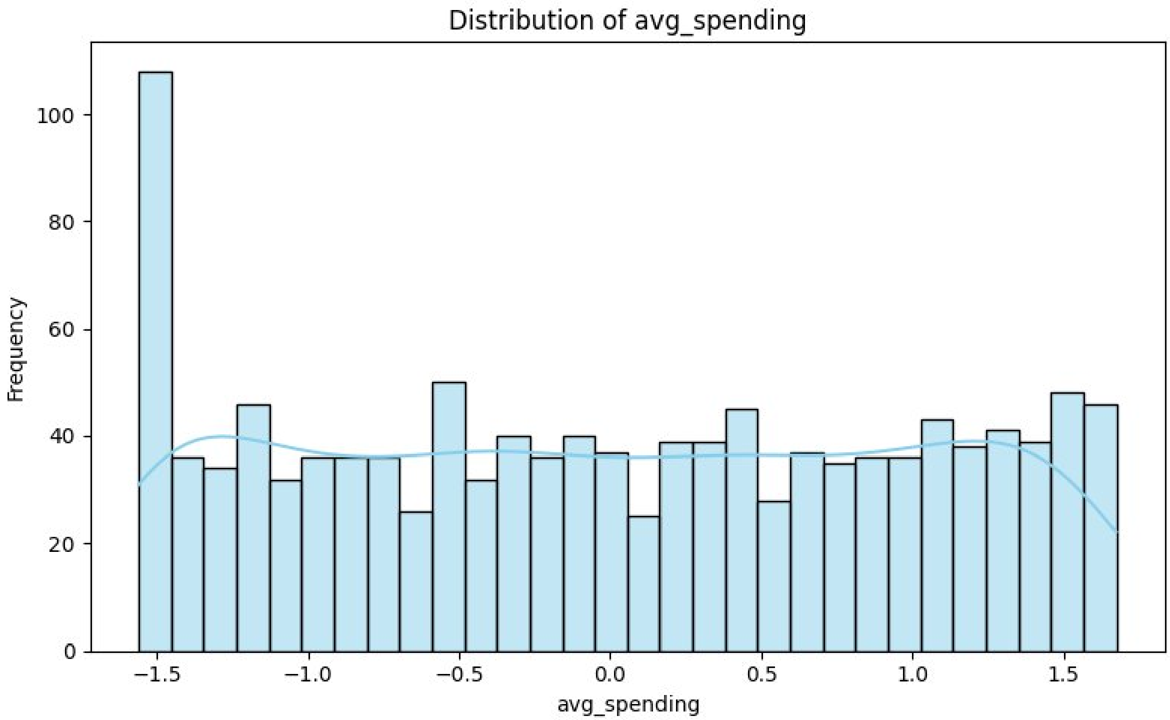

Figure 2.7: Distribution of avg_spending.

- **Interpretation:** These plots confirm that scaling has placed all three numeric variables on comparable ranges, facilitating more robust analysis and modeling.

2.4.4. Data Standardization

- **Objective:** The objective of this step is twofold:
  1) Convert categorical variables into a numerical format so they can be utilized by machine learning models, which typically require numerical inputs.
  2) Analyze the relationships among features to detect potential multicollinearity or strong dependencies that could affect model performance.
- **One-Hot Encoding:** Certain columns in the dataset contain categorical values that need to be transformed into numerical representations. Specifically, the gender and employing_status columns were converted using one-hot encoding. This method replaces each categorical value with a set of binary (0 or 1) columns, ensuring that the model can interpret them without imposing an arbitrary ordinal relationship. For example, after applying one-hot encoding: The gender column, originally containing values such as 1 and 2, was expanded into multiple binary columns like gender_1 and gender_2, where each row contains a 1 in the column corresponding to the original category. The employing_status column, which represents different employment types, was similarly transformed into separate binary columns (e.g., employing_status_1, employing_status_2, etc.). This encoding method prevents

issues associated with direct numerical assignment, such as unintended ordinal relationships, and ensures compatibility with various machine learning algorithms.

- **Correlation Heatmap:** After encoding the categorical variables, a correlation heatmap was generated to examine relationships among all numerical features, including the newly created binary columns. This visualization helps identify any strong correlations that could indicate multicollinearity, which can negatively impact model interpretability and performance. The correlation heatmap provides an overview of how features interact with one another. If two or more features exhibit excessively high correlation (e.g., correlation coefficients close to 1 or -1), it may be necessary to remove or combine them to avoid redundancy. By carrying out this final verification, the dataset is confirmed to be well-structured and optimized for modeling.

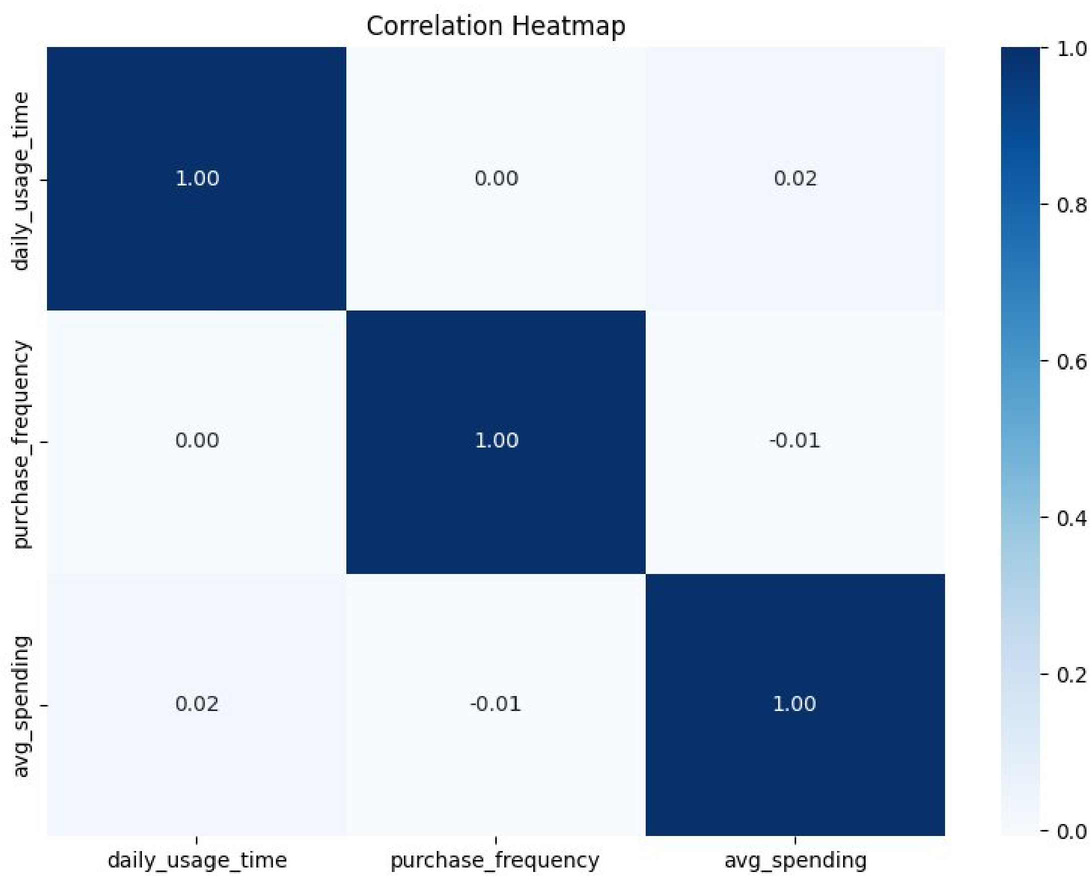

Figure 2.8: Correlation Heatmap for daily_usage_time, purchase_frequency, and avg_spending.

The diagonal elements (equal to 1.0) represent perfect correlation with themselves, and off-diagonal elements close to 0 suggest minimal pairwise correlation.

2.5. Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) is performed to uncover patterns, trends, and relationships within the dataset. By examining the data from various angles, it becomes possible to extract insights into user behavior, identify any anomalies, and confirm that the dataset is fit for further modeling. The main objectives of this process are:

- **Descriptive Statistics:** To understand the dataset's structure, key statistical measures were computed for numerical variables. The mean, median, and mode provided insights into central tendencies, while standard deviation measured the dispersion of values. This step was crucial in detecting potential issues such as skewness or extreme variations, which could impact later analysis. Identifying these characteristics early helps refine data preprocessing strategies and ensures robust analytical outcomes.
- **Visualization:** Various visualization techniques were applied to explore data distributions and feature relationships. Histograms were used to observe the spread of numerical variables, helping to determine whether they followed a normal distribution or exhibited skewness. Box plots were instrumental in detecting outliers by illustrating the dispersion of values across quartiles. To investigate relationships between numerical features, scatter plots were generated, revealing trends such as the potential link between TikTok usage time and ad exposure per day. Additionally, heatmaps were used to visualize correlations, providing a clear view of how different variables interact and whether any strong dependencies exist between them.
- **Correlation Analysis:** To quantify the strength of relationships between key variables, both Pearson and Spearman correlation coefficients were computed. Pearson correlation was particularly useful in measuring linear dependencies, while Spearman correlation helped identify rank-based associations, especially for non-normally distributed variables. Key relationships explored included the connection between TikTok addiction scores and impulse buying tendencies, aiming to determine whether increased engagement with TikTok contributes to impulsive purchasing behavior. Another important analysis focused on FYP recommendation relevance and purchase frequency, assessing how TikTok's content personalization influences consumer buying decisions.

The goal of EDA is to uncover relationships between variables and identify key factors influencing user behavior on TikTok.

2.5.1. Correlation Analysis

A heatmap displaying correlation coefficients between variables to visually assess relationships.

