

Topic Modelling and Sentiment Analysis of  
E-commerce Reviews  
on Scotch Whiskey in China

by

**Amelia Yijia Shao**

An honors thesis submitted in partial fulfillment  
of the requirements for the degree of  
Bachelor of Science  
Business and Economics Honors Program  
NYU Shanghai

May 2026

Professor Marti G. Subrahmanyam  
Professor Christina Wang  
Professor Wendy Jin

Faculty Advisers

Professor Ying Lei

Thesis Adviser

	1
<b>Abstract</b>	<b>2</b>
<b>1 Introduction</b>	<b>3</b>
1.1 Background: The Growing and Transitioning Chinese Whiskey Market	3
1.2 Literature Review and Research Gap	3
1.3 Research Gap	4
1.4 Research Objectives	5
<b>2 Data</b>	<b>5</b>
2.1 Brand and Product Selection	5
2.2 Data Collection	6
2.3 Data Pre-processing	7
<b>3 Empirical Framework</b>	<b>8</b>
3.1 Topic Modeling via BERTopic	8
3.2 Sentiment Analysis Configuration	9
<b>4 Result</b>	
<b>4.1 Overview of Extracted Topics</b>	<b>10</b>
4.2 Correlation and Thematic Relationships	12
4.2.1 The Convergence of Quality and Logistics (“Quality = Packaging”).	12
4.2.2 The “Isolated” Fan Economy.	13
4.2.3 The Dichotomy of E-commerce Discourse: Products vs. Personas.	14
4.3 Sentiment Distribution and the Marketing Volume Paradox	15
4.4 Brand-specific Keyword Analysis and Strategic Alignment	17
4.4.1 The Macallan: Congruence with Premium Positioning.	18
4.4.2 The Balvenie: The E-commerce Promotion Paradox.	18
4.4.3 Johnnie Walker Black Label: The Blurred Brand Narrative.	19
<b>5 Strategic Suggestions</b>	<b>19</b>
5.1 Elevating Logistics to a Core Product Attribute (“The First Touch”)	19
5.2 Mitigating the “Double-Edged Sword” of the Fan Economy	20
5.3 Value-Added Promotions over Price Degradation	20
<b>6 Limitations and Future Research</b>	<b>21</b>
6.1 Sample Selection and Category Mismatch	21
6.2 Data Volume Disparity and Topic Skewing	21
6.3 Platform Noise and the “Fake Review” Phenomenon	21
<b>7 Conclusion</b>	<b>22</b>
<b>Declaration of generative AI and AI-assisted technologies</b>	<b>24</b>
<b>Acknowledgements</b>	<b>24</b>
<b>References</b>	<b>25</b>

**Abstract**

This study investigates the evolving Chinese e-commerce whiskey market by decoding the digital purchasing motivations and post-purchase sentiments of consumers. Utilizing BERTopic modeling and sentiment analysis, the research analyzes online reviews of entry-level 12-year-old whiskies from three major Scotch brands (Johnnie Walker, The Macallan, and The Balvenie). The findings reveal a critical semantic convergence between intrinsic liquid quality and extrinsic delivery services, demonstrating that a flawless logistical experience is essential to perceived product value. Furthermore, the analysis uncovers a distinct “Fan Economy” paradox: while aggressive celebrity-driven marketing generates massive sales volume, it effectively hijacks the core brand narrative and counterintuitively exposes the brand to higher negative sentiment due to operational and logistical bottlenecks. The study concludes with actionable strategic recommendations, advising international brands to elevate logistics to a core product attribute, implement segmented SKU strategies to safely navigate celebrity traffic, and utilize value-added promotions to protect long-term brand equity in a highly localized digital landscape.

**Keywords:** E-commerce, Whiskey Market, Topic Modeling, Sentiment Analysis, Fan Economy, Consumer Behavior

## **1 Introduction**

### **1.1 Background: The Growing and Transitioning Chinese Whiskey Market**

According to market reports, the size of the Chinese Whiskey market had reached 5.5 billion yuan (\$756 million) in 2023, which is a 10% increase from the previous year (Li, 2025). It's also commonly predicted that the demand and sales in China will continue to grow. In this large, indispensable, and foreign market for Scotch Whisky brands, there exists a necessity to better cater to this market.

Nevertheless, the whisky industry is under significant pressure due to global economic instability and the end of post-pandemic growth. The supply-demand equilibrium for a non-essential product like whisky has been disrupted, and thus, many (Scotch) whisky brands are looking for a transformation to win consumers. It's predicted that transformations will happen in terms of packaging, brand image, branding focus, and more (Yewuyong Weishiji Zhinan, 2024). Therefore, in this transitional period of time, understanding the real focus of consumers is crucial for brands to develop an effective strategy to share a piece of cake in the Chinese market.

### **1.2 Literature Review and Research Gap**

The digital transformation has introduced a highly dynamic consumer base in China, where e-commerce consumers are increasingly influenced by localized digital marketing strategies, such as platform-specific subsidies, and the application of the unique "Fan Economy" (Nan & Xu, 2025). Within this context, localization strategy has become a pivotal tool for international brands to significantly enhance brand identity and resonance in the local market (He, 2026). However, as the competitiveness of the e-commerce market intensifies, the

complexity of consumer motivations—ranging from product quality to emotional loyalty toward idols—presents a challenge for brand management.

