ECharts-Based Visualization of the Entire Operation Link for Homestay Platforms

Yihan Xu[1,*], Mengjia Zhang[1], Qingfeng Zhou[1,2]
[1]*Henan University of Technology, Zhengzhou, Henan, China*
[2]*Iflytek Co., Ltd., Hefei, Anhui, China*
** Corresponding Author*

Abstract: With the rapid development of the sharing economy, homestay platforms have become an important part of the tourism industry. Based on Airbnb user behavior data, this paper uses big data technology to conduct multi-dimensional analysis of user behavior. The Hadoop ecosystem is used for data storage and processing, Spark is used for data analysis, and DataX is used to achieve data synchronization from HDFS to MySQL. Finally, ECharts is used to achieve data visualization. The research deeply analyzes user behavior characteristics from dimensions such as user profiles, promotion channels, and seasonal preferences, providing data support for homestay platforms to optimize services and formulate precise marketing strategies. Experimental results show that this analysis method can effectively mine user behavior patterns and provide reliable basis for platform operation decisions.

Keywords: User Behavior Analysis; Big Data Analytics; Hadoop; Homestay Platform; Data Visualization; Echarts

1. Introduction

In recent years, driven by the vigorous development of the tourism industry and the popularization of the sharing economy model, homestay platforms have rapidly emerged globally as a crucial bridge connecting hosts and tourists. As a leading enterprise in the homestay industry, Airbnb operates in 191 countries worldwide and has accumulated massive volumes of user behavior data. These data contain rich commercial value; in-depth analysis can reveal users' behavioral patterns and preference characteristics, providing important references for platforms to optimize services and enhance user experience.

Traditional data analysis methods face numerous challenges when processing massive user behavior data, including issues related to data storage, processing efficiency, and visual presentation. Propelled by the wave of digitalization [1], big data technology, characterized by volume, variety, and velocity [2], has offered a new approach to addressing the aforementioned problems. By leveraging big data technology [3], this study constructs a visual management system that presents complex data in an intuitive and understandable visual format [4], improving the efficiency of information processing [5]. It assists managers in quickly identifying abnormalities for precise intervention, provides a scientific basis for management decisions, and enhances both management efficiency and user experience [6].

Based on the Hadoop ecosystem, combined with the Spark computing framework and ECharts visualization tool, this paper establishes a comprehensive user behavior analysis scheme for homestays, aiming to provide data support for the refined operation of homestay platforms.

2. Visualization Analysis Process

(1) Data Collection and Preprocessing: Acquire Airbnb user behavior data from Kaggle, including basic user information, booking records, and browsing behavior. Implement essential preprocessing steps—data cleaning, deduplication based on user IDs, and targeted missing value handling (imputation or elimination)—to ensure data accuracy and reliability for subsequent analysis.

(2) Data Storage and Management: Store the cleaned data in HDFS (Hadoop Distributed File System) to leverage its scalability and fault tolerance. Build a structured data warehouse with Hive, enabling organized data management and efficient querying to support analytical needs.

(3) Data Analysis and Computing: Utilize Spark for distributed computing to process large

datasets efficiently. Conduct multi-dimensional in-depth analysis, covering user profiles, promotion channel effectiveness, and seasonal booking preferences.
(4) Data Synchronization and Presentation: Sync analysis results from HDFS to MySQL using DataX for stable cross-system data transmission. Provide data access via Java Web APIs and visualize key insights (e.g., conversion rates, behavior trends) with ECharts for intuitive presentation.