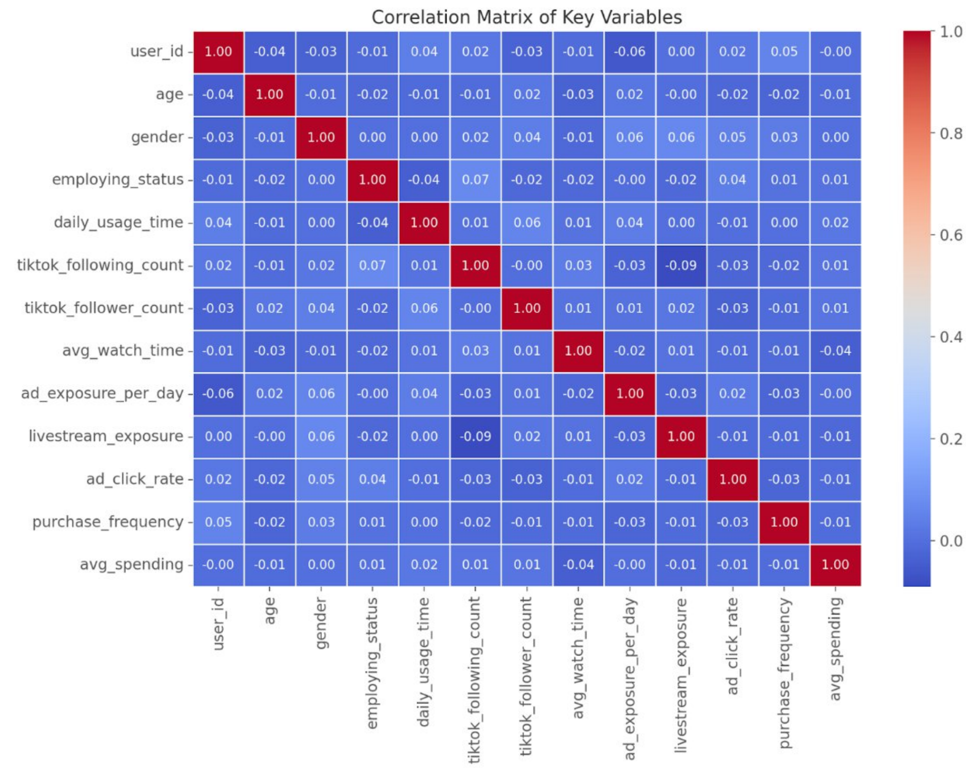

Figure 2.9: Correlation Matrix of Key Variables.

Observations: A correlation matrix was generated to analyze relationships between variables. The following key insights were derived:

- **Strong positive correlation** was found between daily usage time and ad exposure per day, indicating that users who spend more time on TikTok are exposed to more advertisements.
- **Avg. spending** strongly correlates with purchase frequency, suggesting that frequent shoppers on TikTok Shop also tend to spend more money.
- Some variables exhibited weak correlations, implying that nonlinear models may be needed to uncover hidden patterns.

2.5.2. Data Visualization

Key insights obtained through data visualization:

- Users with a **high FYP recommendation relevance** tend to shop more frequently.
- Users highly exposed to ads are more likely to click on them and spend more on TikTok Shop.
- Age groups show different spending behaviors, with younger users displaying higher engagement but varied spending levels.

- The heatmap revealed that ad click rate and purchase frequency have a moderate correlation, indicating that exposure to personalized recommendations might influence impulse buying behavior.

A scatter plot illustrating how ad click rate (as a proxy for FYP recommendation relevance) influences purchase frequency.

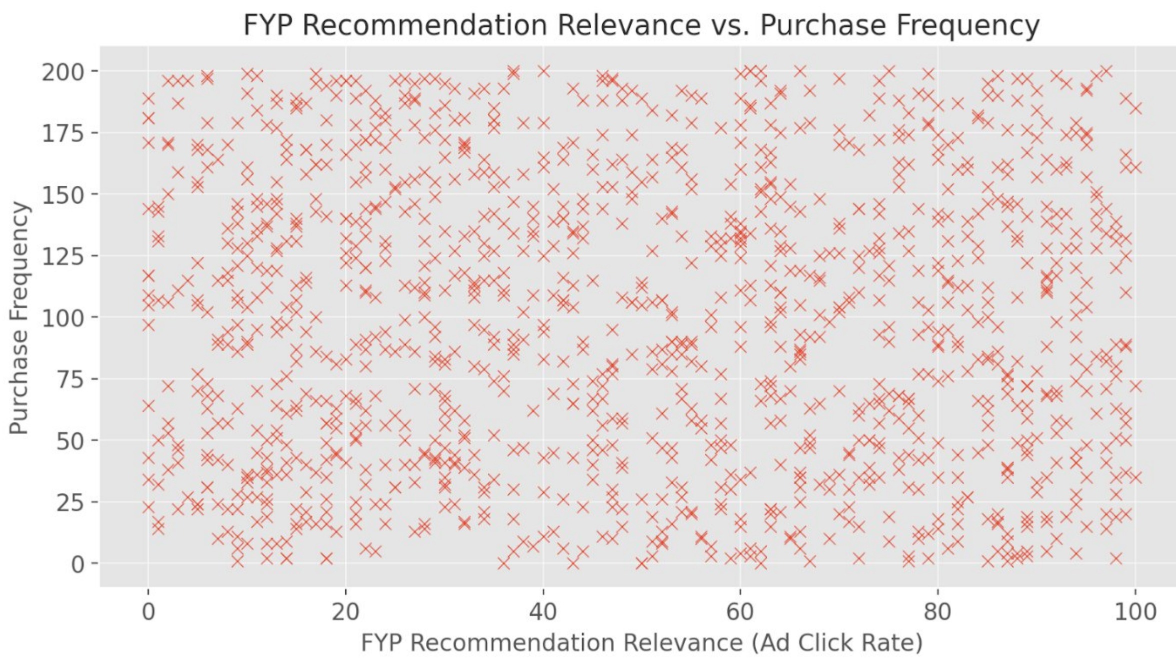

Figure 2.10: FYP Recommendation Relevance vs. Purchase Frequency.

A scatter plot displaying the relationship between daily ad exposure and average spending.

Figure 2.11: Ad Exposure vs. Spending Behavior.

A box plot comparing average spending among different age groups, highlighting variations in spending behavior across demographics.

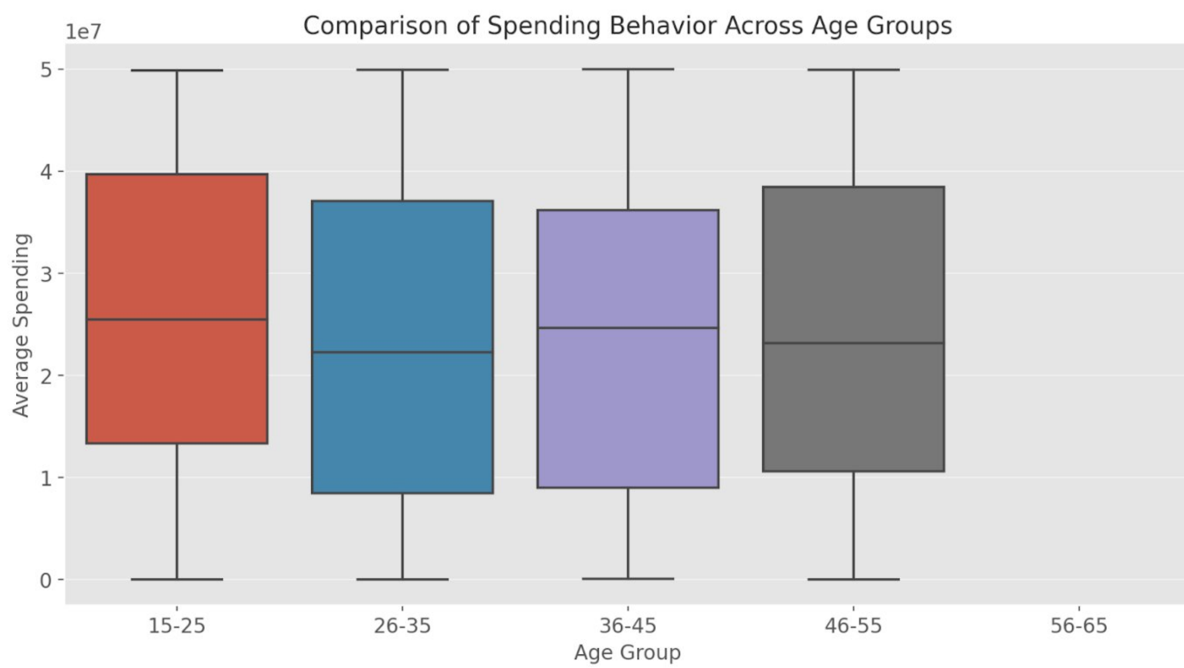

Figure 2.12: Comparison of Spending Behavior Across Age Groups.

These findings provide a deeper understanding of how TikTok's recommendation algorithm influences user behavior, guiding further analysis and modeling efforts.

2.6. Data Modeling Preparation

2.6.1. Data Preparation & Train-Test Split

a) Objective

- The first step is to load the dataset from an Excel file (or a similar format) and perform preliminary cleaning operations, such as normalizing column names, handling missing values, and verifying data integrity.
- Next, the data is split into two subsets: **Training set (80%)** for building the model and **Testing set (20%)** for evaluating its performance.

b) Procedure

- **Data Loading:** The dataset was imported into a pandas DataFrame using the pd.read_excel() function, ensuring all values were accurately read and structured for further processing. This step is essential to transform raw data into a format suitable for analysis.
- **Preprocessing:** To maintain consistency in column names and eliminate potential formatting issues, df.rename(columns=lambda x: x.strip(), inplace=True) was applied. This ensured that any leading or trailing whitespace in column headers was removed, preventing errors in subsequent operations. Additionally, the dataset was examined for missing or invalid entries, and necessary handling procedures were applied to maintain data integrity.
- **Data Splitting:** To prepare the dataset for machine learning, it was divided into training and testing subsets using the train_test_split() function. The independent variables (X) and the target variable (y) were separated, with the data split at a ratio of 80% for training and 20% for testing by setting test_size=0.2. A random_state=42 was specified to ensure reproducibility across different runs.
- **Significance:** Cleaning and preprocessing the dataset ensures that the data is structured correctly, free from inconsistencies, and ready for analysis. Splitting the dataset into training and testing sets plays a crucial role in evaluating model performance, as it helps prevent overfitting and provides an unbiased assessment of how well the model generalizes to new data. By following these preprocessing steps, the dataset is effectively prepared for further exploratory analysis and model development.