It has been proven through research that, along with the increasing competitiveness of the e-commerce market, online reviews, which have a deep impact on consumers' mindset and loyalty, are now essential for business owners to manage (Rachmiani et al., 2024).

Understanding this authentic feedback is highly valuable for the entire chain of identifying dissatisfaction, fixing operational problems, and paving the way for future sales.

To effectively unveil consumer opinions from such massive datasets, topic modeling and sentiment analysis based on text mining are frequently employed, supported by mature machine learning and deep learning techniques (Daza et al., 2024; Zhao et al., 2021). Specifically, BERTopic has gained prominence as a large language model-based approach to topic modeling. By utilizing BERT for vector embedding to capture semantic and contextual information, BERTopic is considered to outperform traditional algorithms (An et al., 2023). Through these advanced models, quantifiably valuable insights can be extracted, enabling brands to better cater to the increasingly complex demands of the Chinese digital consumer.

### **1.3 Research Gap**

In recent years, consumers' reviews in e-commerce have been a frequently examined category. However, there is a marked scarcity of research concerning consumer evaluation or marketing strategies within the spirits domain, especially the specific segment of whisky, and this limitation is particularly pronounced for the Chinese market. Although e-commerce platforms serve as a critical channel for consumers to purchase whisky, accumulating a vast number of reviews for popular or benchmark products, few studies have utilized or analyzed this

data, which can be important to reveal consumer emotions and attitudes. Moreover, as these reviews generate EWOM (Electronic Word-of-Mouth), which is influential to consumer purchase decisions, there exists a need to see through these reviews for brands' reference (Lievonen, 2017), which should also be true in the category of whisky.

## **1.4 Research Objectives**

To bridge this gap, this study employs advanced text mining techniques (including topic modeling and sentiment analysis) on real consumer reviews extracted from major Chinese e-commerce platforms. By analyzing the organic discourse surrounding three distinct entry-level products from Johnnie Walker Black Label, The Macallan, and The Balvenie, the aim of this paper is to: (1) Identify the core semantic themes and purchasing drivers of Chinese e-commerce whiskey buyers; (2) Analyze the semantic correlation between physical product quality and e-commerce service attributes (e.g., logistics and packaging); (3) Evaluate how aggressive localization strategies, specifically the Fan Economy, impact brand sentiment and overall consumer perception compared to traditional product-driven positioning.

## **2 Data**

### **2.1 Brand and Product Selection**

To control for variables as effectively as possible, three mid-to-high-end Scotch whisky brands with distinct market positioning were selected. A primary prerequisite was that each brand maintains an official flagship store on the Taobao/Tmall platforms. Taobao/Tmall is the largest e-commerce platform in China, where brands would be able to open and directly control flagship stores, ensuring the authenticity and stable quality of the products under scrutiny

To capture a comprehensive spectrum of the market, the selected brands represent distinct segments. First, Johnnie Walker Black Label represents Blended Scotch, positioned as globally balanced and accessible (LBB Editorial, 2023). Second, The Macallan represents Luxury Single Malt, positioned as premium, exclusive, and collectible (Alcohol Please, 2025; Elite Traveler, 2023). Finally, The Balvenie represents Artisanal Single Malt, positioned as handcrafted, traditional, and heritage-focused (William Grant & Sons, n.d.).

Regarding target product selection, a data-driven approach was employed: the single top-selling and most-reviewed item within each brand's official store was chosen. These products serve as a proxy to capture and represent current consumer attitudes toward their respective brands.

Notably, while the primary selection criterion was based strictly on sales volume and review density, this process yielded an unanticipated but methodologically advantageous result: all three selected best-sellers shared a 12-year age statement (i.e., Johnnie Walker Black Label 12 Years Old, The Macallan 12 Years Old, and The Balvenie 12 Years Old). This serendipitous alignment provides a natural control for the variable of liquid maturation and product tier, allowing comparability of the selected products across the three distinct brand positionings.

## **2.2 Data Collection**

To capture consumer sentiment surrounding the critical "Double 11" (Singles' Day) shopping festival, the largest yearly e-commerce event in China. Data collection was strategically executed on November 23, 2025. This timing ensured the inclusion of post-purchase evaluations, capturing not only product quality feedback but also unboxing and logistics experiences following the peak transaction period. Automated web scraping techniques were

utilized to exhaustively extract all publicly accessible user reviews from the designated products within the three official flagship stores. This process yielded a total corpus of 3,500 raw reviews. The dataset composition naturally reflects the varying market penetration and sales volumes of the selected segments, comprising 2,000 reviews for Johnnie Walker, 1,000 for The Macallan, and 500 for The Balvenie.