3. Overview of Homestay User Data

3.1 Data Analysis Requirements
(1) User Profile Analysis: Distribution characteristics of user gender, age, and region; proportion of bookings made by users from China to foreign countries.
(2) Promotion Channel Analysis: Monthly new users, registration volume across different user terminals, registration volume through different promotion channels, registration volume from different marketing content, conversion rates of different promotion channels, and conversion rates of different marketing content.
By analyzing basic user information such as age, gender, and device usage, we can understand user profiles. By studying users' accommodation preferences across different seasons and countries, we can identify travel behavior patterns. Through data visualization technology, analysis results are presented intuitively to facilitate understanding and decision-making for platform operators.
(3) User Experience and Interface Design: Airbnb has made significant efforts in product experience and listing aesthetics, but there is still room for improvement. We can analyze users' experience during search, booking, payment, and communication processes to identify areas that need simplification or optimization.
(4) Community and Trust Building: As a community platform, trust is one of the key factors for Airbnb's success. Analyze whether the interaction, evaluation, and trust-building mechanisms between users are sufficiently sound.
(5) Market Expansion and Globalization Strategy: Airbnb operates in 191 countries, but may face different challenges and needs in various markets. We can analyze cultural differences, legal and regulatory requirements, and market competition among users in different regions.
(6) Support and Services for Hosts: Airbnb's success is largely attributed to its ability to attract and support a large number of hosts. Analyze the needs and challenges hosts face when registering, managing, and marketing their listings. Consider providing more convenient listing management tools, optimizing the communication platform between hosts and guests, and offering better customer support and feedback mechanisms to enhance host satisfaction and activity.

3.2 Data Source
The data is sourced from the Airbnb New User Bookings dataset provided by the Kaggle platform, including fields such as user ID, account creation date, first booking date, gender, age, registration method, device type, marketing channel, and destination country. After cleaning, the dataset contains approximately 200,000 valid records. Basic information and consumption type data of homestay users are shown in Table 1.

Table 1. Display of Homestay User Data

Field Type	Meaning
id	User ID
date_account_created	Account creation date
date_first_booking	First booking date
gender	Gender
age	Age
signup_method	Registration method
signup_flow	User registration page
language	Language preference
affiliate_channel	Marketing channel
affiliate_provider	Marketing source provider
first_affiliate_tracked	First marketing ad before registration
signup_app	Registration source app
first_device_type	Device type at registration
first_browser	Browser at registration
country_destination	Destination country
action	Tracking point name
action_type	Operation event type
action_detail	Operation event description
device_type	Device type

3.3 Development Tools and Programming Languages
IntelliJ IDEA, as a powerful integrated

development environment, provides comprehensive support for project development, debugging, and deployment. Combined with the Hadoop high-availability cluster and HDFS distributed file system, a stable and reliable big data storage and computing infrastructure is constructed. Tomcat 9.0 serves as a lightweight Web application server, providing an efficient operating environment for the system. At the data processing level, Spark achieves efficient data analysis and processing through its excellent in-memory computing capabilities, Hive offers an SQL-based data warehouse solution, and DataX realizes efficient data synchronization between HDFS and relational databases.

The entire project integrates multiple programming languages and adopts a front-end and back-end separation architecture. The front-end uses HTML to build the interface foundation, JavaScript to implement interactive logic, and ECharts 5.4 to achieve multi-dimensional data visualization. ECharts provides a variety of chart types, such as line charts, bar charts, pie charts, and maps, which can meet diverse data visualization needs [7]. In practical applications, the principle of "accuracy > clarity > aesthetics" should be followed [8]. The back-end is developed and deployed based on the Java framework [9], capable of processing requests from the front-end and returning corresponding data [10].

4. Data Processing

4.1 Data Cleaning

Spark is used for raw data cleaning, with core tasks including: removing duplicate records based on user IDs; handling missing values for fields such as age, gender, and first booking date through imputation or elimination (see Table 2 for details on missing values in the dataset); filtering out records where ages fall outside the reasonable range (e.g., under 11 or over 80); and standardizing fields like dates and country codes.

Table 2. Description of Missing Values in Fields

Field	Description of Missing Values
date_first_booking	First booking date: 124,544 missing values
gender	Gender: 95,688 missing values
age	Age: 87,991 missing values
first_affiliate_tracked	Registration channel: 6,065 missing values
first_browser	Registration browser: 27,266 missing values
action_type	Operation type: 1,126,204 missing values
action_detail	Operation detail: 1,126,204 missing values

4.2 Data Segmentation

According to analysis requirements, the data is divided into multiple subsets: Classifying users into churned users and retained users based on whether they have completed their first booking; dividing first booking behaviors by four seasons (spring, summer, autumn, winter); and conducting grouped statistics by destination country. The attribute information of churned customers is also analyzed.