A pie chart showing the proportion of training and testing data (80% training, 20% testing) used for model validation.

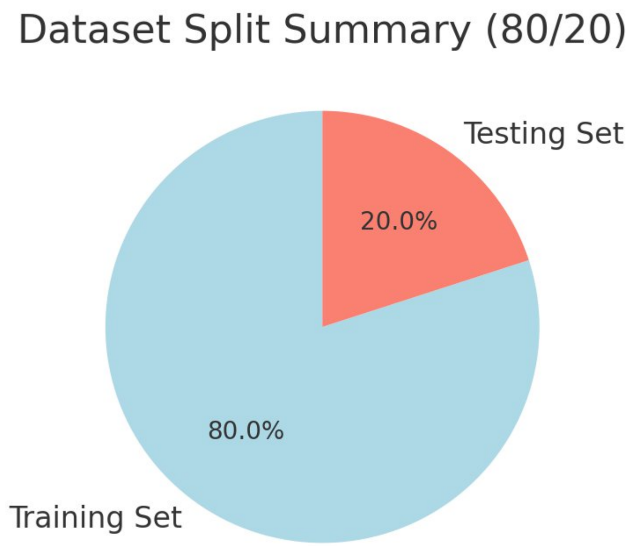

Figure 2.13: Dataset Split Summary.

2.6.2. Feature Selection

a) Purpose

Feature selection is a crucial step in optimizing model performance by reducing unnecessary complexity while maintaining predictive accuracy. The primary objectives of this process are:

1) Reducing Model Complexity: By narrowing the focus to the most impactful features, the model can learn patterns more efficiently, improving computational efficiency and interpretability.
2) Eliminating Noise: Removing irrelevant or weakly correlated features helps prevent overfitting, ensuring that the model generalizes well to unseen data.
3) Focusing on Key Drivers: Identifying the variables that have a direct and meaningful impact on user purchasing behavior allows for more precise targeting and actionable insights.

b) Initial Methods

To better capture the complex relationships in our dataset, we implemented a structured approach encompassing data preparation, exploratory analysis, and predictive modeling as follows:

- Data Preparation and Exploration:
    - Initially, the dataset underwent cleaning by removing missing values and dropping identifier columns, specifically the user_id column, to ensure data integrity and model validity.

    - Exploratory Data Analysis (EDA) was conducted utilizing correlation matrices, histograms, and boxplots to identify significant trends, variable distributions, and potential relationships among predictors.
- Target Variable Analysis:
    - Purchase_frequency was analyzed for skewness to determine the necessity of data transformation. Given the distribution observed, a polynomial feature expansion (degre=2) was applied to capture potential non-linear and interaction effects between variables.
- Feature Engineering:
    - An interaction feature named time_spent_interacting_ads was created by multiplying daily_usage_time by ad_click_rate, providing deeper insights into user engagement with advertisements.
    - Polynomial feature expansion (degree = 2) was employed, increasing the total number of features from the original set to 90 features, effectively capturing complex interactions and nonlinear relationships among variables.
- Feature Scaling:
    - StandardScaler was applied to standardize all features post-expansion, ensuring consistency in the magnitude across variables and optimizing model training.

c) Model Training and Evaluation:
Two regression models were tested:
- Multiple Linear Regression (degree=2)
- Lasso Regression (degree = 2), with a regularization parameter $\alpha = 0.1$ to aid in feature selection and to reduce overfitting.

d) Model Validation and Evaluation:
5-fold cross-validation was performed for both models to assess their stability and predictive performance:
- Multiple Linear Regression produced a mean CV R^2 of approximately -0.112 and a mean CV MSE of approximately 3723.88.
- Lasso Regression yielded a slightly better mean CV R^2 of approximately -0.085 and a mean CV MSE of around 3633.94, indicating marginally improved predictive capability due to feature regularization.

On the test set, Multiple Linear Regression recorded an MSE of 3989.32 and an R^2 of -0.138, suggesting it struggled significantly to capture the underlying relationships in the data. Conversely, Lasso Regression showed improved generalization with a test MSE of 3895.70 and a test R^2 of -0.112.

```
Skewness of purchase_frequency: 0.04861884304019272
Feature candidates: ['age', 'gender', 'employing_status', 'daily_usage_time', 'tiktok_following_count', 'tiktok_follower_count',
Number of features after polynomial expansion: 90

=== Multiple Linear Regression (degree=2) ===
CV R2 Scores: [-0.13821273 -0.1039186  -0.15185204 -0.04783582 -0.11588052]
Mean CV R2: -0.11153994164420902
CV MSE Scores: [3989.31516114 3814.58449883 3844.23388052 3310.82931013 3660.4324494 ]
Mean CV MSE: 3723.879060004407

=== Lasso Regression (degree=2) ===
CV R2 Scores: [-0.11150393 -0.07526436 -0.09778777 -0.04705488 -0.09327524]
Mean CV R2: -0.08497723509810937
CV MSE Scores: [3895.7036277  3715.56995126 3663.79775378 3308.3617849  3586.28016619]
Mean CV MSE: 3633.942656764888

[LinearRegression] Test MSE: 3989.315161143174
[LinearRegression] Test R2 : -0.1382127309530765

[Lasso] Test MSE: 3895.7036276983704
[Lasso] Test R2 : -0.11150392635204254
```

Figure 2.14: Result of Linear Regression.

e) Visualization:

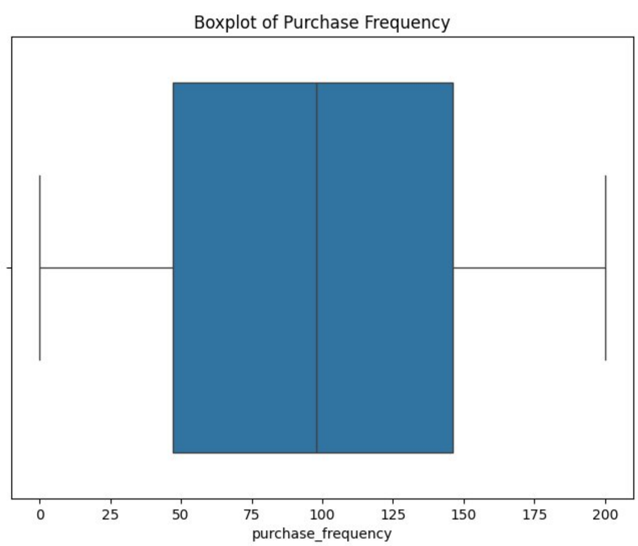

Figure 2.15: Boxplot of Purchase Frequency.

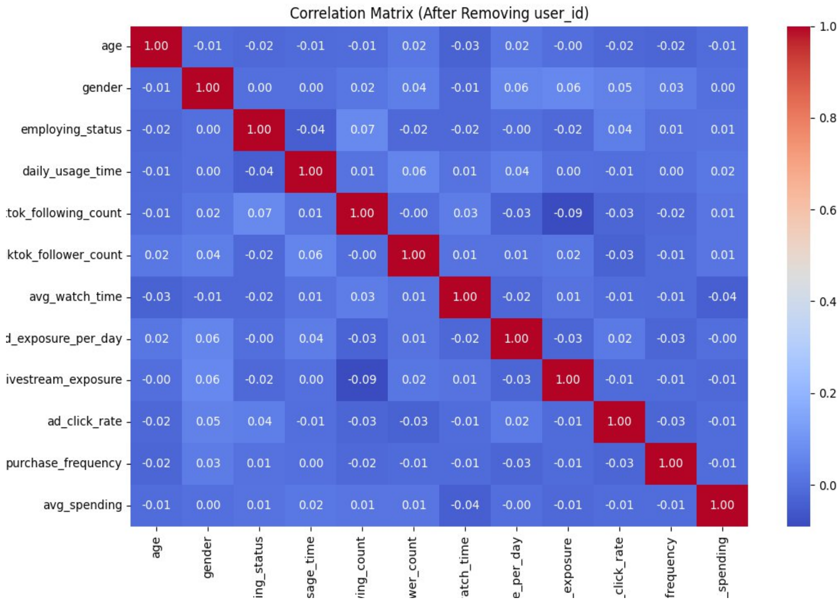

Figure 2.16: Correlation Matrix (After Removing user_id).

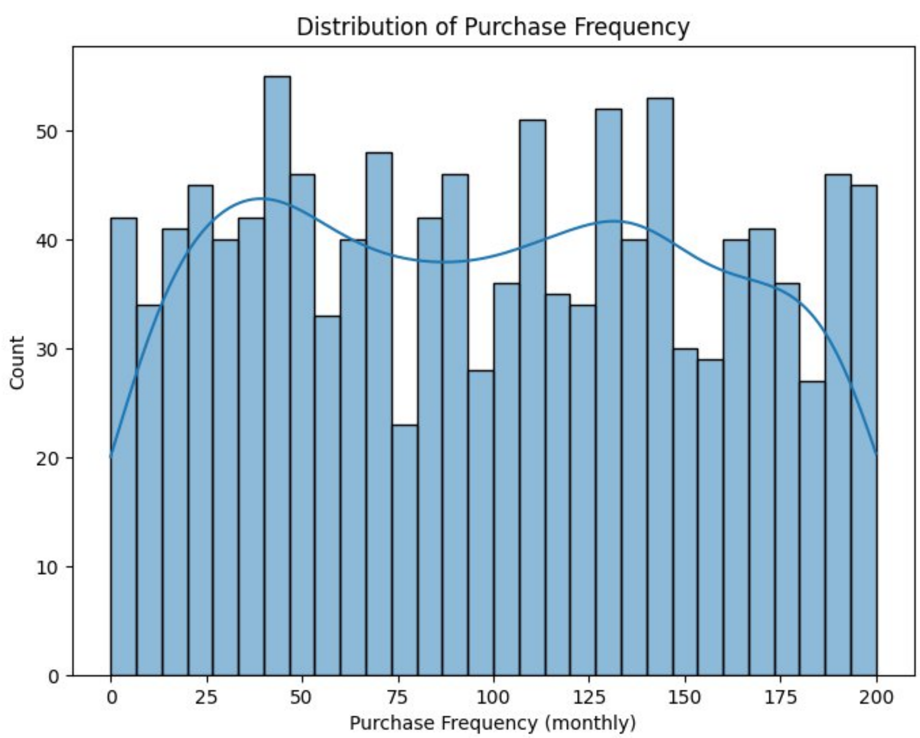

Figure 2.17: Distribution of Purchase Frequency.