### 2.3 Data Pre-processing

Given the linguistic characteristics of Chinese text, which lacks natural spaces between words, accurate text segmentation is a crucial prerequisite for topic modeling. This study employed the Chinese text segmentation module, *Jieba*, which is widely recognized for its efficiency in natural language processing. To enhance the precision of the segmentation in the specific context of the spirits industry, a highly customized, domain-specific lexicon was integrated into the algorithm. This user dictionary ensured that industry-specific terminology (e.g., Sherry cask (雪莉桶), Single malt (单一麦芽), Peat (泥煤味)), platform-specific promotional terms (e.g., Mega-subsidy (百亿补贴)), and brand-specific entities (e.g., Jianci Tan (檀健次), who is the endorser of Johnnie Walker Black Label) were accurately identified as single, meaningful semantic units rather than being incorrectly parsed into fragmented characters.

Following the initial segmentation, a data cleaning protocol was applied to reduce noise. A tailored stop-word dictionary was developed and deployed to filter out syntactical fillers, generic adjectives, and colloquial expressions frequently found in e-commerce reviews (e.g., haha (哈哈), really (真的), just (就是)). This dual approach of customized semantic retention and noise elimination ensured that the refined dataset was highly concentrated on substantive

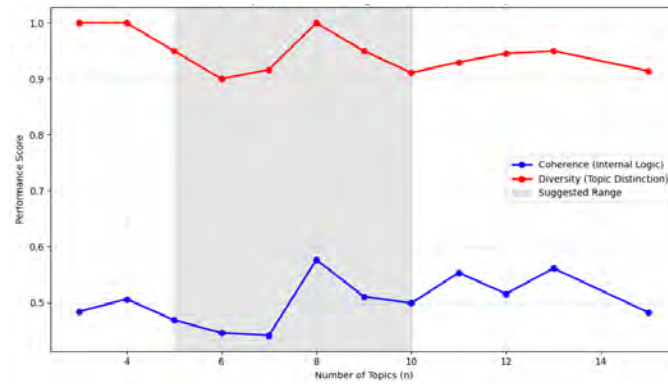
consumer insights, thereby laying a robust foundation for the subsequent unsupervised machine learning analysis.

### **3 Empirical Framework**

#### **3.1 Topic Modeling via BERTopic**

The optimal number of latent topics was definitively established as 8, derived from a rigorous quantitative evaluation of semantic coherence and topic diversity. Initial deployments of the BERTopic algorithm without a predefined topic parameter resulted in excessive fragmentation and semantic overlap among the generated clusters. To systematically resolve this, a hyperparameter tuning protocol was executed to determine the optimal topic count ( $n$ ).

Utilizing the paraphrase-multilingual-MiniLM-L12-v2 embedding model coupled with a Class-based TF-IDF (c-TF-IDF) transformer, the algorithm iterated through predefined topic configurations ranging from 3 to 15. The performance of each configuration was evaluated using two primary metrics: Topic Coherence (measured via the  $c_v$  score) to assess internal semantic consistency, and Topic Diversity to measure the external distinctiveness of vocabulary across clusters. The quantitative evaluation revealed a distinct “double peak” alignment, where both coherence and diversity scores maximized simultaneously at exactly 8 topics. Consequently,  $n=8$  was selected as the optimal hyperparameter, ensuring the resulting themes were both internally unified and mutually exclusive.

**Figure 1***Topic Coherence and Topic Diversity Test*

*Note.* This diagram illustrates the relative performances of each topic number option. When the number of topics is 8, both coherence and diversity get the highest performance score.

### 3.2 Sentiment Analysis Configuration

Following the clustering of semantic themes, sentiment analysis was conducted to quantify consumer attitudes across different brands and topics. The SnowNLP library, optimized for Chinese text processing, was utilized to assign a sentiment probability score ranging from 0 (extremely negative) to 1 (extremely positive) to each review.

By analyzing the density distribution of the scores, a critical threshold was established at 0.3. Scores falling below this 0.3 threshold were strictly classified as negative feedback. This calibrated threshold effectively isolated genuine consumer dissatisfaction from passive neutrality, enabling a highly sensitized extraction of critical operational and product-related pain points.

## 4 Result and Discussion

### 4.1 Overview of Extracted Topics

**Table 1**

*Identified Topics and Core Keywords from Consumer Reviews*

Topic ID & Theme	Core Original Keywords (extract)	English Translation & Cultural Context
Topic #0: Entry-Level & Social	入门, 朋友, 好喝, 喜欢	<b>“Beginner / Friends”</b> : Represents consumers buying for social gatherings or starting their whiskey journey.
Topic #1: Core Product Quality	酒液, 口感, 质量, 整体	<b>“Liquid / Palate / Quality”</b> : Focuses on the objective quality and overall tasting experience of the liquor itself.
Topic #2: Unboxing & Logistics	包装, 物流, 发货, 完好	<b>“Packaging / Logistics / Intact”</b> : The critical e-commerce "first touchpoint." Highlights the demand for safe, premium delivery.
Topic #3: The "Fan Economy"	檀健次, 代言人, 周边, 杯子	<b>“Idol / Ambassador / Merch (Glasses)”</b> : Driven by celebrity endorsement. Purchases are motivated by supporting the idol and collecting exclusive merchandise.
Topic #4: "Daily Sipper" Profile	口粮, 入口, 经典, 味道	<b>“Ration (Daily Sipper) / Classic”</b> : “Ration (口粮)” is a Chinese slang for highly affordable, decent-quality whiskey consumed on a daily basis without financial stress.
Topic #5: Price Sensitivity & Service	百亿补贴, 退款, 客服, 投诉	<b>“Mega-Subsidy / Refund / Complaints”</b> : Highly price-sensitive buyers drawn by platform mega-promotions, often leading to post-purchase disputes.
Topic #6: Loyalty & Business Occasions	回购, 应酬, 聚会, 多次	<b>“Repurchase / Business Socializing”</b> : Indicates strong brand loyalty and consumption during formal business banquets or private gatherings.