5. Visualization Analysis

5.1 Channel Customer Acquisition: Scientifically Positioning High-Value Traffic Entries

In the initial development stage of a homestay platform, it is of strategic importance to acquire customer traffic through diversified channel placements and to identify channels with excellent conversion effects for additional capital investment. Therefore, we utilized bootstrap combined with ECharts to analyze the source channels of new users and their consumption willingness (specifically shown in Figure 1 and Figure 2), aiming to accurately select suitable promotion channels and assist the platform in efficient customer acquisition and resource optimization.

Figure 1 and Figure 2, through the linkage analysis of the three dimensions "Number of Users - Number of Orders - Conversion Rate," construct a complete logical chain for channel effectiveness evaluation: from "Traffic Acquisition" to "Transaction Conversion," and then to "Efficiency Verification." This provides full-process data support for channel placement decisions based on "Traffic Scale - Monetization Capability - Input Efficiency," avoiding decision-making biases caused by single indicators. Data linkage allows for the derivation of a channel tiering strategy: Key investment channels: 'limited' (large user base, high order volume, high conversion rate), 'marketing' (extremely high conversion rate, consider increasing traffic allocation); Maintenance/Observation channels: 'omg tracked' (large user and order volume, but conversion rate slightly lower than 'limited', optimize placement strategy to improve efficiency); Reduce/Optimize channels: 'other', 'product', 'local ops' (small user base, low order volume, or low conversion rate, need to evaluate whether to reduce investment or optimize placement methods).

Figure 1. Multi-Channel Customer Acquisition Chart

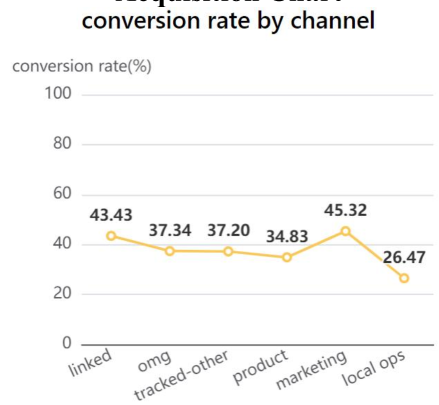

Figure 2. Multi-Channel Conversion Chart

5.2 Customer Profile: Precisely Outlining High-Value User Characteristics

In the homestay industry, customer profile analysis holds key value for precision operations, covering dimensions such as age, gender, nationality, consumption level, and behavioral characteristics. This paper primarily conducts an in-depth analysis of customer attributes based on age and country to clarify the characteristics of the target customer group. Following the customer acquisition phase, formulating customer retention strategies is equally indispensable; furthermore, the scale of high-quality customers is a core factor affecting the long-term development of homestays.

Building a customer geographic-demographic insight system by integrating Figure 3's heatmap (age-country cross-distribution) and Figure 4's customer map (spatial distribution) transforms abstract customer data into intuitive spatial and demographic visualizations. This system serves as a key analytical tool bridging "Customer Insight" and "Business Implementation." Figure 3 reveals the U.S. as the core homestay customer source, an absolute primary market with robust demand; Australia maintains a steady but significantly smaller customer base compared to the U.S.; while Europe, Asia, and other regions have not yet formed effective customer scales, remaining undeveloped markets. Furthermore, the customer base is predominantly young adults aged 21-40, with stable coverage of middle-aged groups, forming a "mature market with broad age distribution." Targeted youthful, social homestay products and personalized marketing activities can thus be designed to resonate with the 21-40 core age group.