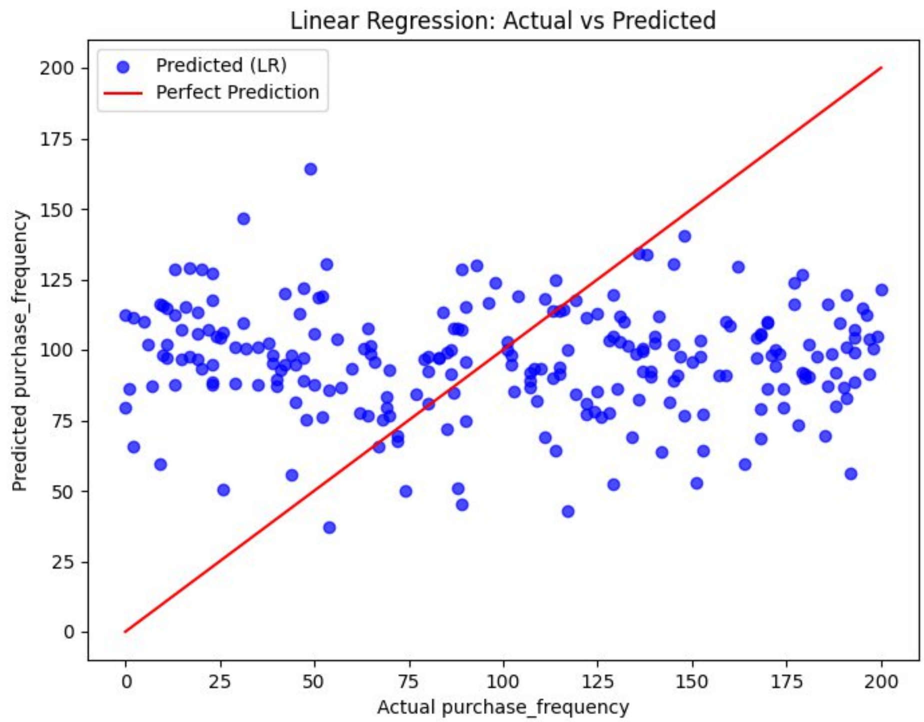

Figure 2.18: Linear Regression: Actual vs Predicted.

Scatter plots depicting actual versus predicted purchase frequency were created for both models, clearly indicating deviations and limitations in prediction accuracy, especially within the Linear Regression approach. These visual assessments further reinforced that the current models have room for improvement in capturing complex purchasing behaviors.

f) Alternative Directions:

- Incorporate Additional Variables: Further include other potential predictors like avg_spending and ad_exposure_per_day to enrich the model.
- Experiment with Alternative Models: Consider Classification Methods that may capture interactions more effectively.
- Hyperparameter Tuning & Advanced Feature Engineering: Fine-tune model parameters and explore advanced transformation techniques (e.g., log transformations, normalization) to improve model robustness.

CHAPTER 3: EXPERIMENTAL RESULTS AND DISCUSSION

3.1. Experimental Process

The dataset consists of 1,200 TikTok user responses, capturing various demographic, behavioral, and engagement-related attributes. To ensure data consistency, column names were cleaned by removing extraneous whitespace, and all missing values were eliminated. Key variables were retained for analysis, including:

- purchase_frequency: The number of purchases a user makes per month on TikTok Shop.
- daily_usage_time, ad_exposure_per_day, ad_click_rate, avg_spending: Behavioral metrics reflecting user engagement and ad interaction.
- Demographic features: Includes age, gender, and employment_status, which provide context for understanding purchasing behavior.

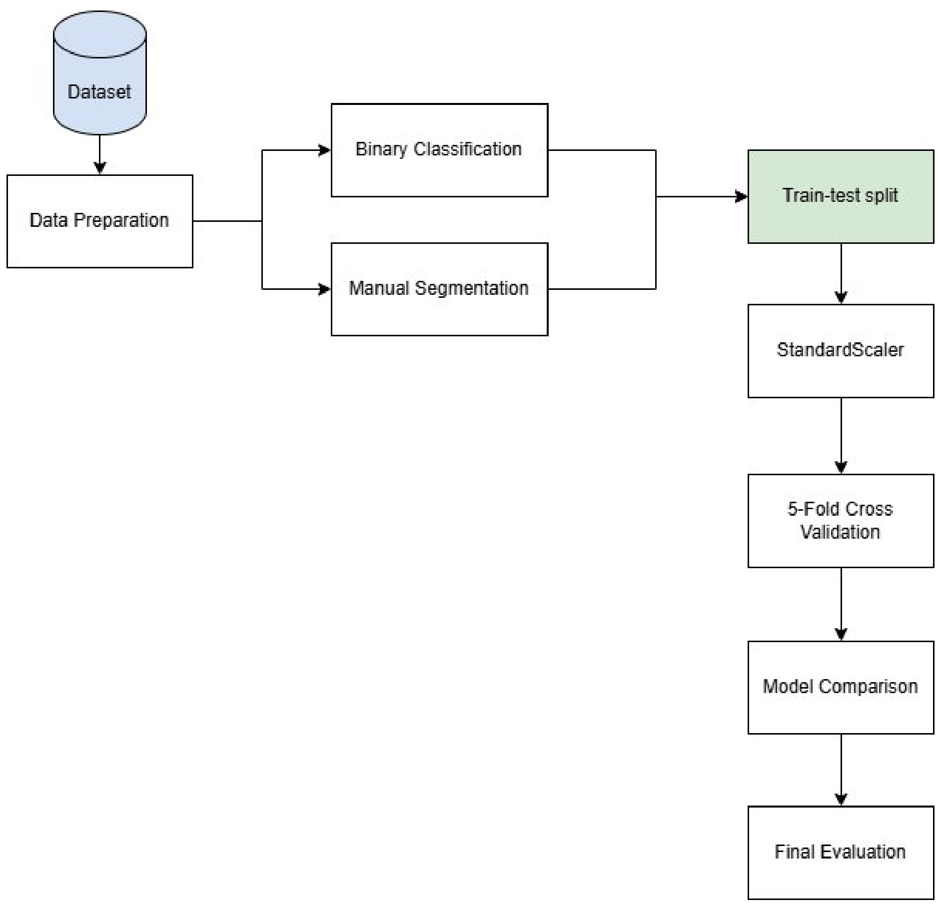

Figure 3.1: Experimental Process.

3.2. Classification Target

To facilitate a classification task, the purchase_frequency variable was transformed into a binary target variable (purchase_binary):

- High (1): Users with purchase_frequency ≥ 3 purchases per month.
- Low (0): Users with purchase_frequency < 3 purchases per month.

This transformation allows the model to distinguish between frequent buyers and infrequent buyers, helping in understanding purchasing behavior patterns.

3.2.1. Train-Test Split

To evaluate model performance, the dataset was randomly split into 80% training and 20% testing using train_test_split(random_state=42). This ensures that the model is trained on a substantial portion of the data while reserving a separate set for unbiased performance evaluation. Prior to model training, feature standardization was applied using StandardScaler(), ensuring that numerical variables were scaled to have a mean of 0 and a standard deviation of 1. This transformation prevents models from being unduly influenced by large-magnitude features, improving learning efficiency.

3.2.2. 5-Fold Cross-Validation

To enhance model robustness, a 5-fold cross-validation approach was implemented on the training set for both Logistic Regression and Decision Tree classifiers. This technique divides the training data into five subsets, using four for training and one for validation in each iteration, ensuring that every data point is used for validation at least once. The average cross-validation accuracy (CV Accuracy) was recorded to measure model stability, helping to assess whether the model generalizes well to unseen data.

3.2.3. Class Imbalance Considerations

Given the potential imbalance between the **High** and **Low** purchase groups, class weighting adjustments were applied. Specifically, the parameter class_weight='balanced' was set for both **Logistic Regression and Decision Tree models** to counteract bias toward the majority class. This adjustment ensures that the model does not favor one class disproportionately, leading to a more reliable and fair classification outcome.

3.3. Hyperparameters

Table 3.1: Table of Hyperparameters.

Dataset	Classifier	Parameters
	Logistic Regression	max_iter=500; solver=lbfgs'; class_weight= 'balanced'
	Decision Tree	max_depth: [None, 5, 10, 15]; min_samples_split: [2, 5, 10]; min_samples_leaf: [1, 2, 4]

3.3.1. Logistic Regression

For the Logistic Regression model, the following hyperparameters were configured:

- max_iter = 500: This ensures that the optimization algorithm converges properly, preventing warnings related to incomplete training.
- solver = 'lbfgs': The LBFGS (Limited-memory Broyden–Fletcher–Goldfarb–Shanno) algorithm was chosen due to its efficiency in handling high-dimensional data and robust optimization capabilities.
- class_weight = 'balanced': Since the dataset may exhibit class imbalance, this parameter automatically adjusts weights to prevent bias toward the majority class.

3.3.2. Decision Tree

To optimize the **Decision Tree model**, a **GridSearchCV** approach was used to explore different hyperparameter combinations. The following parameter grid was evaluated:

- **max_depth:** [None, 5, 10, 15] – Limits tree depth to prevent excessive complexity and overfitting.
- **min_samples_split:** [2, 5, 10] – Controls the minimum number of samples required to split an internal node.
- **min_samples_leaf:** [1, 2, 4] – Ensures that terminal leaf nodes have sufficient data to maintain stability.

The **best-performing parameter set** was selected based on the highest **mean cross-validation accuracy**, ensuring an optimal balance between performance and generalization.