*Note.* The keywords listed are the most representative terms identified by the BERTopic model.

Translations were provided by the researcher to capture cultural nuances (e.g., “Ration” for 口粮, which means this product is more of a daily drink option rather than a collection).

Through the topic modeling analysis of the collected consumer reviews, seven distinct topics emerged, representing a diverse range of purchasing motivations, product perceptions, and consumption scenarios. As detailed in Table 1, these topics span from objective product evaluations (e.g., Topic #1: Core Product Quality) to highly context-specific cultural behaviors unique to the Chinese e-commerce landscape. Notably, the extraction identified culturally nuanced themes such as the “Fan Economy” (Topic #3), driven by celebrity endorsements, and the “Daily Sipper” (“Ration”) (Topic #4), reflecting a demand for affordable, everyday consumption. Additionally, e-commerce-specific concerns, including unboxing experiences (Topic #2) and price sensitivity related to platform subsidies (Topic #5), highlight the critical role of service and logistics in overall customer satisfaction.

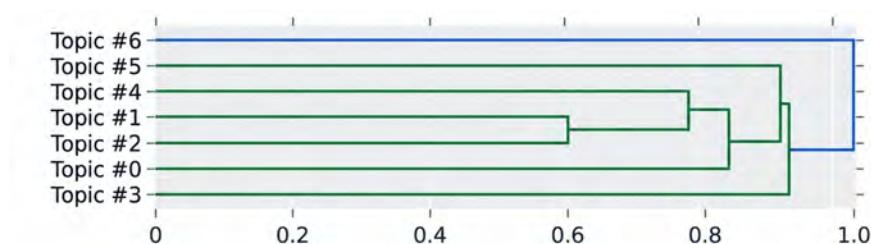
It is important to note the significant disparity in the volume of scraped reviews among the selected brands, with Johnnie Walker contributing a substantially larger portion of the dataset. Rather than a methodological flaw in sampling, this imbalance is a direct, objective reflection of the e-commerce reality during the observation period. The overwhelming data volume for Johnnie Walker is largely attributed to the “Fan Economy” phenomenon—specifically, the intense consumer engagement and concentrated purchasing bursts driven by their celebrity ambassador campaign. Consequently, this disparity itself serves as a key empirical finding, highlighting how celebrity-driven marketing can fundamentally disrupt and dominate organic brand discourse in the digital marketplace, which will be discussed more later.

## 4.2 Correlation and Thematic Relationships

To further understand how these identified topics interact within the consumer's mind, a hierarchical clustering analysis was conducted on the semantic relationships between the topics. The resulting dendrogram (Figure 2) reveals the semantic distances and co-occurrence patterns of the extracted themes, leading to two highly significant findings regarding consumer perception in the Chinese e-commerce whiskey market.

**Figure 2**

*Hierarchical clustering dendrogram*



*Note.* This diagram illustrates the semantic distances between the extracted topics. Please refer to Table 1 for the detailed content of each topic.

### 4.2.1 The Convergence of Quality and Logistics (“Quality = Packaging”).

The most prominent cluster in the analysis occurs between Topic #1 (Core Product Quality: liquid, palate) and Topic #2 (Unboxing & Logistics: packaging, delivery, intactness). These two topics merge at the closest semantic distance (approximately 0.6), indicating a high frequency of co-occurrence in consumer reviews. This suggests a critical behavioral insight: for e-commerce whiskey buyers, the objective quality of the liquor is inseparable from the logistical experience. A premium product is expected to be accompanied by flawless packaging and secure delivery. If the packaging is damaged or the unboxing experience is poor, the consumer's perception of the product's core quality is directly compromised.



### 4.2.3 The Dichotomy of E-commerce Discourse: Products vs. Personas.

The co-occurrence network analysis (Figure 3) further clarifies the semantic landscape by visualizing the core themes around distinct high-density hubs. The visualization reveals a clear dichotomy between the traditional product-oriented discourse and the emerging celebrity-driven discourse.

On one side of the network, a dense cluster around “packaging (包装)”, “liquid (酒液)”, “quality (品质)”, and “palate/taste (口感)” emerges. This area, marked by extensive and tight semantic links, confirms the previous finding: the physical condition of the product upon arrival and its inherent taste profile are intertwined in the consumer's perception of quality in an e-commerce context. These connections represent the rational evaluation process of a typical whiskey consumer focused on intrinsic product attributes.