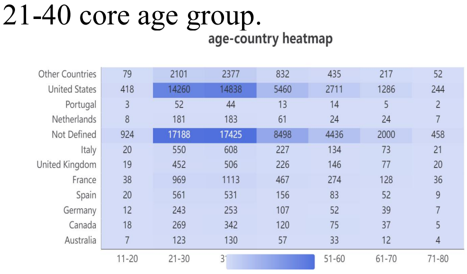

Figure 3. Age-Country Distribution Heatmap

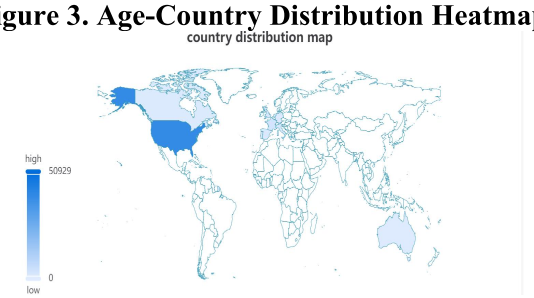

Figure 4. Country Distribution Map

5.3 Tackling Conversion Challenges: Deconstructing Full-Process Bottlenecks for a Profit Breakthrough Path

For homestay platforms with profitability as a core goal, the conversion gap between registered users and paying customers is a key indicator reflecting the platform's operational efficiency and commercial value. Conducting an in-depth analysis of the full "Registration --- Payment" conversion path, precisely identifying core bottlenecks (such as post-registration decision hesitancy, cumbersome booking operations, or inconvenient payment processes), and exploring user behavioral preferences and unmet pain points are crucial for optimizing platform functions, refining operational strategies, and effectively improving conversion efficiency.

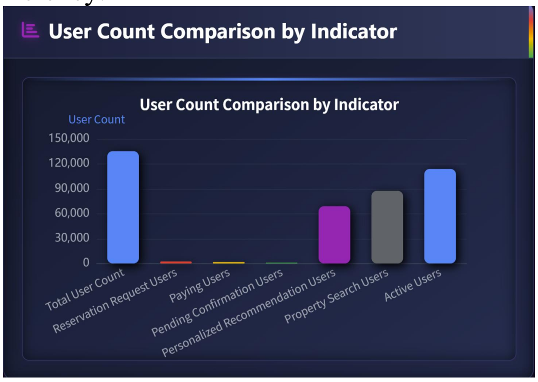

Figure 5. User Distribution across Processes Chart

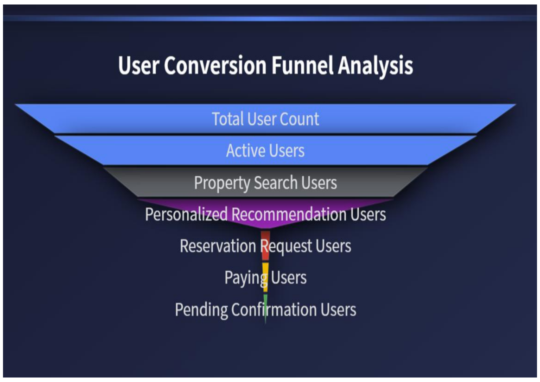

Figure 6. User Conversion Funnel Analysis Chart

In terms of data implications, Figure 5 intuitively shows that the total user base and active users have a solid foundation in scale, but the number of users in conversion stages like "Application for Reservation, Payment, Pending Confirmation" plummets drastically. The funnel model in Figure 6 further visualizes the step-by-step attrition throughout the "Registration-Payment" full chain. Among these, insufficient user activation from "Registration-Active," the decision-making conversion shortcoming from "Active-Reservation," and trust/process obstacles from "Reservation-Payment" are the core bottlenecks restricting conversion efficiency. Based on this, hierarchical optimization strategies need to be formulated for different stages: At the user activation layer, enhance the activity of registered users through personalization and newcomer benefits; At the decision-making conversion layer, optimize property listing display, recommendation algorithms, and the reservation process to reduce user decision-making and operational costs; At the payment trust layer, simplify the payment path and strengthen service trust building. This series of measures serves as both the core lever for efficiently converting registered users into paying users and enhancing platform profitability, and the key path for ensuring the long-term sustainable development of the platform.

5.4 Operational Experience: Optimizing High-Frequency Scenarios as a Payment Conversion Accelerator

This paper focuses on user operation behaviors, addressing pain points such as "cumbersome and time-consuming operations, frequent functional updates causing interference, core operation crashes." Combined with the analysis of Figure 7 "Relationship between Operation Count and Average Duration," the aim is to optimize the efficiency of high-frequency operations, reduce user decision-making costs, and promote smoother payment conversion. From the results perspective, using a bubble chart to visualize the correlation between "Operation Count - Average Duration" establishes a quantitative analysis framework for platform operation experience. It transforms abstract user operation behaviors into an intuitive data mapping of "Frequency - Time Consumption," providing a clear visual decision-making basis for identifying "high-frequency but low-efficiency" operations and locating experience bottlenecks. This assists the platform in deep optimization from the operational level towards user experience and commercial conversion.