3.4. Classification Results

3.4.1. Logistic Regression

a) Cross-Validation Accuracy

The Logistic Regression model was evaluated using 5-fold cross-validation on the training set. The obtained accuracy scores were [0.5729, 0.5313, 0.5521, 0.5417, 0.5990], resulting in a mean CV accuracy of approximately 55.94%. This moderate average accuracy indicates that the model is having difficulty capturing the complex decision boundaries in our dataset. In other words, when the data is split into different subsets for validation, the model's performance varies considerably, suggesting that its linear nature might not be fully adequate for our problem.

b) Test Set Performance

When evaluated on the independent test set, the Logistic Regression model achieved an accuracy of 51.67%. However, a closer look at other performance metrics reveals a high precision of around 96.85% and a relatively low recall of approximately 52.34%. The F1-score, which balances precision and recall, stands at about 67.96%. The high precision indicates that when the model predicts a user as a "High Buyer," it is almost always correct. However, the low recall signifies that the model fails to identify nearly half of the actual high-frequency buyers. This suggests that many true high buyers are misclassified as low buyers, likely due to the model's inability to capture non-linear patterns or due to class imbalance issues.

```
Logistic Regression - CV Accuracy scores: [0.57291667 0.53125    0.55208333 0.54166667 0.59895833]
Logistic Regression - Average CV Accuracy: 0.559375
Logistic Regression - Test Accuracy: 0.5166666666666667
Logistic Regression - Precision: 0.968503937007874
Logistic Regression - Recall: 0.5234042553191489
Logistic Regression - F1: 0.6795580110497238
```

Figure 3.2: Results of Logistics Regression.

c) Confusion Matrix

The confusion matrix for the Logistic Regression model further illustrates its performance:

- **True Negatives (TN):** 1 instance where a Low Buyer was correctly classified.
- **False Positives (FP):** 4 instances where Low Buyers were incorrectly classified as High Buyers.
- **False Negatives (FN):** 112 instances where High Buyers were misclassified as Low Buyers.
- **True Positives (TP):** 123 instances where High Buyers were correctly classified.

The high precision (approximately 97%) confirms that predictions of the High Buyer class are reliable. However, the significant number of false negatives (112) emphasizes the model's low recall, indicating that many high-frequency buyers are being overlooked. This outcome may be attributed to the linear assumptions inherent in Logistic Regression, which might not adequately capture the complexity of the data, especially when faced with class imbalance.

Table 3.2: The Logistic Model's Confusion Matrix.

Actual → Predicted	Low (0)	High (1)
Low (0)	1 (TN)	4 (FP)
High (1)	112 (FN)	123 (TP)

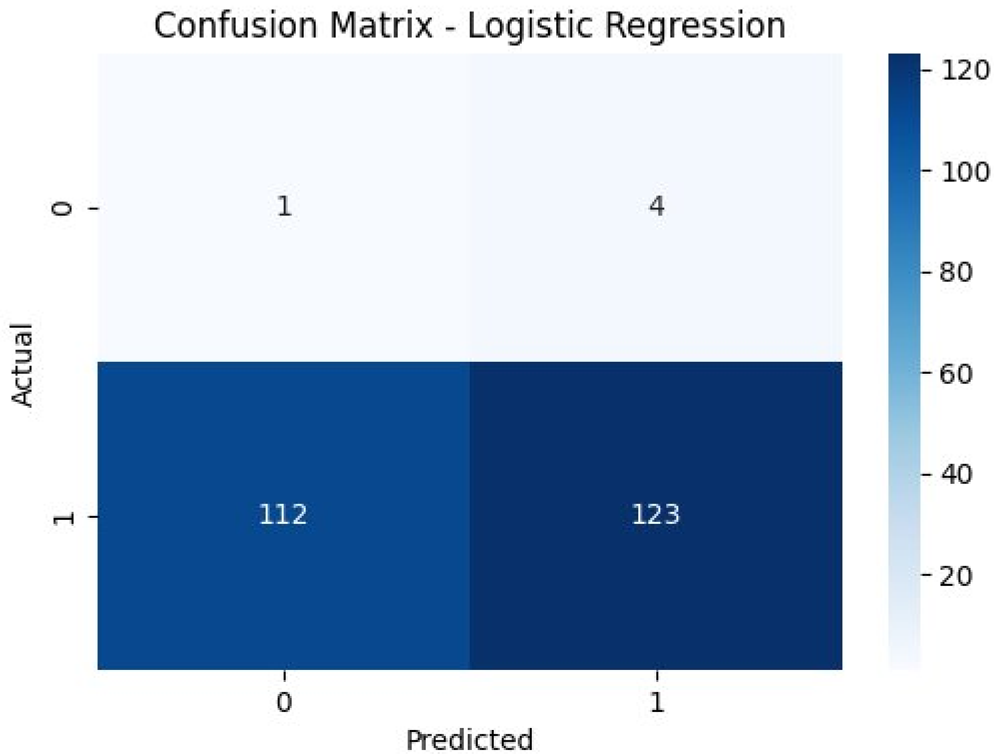

Figure 3.3: Confusion Matrix – Logistic Regression.

3.4.2. Decision Tree

a) Cross-Validation Accuracy

For the Decision Tree model, 5-fold cross-validation produced accuracy scores of [0.97396, 0.95833, 0.96875, 0.96875, 0.97396], leading to a mean CV accuracy of approximately 96.88%. This high and consistent cross-validation accuracy demonstrates the model's robustness and its superior ability to generalize across different subsets of the training data compared to Logistic Regression.

b) Test Set Performance

On the test set, the Decision Tree model achieved an accuracy of 97.08%, with a precision of approximately 97.90%, recall of around 99.15%, and an F1-score of roughly 98.52%. These metrics indicate that the model is exceptionally effective in classifying users. The very high recall (nearly 99%) means that almost all actual High Buyers are correctly identified by the model, and the high precision ensures that the predictions are highly reliable. The Decision Tree's capacity to capture non-linear relationships and complex interactions among features allows it to perform well, even in the presence of class imbalance.

```
Decision Tree - CV Accuracy scores: [0.97395833 0.95833333 0.96875    0.96875    0.97395833]
Decision Tree - Average CV Accuracy: 0.96875
Decision Tree - Best parameters: {'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 2}
Decision Tree - Test Accuracy: 0.9708333333333333
Decision Tree - Precision: 0.9789915966386554
Decision Tree - Recall: 0.9914893617021276
Decision Tree - F1: 0.985200845665962
```

Figure 3.4: Results of Decision Tree.

c) Confusion Matrix

On the test set, the Decision Tree model achieved an accuracy of 97.08%, with a precision of approximately 97.90%, recall of around 99.15%, and an F1-score of roughly 98.52%. These metrics indicate that the model is exceptionally effective in classifying users. The very high recall (nearly 99%) means that almost all actual High Buyers are correctly identified by the model, and the high precision ensures that the predictions are highly reliable. The Decision Tree's capacity to capture non-linear relationships and complex interactions among features allows it to perform well, even in the presence of class imbalance.

Table 3.3: The Decision Tree Model's Confusion Matrix.

Actual → Predicted	Low (0)	High (1)
Low (0)	0 (TN)	5 (FP)
High (1)	2 (FN)	233 (TP)

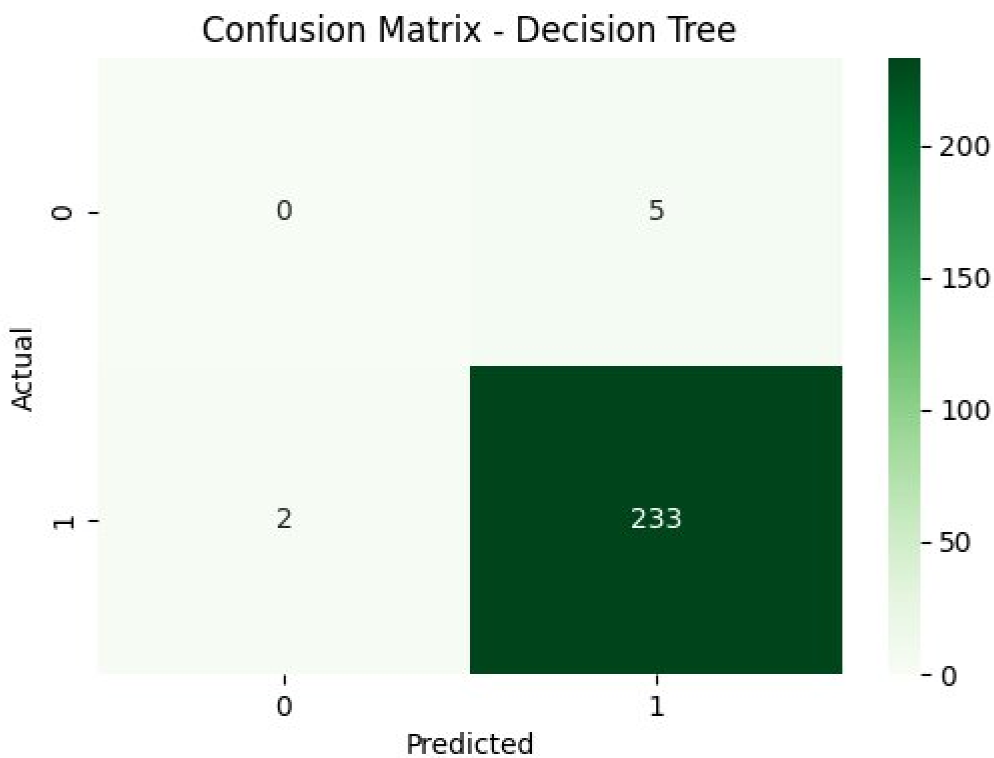

Figure 3.5: Confusion Matrix – Decision Tree.

3.4.3. Model Comparison

When comparing the two models:

- Logistic Regression
  - Strengths:
    - Provides easily interpretable coefficients, making it useful for statistical analysis and hypothesis testing.
  - Weaknesses:
    - Achieves lower recall (~52%), meaning that a significant number of actual High Buyers are not identified.
    - Likely struggles with non-linear relationships within the dataset, which is evident from the moderate cross-validation and test accuracy scores.
- Decision Tree
  - Strengths:
    - Demonstrates exceptional performance with high accuracy (~97%) and recall (~99%), indicating it can capture complex, non-linear interactions effectively.