In contrast, a visually isolated and exceptionally dense cluster forms around “Jianci Tan (檀健次)”. This cluster is characterized by high-weight connections among terms such as “ambassador (代言人)”, “support (支持)”, “exclusive merchandise/glasses (周边/杯子)” (In this case, the “exclusive merchandise” is a glass with a print of Jianci Tan’s hand, which will be given as a gift alongside the purchase), and “Black Label (黑牌)”. This separation confirms the visual representation from the hierarchical analysis (Figure 1).

Crucially, the massive size and monolithic density of the “Jianci Tan” cluster are a direct reflection of the data structure. Due to the campaign-driven nature of the Fan Economy, a disproportionately large volume of reviews is highly standardized and repetitive in content (e.g., standard scripts, such as “support our endorser Jianci Tan” (支持代言人檀健次)). These comments dominate the corpus, largely attributed to Johnnie Walker. This visual phenomenon

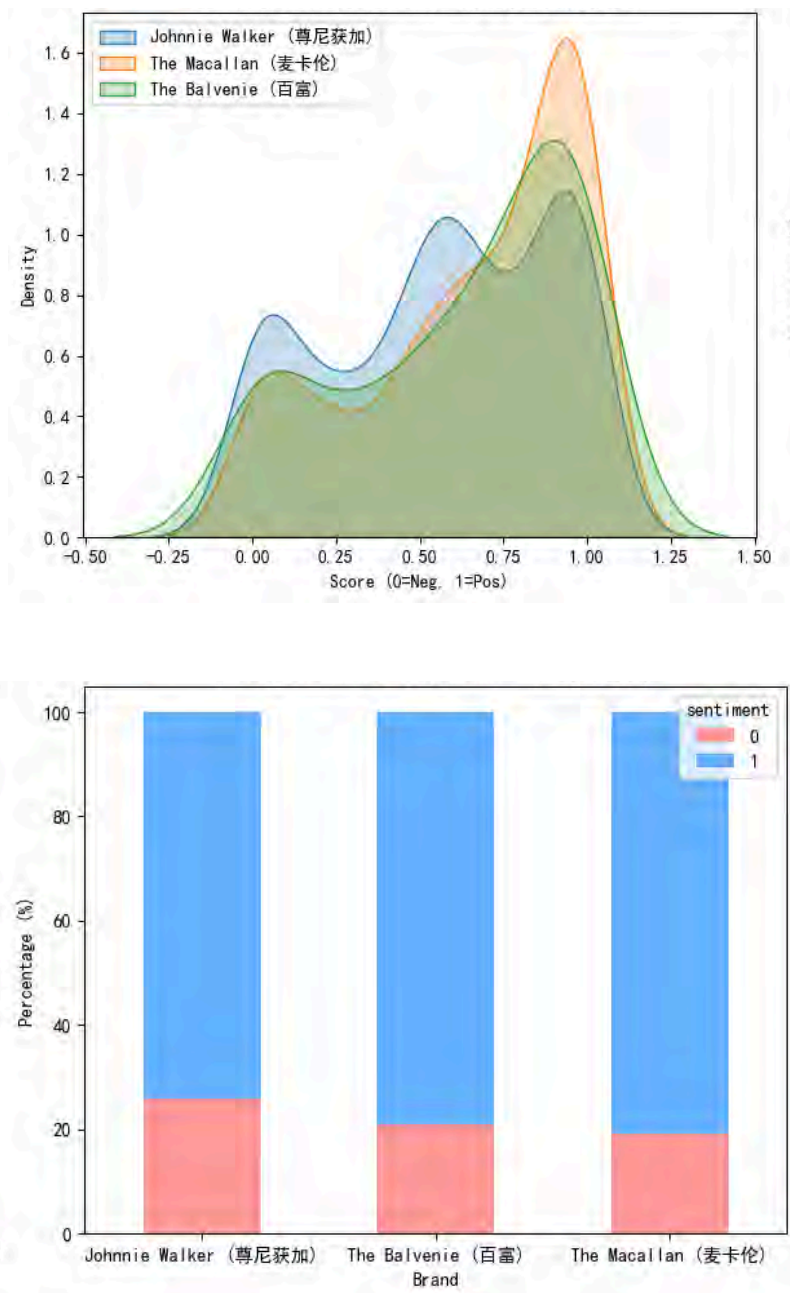
essentially maps the ‘parallel narrative’ of the Fan Economy consumption: a massive parallel world of consumption driven by emotional attachment to a celebrity and the pursuit of exclusive merchandise, rather than a deep engagement with the intrinsic quality of the whiskey itself.

### **4.3 Sentiment Distribution and the Marketing Volume Paradox**

The sentiment score density (Figure 4, up) demonstrates that while all three brands exhibit bimodal distributions (peaks near both 0.0 and 1.0), Johnnie Walker displays a considerably higher density peak in the extreme negative region (approaching score 0.0) compared to The Macallan and The Balvenie. This visual observation is quantified in the stacked bar chart (Figure 4, down), revealing that Johnnie Walker holds the highest proportion of negative reviews (approximately 26.1%), while The Macallan enjoys the highest positive ratio (80.7%).