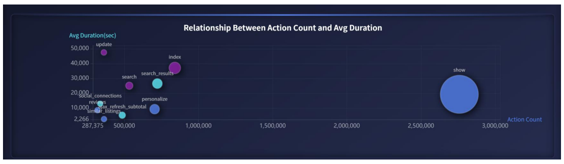

Figure 7. Operation Count vs. Average Duration Chart

Regarding the data-level significance of Figure 7: High-Frequency & Efficient Operations (High operation count + Short average duration): Represented by the "show" operation, with an operation count close to 3,000,000 and an average duration significantly lower than other operations. These operations represent the platform's core and efficient user interaction scenarios, indicating that their process design and response speed have reached a relatively good experience standard. They can serve as reference benchmarks for optimizing other operations, while their stability needs to be ensured to avoid crashes affecting the key user experience before payment decisions. High-Frequency & Low-Efficiency Operations (Relatively high operation count + Relatively long average duration): Such as "search_results", "index" operations, with operation counts in the high range and relatively long average durations. These operations belong to high-frequency user scenarios (e.g., search result loading, homepage indexing), but the time consumption issue can easily trigger user frustration, representing the core sticking point for operational experience optimization – technical optimization (e.g., caching strategies, page lightweight) is needed to reduce time consumption, avoiding decreased user payment willingness due to cumbersome operations. Low-Frequency but Time-Consuming Operations: Such as the "update" operation, with an average duration exceeding 40,000 seconds but a low operation count. Although these operations are used infrequently, excessively long single operation times can severely impact users' perception of the platform's technical stability. It is necessary to investigate whether there are functional redundancies or technical flaws, avoiding damage to the user's overall trust in the platform due to the poor experience of individual operations.

6. Conclusion

Based on the data analysis and visualization research of the Airbnb homestay platform, this paper conducts in-depth analysis from four dimensions: channel acquisition, user portrait, conversion optimization, and operation experience. Through a complete logical chain of "scientifically locating high-value traffic sources, accurately outlining high-value user characteristics, deciphering full-lifecycle conversion bottlenecks, and optimizing payment experience in high-frequency scenarios", it not only achieves strategic breakthroughs in the operational development of homestay platforms (such as improved customer acquisition efficiency, precise user operation, optimized conversion paths, and upgraded user experience) but also forms quantifiable commercial gains in economic value dimensions (such as improved ROI of resource investment, increased payment conversion rate, and extended user lifetime value). This study provides a full-lifecycle solution with both practical guidance and commercial value for the refined operation and profit growth of homestay platforms.

References

[1] JIN Ting. Dilemmas of Higher Vocational Education under the Background of Digital Transformation and Narrative Empowerment. Chinese Vocational and Technical Education, 2024(33): 59-66.
[2] MENG Xiaofeng, CI Xiang. Big Data Management: Concepts, Technologies and Challenges. Journal of Computer Research and Development, 2013, 50(1): 146-169.
[3] ZHANG Ming, ZHANG Yichun. Research on Constructing Teaching Quality Monitoring System in Higher Vocational Colleges Based on Big Data Technology. Chinese Vocational and Technical Education, 2021(35): 19-23.
[4] SHEN Enya. Big Data Visualization Technology and Its Application. Science & Technology Review, 2020, 38(3): 68-83.
[5] MENG Qinyuan. Research on Information Visual Communication Design Based on the Background of Chinese Intangible Cultural Heritage Inheritance. Screen Printing Industry, 2024(3): 71-73.
[6] LI You, WU Zigu-li. Application of Human-Computer Interaction Technology Based on Information Visualization in Visual Communication Design. Automation & Instrumentation, 2024(2): 182-186.
[7] REN Ni, WU Qiong, LI Huiquan. Analysis and Research on Data Visualization Technology. Electronic Technology & Software Engineering, 2022(16): 180-183.
[8] LIAO Z K, ZHANG C C. Generative AI makes for better scientific writing but beware the pitfalls. Nature, 2024, 631(8021): 505.
[9] FAN Luqiao, GAO Jie, DUAN Banxiang. Domestic Popular Tourist Attractions Data Visualization System Based on Python+ Flask+ ECharts. Modern Electronics Technique, 2023, 46(9): 126-130.
[10] CHEN Jiafa, HUANG Yujing. Application of Flask Framework in Data Visualization. Fujian Computer, 2022, 38(12): 44-48