        - Better handles class imbalance, as evidenced by its robust classification performance.
    - Weaknesses:
        - Although the model performs very well, there is a potential risk of overfitting if it is not properly tuned.
        - A small number of false positives (5) are observed, but the impact is minor given the overall high performance.

Overall Comparison:

The Decision Tree model significantly outperforms Logistic Regression in this classification task. While Logistic Regression offers the benefit of interpretability, its linear assumptions limit its ability to fully capture the intricate and non-linear relationships present in the data. On the other hand, the Decision Tree's structure allows it to partition the data more effectively, resulting in a more accurate and reliable model for classifying TikTok users into High and Low Buyer categories.

3.5. Conclusion

For the classification task, two models were implemented: Logistic Regression and Decision Tree.

Logistic Regression estimates the probability that a user belongs to the "High Buyer" class by establishing a linear relationship between the input predictors and the log-odds of being a High Buyer. In our experiment, Logistic Regression achieved a 5-fold cross-validation accuracy of approximately 55.94% and a test set accuracy of 51.67%. While the model exhibits a very high precision of around 96.85% – meaning that when it classifies a user as a High Buyer, it is almost always correct – it suffers from a low recall of about 52.34%. This low recall indicates that the model fails to identify nearly half of the actual High Buyers, which suggests that the linear assumptions underlying Logistic Regression may not be sufficient to capture the more complex, non-linear relationships within our dataset, especially given the class imbalance between High and Low Buyers.

In contrast, the Decision Tree model, which constructs a series of "if-then" rules to partition the dataset based on feature thresholds, performed significantly better. The Decision Tree achieved a cross-validation accuracy of approximately 96.88% and a test set accuracy of 97.08%. Its precision is around 97.90%, and it impressively reaches a recall of approximately 99.15%. These results indicate that the Decision Tree not only correctly classifies almost all High Buyers but also manages to balance the trade-off between precision and recall very effectively. The superior performance of the Decision Tree suggests that it is more capable of capturing the non-linear interactions among the variables and better handling the class imbalance inherent in our dataset.

Additionally, our analysis was deepened by manually categorizing users into three distinct buyer segments based on purchase frequency. Specifically, High Buyers are defined as those with a purchase frequency exceeding 100, Mid Buyers have purchase frequencies ranging from 30 to 100, and Low Buyers are those with fewer than 30 purchases. Our segmentation analysis revealed that 592 users fall into the High Buyer group, 432 into the Mid Buyer group, and 176 into the Low Buyer group. Each segment displayed distinct behavioral characteristics when comparing key features like daily usage time, ad click rate, and average spending, thereby providing deeper insights into the diverse purchasing behaviors within our user base.

Overall, while Logistic Regression provides valuable interpretability through its easily understood coefficients, it falls short in effectively capturing the full complexity of the data, as evidenced by its low recall. The Decision Tree model, on the other hand, outperforms Logistic Regression by achieving significantly higher accuracy and recall, making it the more robust choice for classifying TikTok users into High and Low Buyer categories. These findings not only answer our research questions regarding the impact of TikTok's recommendation algorithm on purchasing behavior but also deliver actionable insights that can help businesses refine their marketing strategies to better target and engage with different segments of their user base.

CHAPTER 4: CONCLUSION

4.1. Research Discussion

4.1.1. Evaluation and Discussion

The research sought to investigate how the recommendation algorithm of TikTok impacts user behavior – specifically, purchasing behavior – through classification methods. The two models used were Logistic Regression and Decision Tree. Logistic Regression showed moderate total accuracy, though its performance was characterized by a very high precision (96.85%) but significantly lower recall (52.34%) (mentioned Chapter 3). This disparity indicates that although the model is good at predicting “High Buyers,” it misses nearly half of them. The constraints within Logistic Regression inherent in modeling complex, non-linear interdependencies, further compounded by a class imbalance issue, were starkly evident.

On the other hand, the Decision Tree model had outstanding performance with cross-validation and test accuracies of approximately 96.88% and 97.08% respectively. Its precision at 97.90% and almost perfect recall at 99.15% clearly show its performance in extracting intricate interactions between behavior indicators. The improved performance of the Decision Tree reflects its ability to handle non-linearity and class imbalance and thus the ability to provide reliable classification of “High Buyers.”

4.1.2. Responding Research Questions

RQ1: What are the determinants of attitudes and behavior intentions of TikTok consumers, and how do these determinants interact with one another?

Our experiment analysis identified several key factors that significantly impact users’ purchasing behavior on TikTok. A prime among these are:

- Daily Usage Time: Longer time spent on the site translates to more exposure to content and ads, leading to a higher chance of impulse buying.
- Ad Exposure per Day: Increased frequency of ad exposure means that users are repeatedly being exposed to promotional content, driving brand recall and affecting their purchasing behavior.
- Ad Click Rate: Utilized as an estimate of user engagement with recommended content. A high click rate indicates that the users don’t merely view but also engage with the ads, therefore suggesting an increased purchase intent.
- Average Spending: This reflects users’ purchasing potential and buying inclination, as a direct measure of consumer action.

These factors are significant separately but are extremely correlated as well. For instance, increased use every day corresponds with more ad exposure, and hence increased probabilities of clicks and subsequently higher expenditures. The great recall of ~99.15% of the Decision Tree model augments the aspect that these factors, taken together, establish a feedback system whereby the involvement of the user induces more personalization of suggestions as well as impact on the buying frequency. Plots such as the correlation heatmaps and scatter plots in our EDA corroborate these relations, with the implication that whenever one indicator (e.g., duration of use on a day-to-day basis) increases, so do others (e.g., ad click-through rate and mean spend).

RQ2: Are buying intentions and attitude among TikTok users through TikTok Shop truly impacted by such search recommendation algorithms?

The findings of the experiment lend clear support for the view that TikTok's recommendation algorithm is a strong determinant of users' attitudes and buying intentions. The high-performance metrics of the Decision Tree model provide strong evidence that algorithm-based engagement metrics are predictive of purchasing behavior. Specifically, high recall and accuracy suggest that after being exposed to targeted recommendations – represented by variables like ad exposure and ad click rate – users are much more likely to become high-frequency buyers. While the Logistic Regression model had lower overall accuracy (51.67%), its greater precision (96.85%) in classifying "High Buyers" further supports the conclusion that the inherent algorithmic properties of the underlying models are good predictors of user interactions and consequent purchasing behavior. These results affirm the reality that the algorithm on TikTok not only generates content that is customized to engage the largest possible number of users but also directly influences the purchase decision through TikTok Shop.

RQ3: What are the implications for managers in companies to tap into to conduct business in the marketplace of TikTok Shop?

Pragmatic implications of our findings for managers are presented as follows.

Firms can invest in programs that use algorithm-based insights to target highly active user segments. Since the Decision Tree model is identifying "High Buyers" intensely, marketing can be targeted to these segments in order to obtain maximum conversion rates and revenues. With algorithmic personalization, firms can customize their content and offers based on precise behavioral information, thereby optimizing the overall effectiveness of their marketing campaigns.

As ad exposure and click-through rates have extensive control over consumer buying behavior, companies need to organize advertising campaigns to optimize ad relevance and interactivity. The approach can involve dynamic placing of ads and real-time optimization of content so that it is more likely to capture people's attention. Effective advertising models ensure that marketing content appeals to targeted consumers and increases the likelihood of engagement and follow-up buying behavior.

The clear-cut segregation among High, Mid, and Low Buyers – distinct from both classification outcomes and manual segregation – provides a robust context for resource allocation. With such segmentation, company resources can be employed to get Mid Buyers elevated to High Buyers through differentiated promotion, tailor-made offers, and diversified communications strategies, and strategies can also be designed to recover Low Buyers. Segmentation-based resource allocation helps companies focus their resources in places that will return to them the most.

The limitations seen with traditional linear models like Logistic Regression make the adoption of more sophisticated, non-linear predictive models even more essential. Decision Trees and ensemble methods, for instance, have proven to be superior in performance through their ability to effectively model complex user behaviors and non-linear relationships. With sophisticated predictive models, more accurate and actionable insights can be derived, leading eventually to improved strategic decision-making.

To better optimize marketing plans, firms need to also look to enhance data collection and integration processes. Incorporating other variables related to content relevance – FYP_relevance – and psychological variables like FOMO and sentiment analysis from consumer reviews can lead to more detailed insights. With the enhanced data integration method, more effective customer engagement strategies can be formulated that are more precisely aligned with real drivers of purchase behavior.

Finally, with increasing concerns regarding data privacy and ethical usage of AI, managers have to balance personalization's benefits with robust safeguards for the safeguarding of user data. Clear data open practices and clear-cut processes for obtaining user consent need to be incorporated in any policy utilizing algorithmic recommendations. Companies can ensure this way that personalization activities are ethically conducted while maintaining user confidence and respecting privacy.

In summary, the study proves that TikTok's algorithmic recommendation plays a crucial role in influencing user behavior, especially purchasing behavior. The better performance of the Decision Tree model gives a sound foundation for explaining and predicting such behaviors, thus providing useful insights for strategic marketing and operational optimization in the TikTok Shop market.