**Figure 4**

*Sentiment distributions comparing among and within the three selected brands*



*Note.* This diagram illustrates a paradox between massive marketing volume and high negative sentiment ratios. 1 stands for positive feedback, and 0 stands for negative feedback.

This finding is particularly intriguing when cross-referenced with the previous thematic network analysis. Although Johnnie Walker's discourse is dominated by the massive and positive "Fan Economy" cluster (driven by celebrity endorsement), this sheer volume of positive standardized scripts failed to insulate the brand from a higher rate of severe dissatisfaction.

The density distribution offers an explanation: while Johnnie Walker's massive fan reviews are technically categorized as positive, they tend to be standardized, low-sentiment-density expressions (e.g., standard "support" scripts as discussed), contributing to a more diffused positive peak compared to The Macallan's highly concentrated positive sentiments. Looking more into the raw data, more clues can be found. Some comments were related to the missing of the expected merchandise, for example, "There's no Jianci Tan's photo card...I'm really disappointed...And they said they cannot deal with this...", and "There's no packaging box. And no promised gift. I bought this only to support Jianci Tan. I don't even drink this type of alcohol. That's so lame."

In conclusion, the results suggest that Johnnie Walker's celebrity-driven campaign attracted a vast and heterogeneous audience, including many casual e-commerce buyers. When logistics or service issues (identified in Topic #2 and Topic #5, such as damaged merchandise or subsidy disputes) occurred under such high-traffic conditions, they triggered intense and concentrated negative feedback from practical-minded consumers. Essentially, while the Fan Economy generated high noise, it objectively creates logistical complexity and also increases the brand's vulnerability to the service-based negative sentiment.

#### 4.4 Brand-specific Keyword Analysis and Strategic Alignment

**Table 2**

*Brand-Specific Keyword Extraction and Strategic Alignment*

<b>Brand</b>	<b>Core Chinese Keywords (extract)</b>	<b>English Translation &amp; Strategic Insight</b>
The Macallan	品质, 质量, 口感	“Quality & Palate” (Product-Driven): Reviews are highly concentrated on the liquid's objective quality and tasting notes. Indicates a mature, product-focused consumer base.
The Balvenie	赠品, 活动, 值得	“Gifts & Promotions” (Value-Driven): Keywords highlight campaigns and extra merchandise. Indicates a highly price-sensitive audience looking for the best “deal”.
Johnnie Walker (Black Label)	檀健次, 支持, 代言人, 杯子	“Idol & Merch” (Fan Economy-Driven): The discourse is entirely hijacked by the celebrity spokesperson (Tan Jianci) and exclusive merch (Glasses). Product attributes like “Taste” are heavily overshadowed.

*Note.* Extracted keywords reveal the primary drivers of consumer discourse for each brand, highlighting the contrast between product-driven, value-driven, and fan-economy-driven purchasing behaviors.

To further contextualize the findings, a brand-specific keyword extraction was conducted (Table 2) to evaluate how effectively each brand’s intended market positioning translates into actual consumer discourse within the e-commerce environment. The results reveal varying degrees of alignment.

#### **4.4.1 The Macallan: Congruence with Premium Positioning.**

For The Macallan, a brand traditionally positioned in the ultra-premium and luxury single malt segment, consumer discourse remains highly product-driven. Keywords such as “quality (品质/质量)” and “palate (口感)” dominate the semantic space. This indicates a strong alignment between the brand's heritage-focused positioning and consumer perception. Buyers of The Macallan in this dataset represent a more mature, traditional whiskey consumer base that prioritizes the intrinsic, objective quality of the liquid over peripheral marketing tactics.

#### **4.4.2 The Balvenie: The E-commerce Promotion Paradox.**

The Balvenie, while also a premium single malt known for its artisanal craftsmanship, presents a slightly different consumer profile in the e-commerce context. The prevalence of keywords like “gifts (赠品)”, “promotional events (活动)”, and “worth it (值得)” suggests a heavily value-driven audience. While the brand maintains its high-quality image, the actual conversion trigger on digital platforms appears tightly linked to promotional mechanics and perceived deal value. This highlights an e-commerce paradox where even craft brands are subject to the highly price-sensitive and promotion-oriented nature of the platform's ecosystem.

#### **4.4.3 Johnnie Walker Black Label: The Blurred Brand Narrative.**

The most striking deviation occurs with Johnnie Walker. As a globally recognized blended Scotch, its traditional positioning revolves around accessibility and the “Keep Walking” lifestyle. However, the data reveal that its e-commerce discourse has been almost entirely overtaken by its localized marketing strategy. The semantic landscape is saturated with terms like “Jianci Tan (檀健次)”, “support (支持)”, “ambassador (代言人)”, and “glasses/merch (杯子)”.

This demonstrates a clear Fan Economy-driven purchasing behavior. The celebrity endorsement successfully generated massive visibility and sales, but strategically, the brand narrative was effectively “hijacked” by the idol's persona. Intrinsic product attributes such as taste or heritage were heavily overshadowed by the desire to collect exclusive merchandise and demonstrate loyalty to the celebrity. This confirms that in the Chinese e-commerce ecosystem, leveraging top-tier idols can create explosive short-term conversion, but it fundamentally shifts the consumer's focus away from the core product.

## **5 Strategic Suggestions**

### ***5.1 Elevating Logistics to a Core Product Attribute (“The First Touch”)***

Based on the findings: Topic #1 and Topic #2 clustering (Quality = Packaging). In the e-commerce environment, a whiskey's quality is no longer judged solely by its liquid. The unboxing experience and the physical integrity of the delivery constitute the consumer's first touch. Brands must stop treating logistics as a mere operational backend and start treating it as a core product feature. Strategic investments should be directed toward premium, shock-proof packaging that prevents leakage and damage, as well as elevating the aesthetic unboxing experience to match the premium nature of the spirit. A flawless delivery is the prerequisite for a positive taste evaluation.

### ***5.2 Mitigating the “Double-Edged Sword” of the Fan Economy***

Based on the finding: The Johnnie Walker sentiment paradox and narrative hijacking. While celebrity endorsements (like Jianci Tan) generate explosive short-term traffic and sales, they risk shadowing the brand narrative and exposing the brand to intense negative sentiment

from operational bottlenecks. To harness this traffic safely, brands should adopt a “Segmented SKU Strategy.” Exclusive merchandise and idol-related bundles should be separated from core heritage product lines. Furthermore, before launching a Fan Economy campaign, brands must preemptively scale up their customer service and logistics capacities to handle the massive traffic spike, thereby preventing the severe post-purchase dissatisfaction observed in the data.

### ***5.3 Value-Added Promotions over Price Degradation***

Based on the finding: The Balvenie’s e-commerce paradox and Topic #5 (Price Sensitivity). Chinese e-commerce consumers are highly promotion-driven, even within the premium single malt category. However, participating in direct price-slashing (e.g., Mega-Subsidies) can fatally dilute a luxury brand’s equity and trigger refund disputes. Brands should pivot from direct discounting to “Value-Added Promotions” (e.g., offering high-quality branded glassware or tasting miniatures as gifts). This satisfies the e-commerce consumer’s psychological need for a “deal” while preserving the premium price point and brand prestige.

## **6 Limitations and Future Research**

### ***6.1 Sample Selection and Category Mismatch***

The first limitation pertains to the product selection. To maintain a focused scope, the study selected only one specific entry-level SKU per brand. Consequently, the extracted consumer perceptions cannot fully represent the overarching brand equity or the entirety of the brands' product portfolios. Furthermore, the selected SKUs include a mix of categories (two Single Malts vs. one Blended Scotch) with slight price variations. Because these are strictly entry-level products, the findings primarily reflect the behaviors of mass-market or novice

consumers, excluding the nuanced purchasing motivations of whiskey connoisseurs or high-end collectors. Future research should aim for stricter categorization matching (e.g., comparing 12-year Single Malts exclusively) and expand to premium tiers.

## **6.2 Data Volume Disparity and Topic Skewing**

As noted in the methodology, the scraped dataset is heavily unbalanced. Johnnie Walker’s review volume significantly outnumbered the others due to the explosive traffic generated by the Jianci Tan campaign. While this accurately reflects the reality of the “Fan Economy,” it caused the celebrity-related keywords to form a massive, isolated topic (Topic #3) that heavily skewed the overall corpus. This dominance may have inadvertently suppressed quieter, organic conversations regarding the intrinsic qualities of the other two brands.

## **6.3 Platform Noise and the “Fake Review” Phenomenon**

A common challenge in e-commerce text mining is the prevalence of “brushing scams (刷单)” or artificially generated positive reviews designed to boost store rankings on platforms like Taobao. It is highly likely that the dataset contains a certain percentage of these fake positive verbatim. The most critical operational insights—specifically the correlation between packaging and perceived quality (Figure 1), and the sentiment paradox regarding logistics and customer service (Figure 3)—were primarily derived from the specific, detailed, and often negative consumer feedback. However, logically, the core strategic significance of this study remains robust. Because artificially generated reviews are overwhelmingly generic and positive, their presence does not invalidate the structural findings related to consumer pain points and operational bottlenecks.

## 7 Conclusion

This study utilized BERTopic modeling and sentiment analysis on 3,500 Taobao consumer reviews to decode the purchasing motivations and post-purchase sentiments surrounding three entry-level Scotch whiskies (Johnnie Walker, The Macallan, and The Balvenie) in China. Despite the methodological boundaries previously discussed, the research yields robust, actionable insights for international spirits brands navigating the digital marketplace.

The empirical findings directly inform three strategic imperatives:

First, the semantic convergence of liquid quality and delivery dictates that consumers no longer separate the physical product from its logistical handling. Consequently, brands must elevate logistics and premium unboxing to core product attributes to protect the spirit's perceived value.

Second, the analysis revealed a “Fan Economy” paradox: while celebrity endorsements generate explosive traffic, they blur the brand narrative and expose the company to concentrated negative sentiment due to operational strain (e.g., missing merchandise). To safely harness this traffic, brands are advised to adopt a segmented SKU strategy, isolating idol-driven bundles from heritage product lines to safeguard long-term brand equity.