4.2. Dashboard Implementation

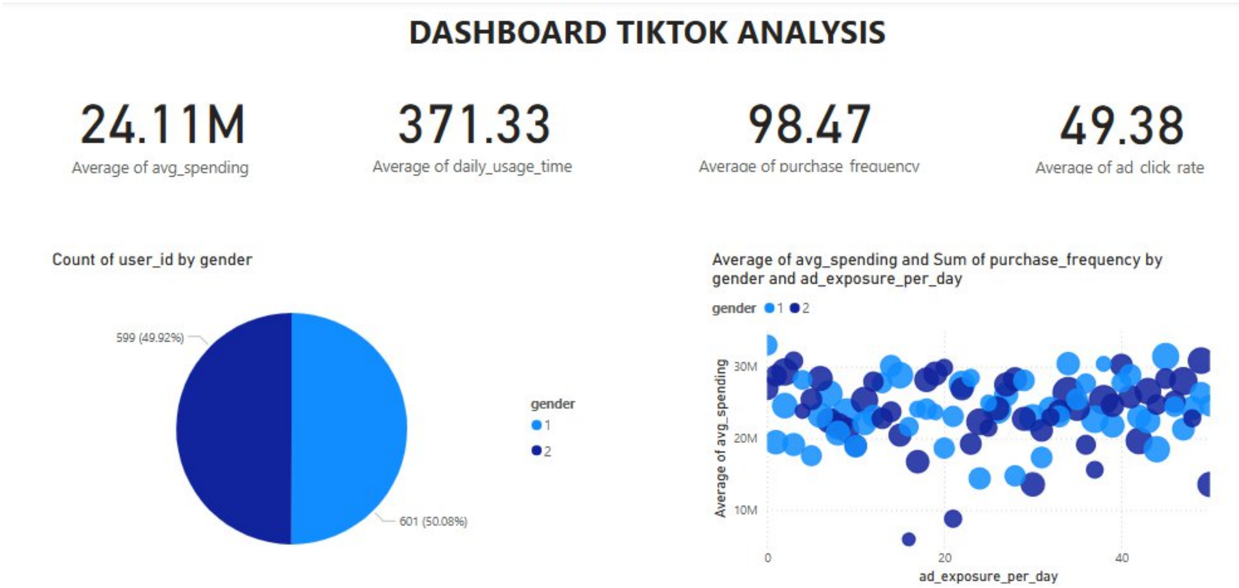

Figure 4.1: TikTok Analysis Dashboard Insights.

The TikTok analysis dashboard effectively summarizes key user engagement and purchasing behaviors, offering clear insights to stakeholders. It highlights essential metrics that capture user interactions with TikTok's content and advertisements, as well as their purchasing patterns.

- **Average of Average Spending (24.11M):** This substantial average spending value underscores the strong purchasing activity of TikTok users, indicating the platform's significant commercial potential.
- **Average of Daily Usage Time (371.33 minutes):** With an average daily usage exceeding six hours, users demonstrate high engagement with the platform. This metric strongly suggests that users are receptive to personalized marketing and recommendation strategies due to extensive interaction periods.
- **Average Purchase Frequency (98.47 purchases):** This high frequency confirms a highly active consumer base. Frequent purchasing indicates successful engagement strategies, likely influenced by targeted advertising and relevant content recommendations.
- **Average Ad Click Rate (49.38%):** Nearly half of ad exposures result in user interactions. This high click rate demonstrates the effectiveness of TikTok's targeted ads in capturing user interest and potentially driving purchasing decisions.
- **Gender Distribution (Pie Chart):** The user base is almost evenly split, with Female (Gender 2) slightly leading (50.08%) compared to Male (Gender 1)

(49.92%). This balanced distribution suggests a broad appeal of the platform and indicates marketing strategies could be equally effective for diverse demographics.

- **Scatterplot Analysis (Average Spending vs. Ad Exposure):** This visualization provides further clarity on the relationship between ad exposure and purchasing behavior. Notably, a positive correlation can be observed: users who experience higher daily ad exposure generally exhibit higher average spending. The distribution suggests targeted ads significantly enhance consumer engagement and spending habits. Gender-wise differentiation, represented by two colors, also allows marketers to tailor gender-specific ad strategies, optimizing the impact of advertisements.

Overall, this dashboard offers valuable business intelligence by clearly indicating high user engagement levels, effective ad strategies, and promising consumer spending behaviors. It suggests actionable insights for marketers on TikTok, reinforcing the platform's strategic value in digital commerce and targeted advertising efforts.

4.3. Research Limitations and Future Work

4.3.1. Research Limitations

While our classification methods have achieved encouraging results, this research faces a number of limitations that are worth exploring further. First, although Decision Tree model performance was excellent in predicting "High Buyers," our investigation was largely based on behavioral metrics – ad exposure per day, daily usage time, ad click rate, and average spending – without incorporating content-related or psychological variables (e.g., FYP relevance, FOMO, or sentiment analysis). This focus, while effective for classification, might overlook more granular aspects of user activity that can inform purchase.

The other concern is the inherent challenge of class imbalance in the data. In spite of techniques like class weighting, the imbalance still impacted the precision of models like Logistic Regression, which was not able to learn from the whole range of user activity. Moreover, the current study's segmentation strategy – largely based on predefined purchase frequency cut-offs – has the potential to oversimplify the multifaceted continuum of user engagement. Manual segmentation, while useful for preliminary analysis, is not capable of recognizing underlying patterns that could exist among the diverse user population.

Future studies should examine the application of unsupervised learning methods, particularly clustering algorithms, in a bid to make inferences that go beyond the surface level of TikTok user behavior.

4.3.2. Future Work: Examining Clustering Techniques

One promising line of future research is the application of clustering methods to divide up TikTok users in a less overt and more data-driven fashion. Unlike supervised classification, clustering can find naturally occurring groupings in the data without applying preconceived labels, thus revealing underlying behavioral tendencies that might otherwise be masked. For instance, methods such as K-Means, hierarchical clustering, or density-based clustering (e.g., DBSCAN) can be employed to analyze sub-segments of the user population.

To effectively utilize a clustering approach, future work must first enhance the feature engineering process. This can involve adding additional dimensions such as content interaction metrics (e.g., relevance of FYP), user comment sentiment scores, and psychological factors such as FOMO or user satisfaction. A more complete feature set will paint a more accurate portrait of the drivers of user behavior on TikTok, and will most likely lead to more insightful clusters.

The clustering analysis can yield actionable data by revealing subgroups with distinctive behavior patterns. For example, clustering could determine that among the broader "Mid Buyer" category, there exist sub-segments with varying levels of activity or responsiveness to specially targeted ads. This type of information can be utilized to design highly focused marketing campaigns to convert these subgroups into "High Buyers" or re-activate "Low Buyers."

In addition, more advanced clustering techniques – i.e., ensemble clustering or multimodal models combining clustering and supervised learning – may also further refine user segmentation. These methods will allow for more explainability and develop a more substantial understanding of dynamics in play. By comparing cluster outcomes with those of classification, businesses can then gain a better understanding of the user behavior at play, thereby eliminating the divide between predictive brilliance and actionable strategic insight.

In summary, although the present study confirms the significance of TikTok's recommendation algorithm through classification models, the integration of clustering methods in the future can potentially shed more light on the complex, multidimensional facets of user behavior. In addition to enhancing segmentation quality, this addition will also help influence more subtle and effective marketing campaigns on platforms like TikTok Shop.

4.4. Study Reflections

Key lessons included the crucial studies for advanced non-linear models, rigorous data preprocessing, and rich visual analytics, alongside integrating quantitative and qualitative methods, with future work focusing on clustering to uncover hidden consumer patterns. It's essential to employ sophisticated, custom analytical approaches, harness detailed visual insights, and uphold meticulous data management practices.

Firstly, the comparison between Logistic Regression and Decision Tree models emphasized the importance of selecting appropriate analytical tools that match data complexity. The significant performance gap demonstrated that linear models are not able to properly manage non-linear relationships and class imbalances present in user behavior data. The outcome confirms the lesson that more sophisticated, non-linear techniques like Decision Trees or ensemble algorithms are typically required to properly capture subtle interactions for complicated, real-world datasets.

Second, our excessive use of visualizations throughout the Exploratory Data Analysis (EDA) step proved to be highly beneficial. Histograms, box plots, scatter plots, and heatmaps did not just inform us of the underlying structure and distribution of key variables but also of high interdependencies between the measures of user engagement. The visualization tools played a central role in advising feature selection and model validation, thereby affirming the importance of quality data visualization to the derivation of actionable findings and results communication to stakeholders.

Third, integrating supervised classification and hand segmentation produced a more detailed comprehension of user activity. Whereas predictive accuracy came from data produced by classification models, hand segmentation allowed us to contextualize the quantitative results in the framework of real purchasing behaviors. This two-way strategy marked the power of unifying data analysis with qualitative analysis to develop more advanced marketing strategies.

Further, the project underlined the necessity of proper data preprocessing and feature engineering. Dealing with missing values, treating outliers, and scaling features were all critical operations that safeguarded the integrity of our analysis. Not only did these procedures help improve model performance, but they also made our findings more reliable and demonstrated how careful data preparation is the building block of good research.

Finally, the study has revealed the potential benefits of expanding the research framework through the application of unsupervised learning techniques, such as clustering. Although still in its nascent stages in this study, preliminary findings in clustering indicate promise in uncovering hidden unseen patterns of behavior, beyond the reach of classification methods alone. It offers a promising avenue for future studies, in which the integration of clustering methodologies may lead to even more effective and more focused marketing activity.

Overall, the lessons highlight the significance of using advanced, bespoke analytical methods, leveraging rich visual analytics, and adhering to rigorous data handling practices. They also mirror the value of a multi-faceted research methodology that involves predictive modeling along with exploratory methods to achieve an in-depth understanding and mastery of digital consumer behavior.