Third, the prevalence of price-sensitive discourse highlights a highly promotion-driven e-commerce culture. While it's important to consider consumers' demand for cost-efficiency, companies should deploy value-added promotions instead of blindly engaging in brand-diluting price wars to satisfy the consumer's psychological demand for deals while preserving premium positioning.

Ultimately, this research provides a vital empirical framework for enterprise campaign management. It underscores that while aggressive localization and celebrity-driven traffic are powerful tools for market penetration in China, their success relies entirely on a brand's operational readiness. By continuously leveraging consumers' first-hand feedback to identify and resolve underlying service bottlenecks, brands can effectively translate short-term digital traffic into sustainable, long-term brand loyalty.

**Declaration of generative AI and AI-assisted technologies**

During the process of data analysis and writing, the author used Gemini 3 Pro to assist coding in Python and enhance readability. The author reviewed and edited the content and takes full responsibility for the content.

**Acknowledgements**

I would like to express my sincere gratitude to my thesis advisor, Professor Ying Lei, who supported me throughout the whole process with clear instructions on every step, encouraging me to grasp new skills and step out of my comfort zone. Without her help, I would not be able to complete a relatively solid research as my undergraduate graduation thesis.

I also want to thank the Business and Economics Honors Program at NYUSH, which allows me to allocate a whole year to have a taste of academia and research. Thank all of my classmates and Xinyi, the program assistant, for spending all the Friday mornings together.

Lastly, many thanks to myself. Thank you for applying and enrolling yourself in the program, even though it sounded so intimidating at the beginning. And thank you for trying out the quantitative research methods and finding out that it is not too late to pick up some useful coding skills. And very importantly, as the only marketing student in the program, you survived all these and defended the honor of your major.

## References

- An, Y., Oh, H., & Lee, J. (2023). Marketing Insights from Reviews Using Topic Modeling with BERTopic and Deep Clustering Network. *Applied Sciences*, 13(16), 9443. <https://doi.org/10.3390/app13169443>
- Daza, A., González Rueda, N. D., Aguilar Sánchez, M. S., Robles Espíritu, W. F., & Chauca Quiñones, M. E. (2024). Sentiment Analysis on E-Commerce Product Reviews Using Machine Learning and Deep Learning Algorithms: A Bibliometric Analysis, Systematic Literature Review, Challenges and Future Works. *International Journal of Information Management Data Insights*, 4(2), 100267. <https://doi.org/10.1016/j.ijime.2024.100267>
- He, H. (2026). Localization strategies and brand recognition of multinational firms. *Finance Research Letters*, 97, 109761. <https://doi.org/10.1016/j.frl.2026.109761>
- LBB Editorial. (2023, November 1). *Understanding the Blend of the Johnnie Walker Brand* | *LBBOnline*. <https://lbbonline.com/news/understanding-the-blend-of-the-johnnie-walker-brand>
- Lievonen, M. (2017). Consumer Emotions and E-commerce: A Literature Review. *BLED 2017 Proceedings*. <https://aisel.aisnet.org/bled2017/24>
- Maikalun Macallan weishiji: Weihe bei yu wei “weishiji zhi wang”? Tansuo qi liu da biao zhu [Macallan Whisky: Why Is It Hailed as the “King of Whisky”? Exploring Its Six Pillars]. (2025, August 3). Alcohol Please. <https://alcoholpleasehk.com/en/blogs/information/macallan-whisky-guide>
- Nan, D., & Xu, W. (2025). *Fan Economy in the Social Media Age: New Marketing Approaches*. 485–494. [https://doi.org/10.2991/978-94-6463-734-2\\_56](https://doi.org/10.2991/978-94-6463-734-2_56)

Rachmiani, R., Oktadinna, N., & Fauzan, T. (2024). The Impact of Online Reviews and Ratings on Consumer Purchasing Decisions on E-commerce Platforms. *International Journal of Management Science and Information Technology*, 4, 504–515.

<https://doi.org/10.35870/ijmsit.v4i2.3373>

*Unique among single malts: The Balvenie | William Grant & Sons.* (n.d.). Retrieved April 29, 2026, from

<https://www.williamgrant.com/gb/stories/the-balvenie-the-most-handcrafted-of-single-malts/>

Yingxue, L. (2025, May 7). POPULARITY OF NATION'S MATURING WHISKY INDUSTRY UNCORKED. *China Daily*.

<https://www.chinadaily.com.cn/a/202505/07/WS681ac931a310a04af22bde01.html>

Yewuyong Weishiji Zhinan. (2024, December 12). Sugelan weishiji chanye 2025 nian qushi zhanwang (shang): Zhuanxing qi jijiang lailin! [Scotch Whisky Industry 2025 Trend Outlook (Part 1): The Transition Period Is Coming!]

<https://ch-9.net/article/wOMTkCf1K5>

Zhao, H., Liu, Z., Yao, X., & Yang, Q. (2021). A machine learning-based sentiment analysis of online product reviews with a novel term weighting and feature selection approach.

*Information Processing & Management*, 58(5), 102656.

<https://doi.org/10.1016/j.ipm.2021.102656>