REFERENCES

Alexander, J. (2020). TikTok reveals some of the secrets, and blind spots, of its recommendation algorithm. The Verge. https://www.theverge.com/2020/6/18/21296044/tiktok-for-you-page

Anh, Q. (2024). Việt Nam nằm trong top 5 quốc gia có lượng người dùng Tiktok nhiều nhất thế giới [The number of Tik Tok users in Vietnam was in top 5 of the world] . Nhịp Sống Kinh Tế Việt Nam & Thế Giới. https://vneconomy.vn/viet-nam-nam-trong-top-5-quoc-gia-co-luong-nguoi-dung-tiktok-nhieu-nhat-the-gioi.htm

Backlinko. (2024). TikTok Statistics you need to know. Backlinko. https://backlinko.com/tiktok-users

Baek, Y. M., Bae, Y., & Jang, H. (2013). Social and Parasocial Relationships on Social Network Sites and Their Differential Relationships with Users' Psychological Well-Being. Cyberpsychology, Behavior, and Social Networking, 16(7), 512–517. https://doi.org/10.1089/cyber.2012.0510

Bandy, J., & Diakopoulos, N. (2020). #TulsaFlop: A Case Study of Algorithmically-Influenced Collective Action on TikTok. ArXiv:2012.07716 [Cs]. https://arxiv.org/abs/2012.07716

Barve, G. (2016). Social Media and its Effects on Society. International Journal of School and Cognitive Psychology, 03(01). https://doi.org/10.4172/2469-9837.1000166

Bhandari, A., & Bimo, S. (2022). Why's Everyone on TikTok Now? The Algorithmized Self and The Future of Self-Making on Social Media. Social Media + Society, 8(1), 1–11. https://doi.org/10.1177/20563051221086241

Braghieri, L., Levy, R., & Makarin, A. (2021). Social Media and Mental Health. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3919760

Brindmark, V., & Wickström, G. (2023). Algorithms in feed-based SNS and their influence on digital platform User Habits : A quantitative study of the link between algorithms in feed-based SNS and user habits amongst high-school and university students in Sweden. DIVA. https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1768646&dswid=-6836

Ceci, L. (2023, October 10). TikTok global downloads worldwide 2023. Statista. https://www.statista.com/statistics/1377008/tiktok-worldwide-downloads-quarterly/

Chen, Z., & Shi, C. (2022). Analysis of Algorithm Recommendation Mechanism of TikTok. International Journal of Education and Humanities, 4(1), 12–14. https://doi.org/10.54097/ijeh.v4i1.1152

Davis, J. P. (2019). The TikTok Strategy: Using AI Platforms to Take Over the World. INSEAD Knowledge. https://knowledge.insead.edu/entrepreneurship/tiktok-strategy-using-ai-platforms-take-over-world

Digital 2024: 5 billion social media users. (2024). We Are Social. https://wearesocial.com/us/blog/2024/01/digital-2024-5-billion-social-media-users/

Giải mã thói quen trực tuyến của người Việt: Báo cáo Connected Consumer Q2 2023. (2023). Tomorrow Marketers. https://blog.tomorrowmarketers.org/thoi-quen-truc-tuyen-nguoi-viet-connected-consumer/

Giải mã xu hướng sử dụng mạng xã hội của GenZ toàn cầu - Báo cáo từ GWI 2023. (2023). Tomorrow Marketers. https://blog.tomorrowmarketers.org/xu-huong-su-dung-mang-xa-hoi-cua-genz-toan-cau-report-gwi/

Grandinetti, J. (2021). Examining embedded apparatuses of AI in Facebook and TikTok. Ai & Society, 38(4), 1–14. https://doi.org/10.1007/s00146-021-01270-5

Hagar, N., & Diakopoulos, N. (2023). Algorithmic indifference: The dearth of news recommendations on TikTok. New Media & Society. https://doi.org/10.1177/14614448231192964

Han, J. (2024). The Impact of Personalized Recommendations on Consumer Purchase Decisions on TikTok – A Case Study of College Students. Interdisciplinary Humanities and Communication Studies, 1(6). https://doi.org/10.61173/bj0e1289

How TikTok recommends videos #ForYou. (2020). TikTok. https://newsroom.tiktok.com/en-us/how-tiktok-recommends-videos-for-you

Ivie, E. J., Pettitt, A., Moses, L. J., & Allen, N. B. (2020). A Meta-Analysis of the Association Between Adolescent Social Media Use and Depressive Symptoms. Journal of Affective Disorders, 275(1), 165–174. https://doi.org/10.1016/j.jad.2020.06.014

Jingtian, H. (2024). The Impact of Personalized Recommendations on Consumer Purchase Decisions on TikTok – A Case Study of College Students. Dean & Francis, 1(6). https://doi.org/10.61173/bj0e1289

Kang, H., & Lou, C. (2022). AI agency vs. human agency: understanding human–AI interactions on TikTok and their implications for user engagement. Journal of Computer-Mediated Communication, 27(5). https://doi.org/10.1093/jcmc/zmac014

Kemp, S. (2024). Digital 2024: Vietnam. DataReportal. https://datareportal.com/reports/digital-2024-vietnam

Kim, N. (2022). [Infographic] Cơn sốt #TikTokMadeMeBuyIt và bài học dành cho marketer. Advertising Vietnam. https://advertisingvietnam.com/infographic-con-sot-tiktokmademebuyit-va-bai-hoc-danh-cho-marketer-l19451

Klug, D., Qin, Y., Evans, M., & Kaufman, G. (2021). Trick and Please. A Mixed-Method Study On User Assumptions About the TikTok Algorithm. Trick and Please. A Mixed-Method Study on User Assumptions about the TikTok Algorithm, 84–92. https://doi.org/10.1145/3447535.3462512

Knight, W. (2019). This Technique Can Make It Easier for AI to Understand Videos. WIRED. https://www.wired.com/story/technique-easier-ai-understand-videos/

Kowal, J., & Klebaniuk, J. (2023). Psychological Impact of Social Media Use on Quality of Life and Human Potentiality. https://www.researchgate.net/publication/362983735_Psychological_Impact_of_Social_Media_Use_on_Quality_of_Life_and_Human_Potentiality

Kumar, M. (2023). TikTok Profiles Data. Kaggle. Retrieved from https://www.kaggle.com/datasets/manishkumar7432698/tiktok-profiles-data

Lovejoy, B. (2021). How TikTok's algorithm works: A fascinating and disturbing analysis. 9to5Mac. https://9to5mac.com/2021/07/28/how-tiktoks-algorithm-works/

Ma, Y., & Hu, Y. (2021). Business Model Innovation and Experimentation in Transforming Economies: ByteDance and TikTok. Management and Organization Review, 17(2), 1–7. https://www.cambridge.org/core/journals/management-and-organization-review/article/business-model-innovation-and-experimentation-in-transforming-economies-bytedance-and-tiktok/057C8387EC953C14ED10AA8996F53947

Martínez-Martínez, F. D., González-García, H., & González-Cabrera, J. (2022). Perfiles de redes sociales de estudiantes, necesidades psicológicas básicas, autoconcepto e intención de ser físicamente activo. Behavioral Psychology/Psicología Conductual, 30(3), 757–772. https://doi.org/10.51668/bp.8322310s

Mahmood, M. A. (2022). TikTok Dataset. Kaggle.com. https://www.kaggle.com/datasets/muhammadanasmahmood/tiktok-dataset

Morán-Pallero, N., & Felipe-Castaño, E. (2021). Autoconcepto en las redes sociales y su relación con el afecto en adolescentes. Behavioral Psychology/Psicología Conductual, 29(3), 611–625. https://doi.org/10.51668/bp.8321306s

Newton, C. (2020). Three takeaways from a visit to TikTok's new transparency center. The Verge. https://www.theverge.com/interface/2020/9/11/21430822/tiktok-transparency-visit-tour-algorithms-for-you-page

Nguyen, X. N. (2018). GIỚI THIỆU NGHIÊN CỨU THEO PHƯƠNG PHÁP HỖN HỢP. INTRODUCTION to MIXED METHODS RESEARCH. https://doi.org/10.13140/RG.2.2.20552.57604

Phạm, Q. H. (2023). Thống kê Shopify mới nhất và các sự thật thú vị về Shopify. Magenest - One-Stop Digital Transformation Solution. https://magenest.com/vi/thong-ke-shopify/

Przybylski, A. K., Murayama, K., DeHaan, C. R., & Gladwell, V. (2013). Motivational, emotional, and Behavioral Correlates of Fear of Missing out. Computers in Human Behavior, 29(4), 1841–1848. https://doi.org/10.1016/j.chb.2013.02.014

Schellewald, A. (2022). Theorizing "Stories About Algorithms" as a Mechanism in the Formation and Maintenance of Algorithmic Imaginaries. Social Media + Society, 8(1). https://doi.org/10.1177/20563051221077025

Social, S. (2024). 27 Số Liệu Thống Kê TikTok Mới Nhất 2024. GIGAN JSC. https://gigan.vn/27-so-lieu-thong-ke-tiktok-2024/

Talu, U., & Denis, N. (2023). Turkey 2023 Elections on Tik Tok: Insights from Computer Vision and Natural Language Processing. https://studenttheses.uu.nl/bitstream/handle/20.500.12932/45356/Talu%2c%20U._9554041_Thesis_revised_FINAL.pdf

Taylor, S. H., & Anthony, C. Y. (2024). The lonely algorithm problem: the relationship between algorithmic personalization and social connectedness on TikTok. Journal of Computer-Mediated Communication, 29(5). https://doi.org/10.1093/jcmc/zmae017

Trần, T. T. T., & Lê, T. H. (2024). ẢNH HƯỞNG CỦA TIKTOK ĐẾN THÓI QUEN TIÊU DÙNG CỦA SINH VIÊN HIỆN NAY. Tâm lý – Giáo dục. https://tamlygiaoduc.com.vn/wp-content/uploads/2024/09/15-22.pdf

Tổng hợp 04 xu hướng tiêu dùng nổi bật năm 2024. (2023, December 26). Tomorrow Marketers. https://blog.tomorrowmarketers.org/bao-cao-tong-hop-xu-huong-tieu-dung-2024/

Trung, H. (2025). Dân số thế giới đạt 8,09 tỷ người vào ngày đầu năm mới 2025. Nhân Dân. https://nhandan.vn/dan-so-the-gioi-dat-809-ty-nguoi-vao-ngay-dau-nam-moi-2025-post853695.html

Vũ, H. L. (2022). Tại sao con người luôn muốn có cảm giác "thuộc về"? Vietcetera. https://vietcetera.com/vn/tet-den-di-tim-cam-thuc-thuoc-ve

Zhang, M., & Liu, Y. (2021). A commentary of TikTok recommendation algorithms in MIT Technology Review 2021. Fundamental Research, 1(6), 846–847. https://doi.org/10.1016/j.fmre.2021.11.015

Zhao, H., & Wagner, C. (2022). How TikTok leads users to flow experience: investigating the effects of technology affordances with user experience level and video length as moderators. Internet Research, 33(2). https://doi.org/10.1108/intr-08-2021-0595

Zhao, Z. (2021). Analysis on the "Douyin (Tiktok) Mania" Phenomenon Based on Recommendation Algorithms. E3S Web of Conferences, 235(03029), 1–10. https://doi.org/10.1051/e3sconf/202123503029

Zhou, R. (2024). Understanding the Impact of TikTok's Recommendation Algorithm on User Engagement. International Journal of Computer Science and Information Technology, 3(2), 201–208. https://doi.org/10.62051/ijcsit.v3n2.24