

Construction of Segmented Market User Personas and Demand Analysis Based on SQL

Wankun Shang¹

¹ Nke Safety Apparel, United States

Correspondence: Wankun Shang, Nke Safety Apparel, 16616 Chandler RD East Lansing, MI 48823, United States.

Received: November 29, 2025; Accepted: December 21, 2025; Published: December 22, 2025

Abstract

In the context of the digital economy, enterprises are gradually shifting their understanding of segmented market user demand from experience-driven approaches to data-driven methods. User personas, as an important tool linking user behavior data with business decision-making, play a key role in refined operations and market analysis. Taking business data stored in relational databases as the research object, this study constructs segmented market user personas based on SQL technology and conducts user demand analysis on this basis. First, combining user persona theory with segmented market theory, the dimensions of user characteristics and the indicator system are systematically designed. Then, SQL is used to clean, integrate, and aggregate multi-source data to extract key indicators such as demographic attributes, behavioral characteristics, and consumption preferences, thereby enabling the construction of segmented market user personas. Finally, by comparing the persona characteristics of different user groups, differentiated user demands are identified and corresponding business implications are summarized. The results show that the SQL-based user persona approach has advantages such as low implementation cost, strong operability, and ease of deployment. It can effectively support user demand analysis in segmented markets and provide data support for precise enterprise decision-making and resource allocation.

Keywords: SQL, segmented market, user persona, demand analysis, data-driven decision-making

1. Introduction

With the rapid development of information technology and the continuous accumulation of data resources, enterprises are gradually shifting from product- and channel-centered extensive management to user-centered refined operations. In this context, accurately understanding user characteristics and demands based on objective data has become essential for improving market responsiveness and decision-making quality. As market structures grow more complex and user demands become increasingly diversified, traditional approaches based on experience or simple statistical summaries are no longer sufficient to describe the real profiles of users in different segmented markets. This has created a strong demand for systematic and data-driven analytical methods. User personas serve as an effective tool for integrating multidimensional user information from different business systems into structured representations, thereby supporting user behavior analysis and demand identification. However, existing persona construction methods often rely on big data platforms or complex algorithmic models, which increase technical complexity and implementation costs and limit their adoption in small- and medium-sized systems. In contrast, SQL is a mature and widely used data processing language in relational databases, offering advantages such as stability, interpretability, and ease of maintenance, making it a practical solution for persona construction in real business environments. Although segmented market research has been widely discussed, studies focusing on SQL-based persona construction for demand analysis remain limited. Therefore, this study proposes an SQL-based approach to constructing segmented market user personas and analyzing user demands. By designing a structured indicator system and implementing data processing and feature extraction through SQL, this research provides a feasible and practical framework to support data-driven decision-making in enterprises.

2. Theoretical Foundations and Related Research Review

2.1 User Persona and Segmented Market Theory

User personas originated in the fields of precision marketing and user research. Their core purpose is to construct structured representations of user groups or individuals by systematically integrating multidimensional user characteristics, thereby supporting behavior prediction and decision-making. A user persona is not a simple aggregation of attributes, but a comprehensive description of users' overall characteristics in specific business

contexts, incorporating demographic information, behavioral patterns, preference features, and value attributes. With the development of data-driven analysis, user personas have gradually evolved from qualitative descriptions to quantitative, indicator-based representations that emphasize computability, interpretability, and continuous updating. Existing studies generally divide user personas into two main levels: basic attribute personas and behavioral personas. Basic attribute personas describe relatively stable characteristics such as age, gender, region, and occupation, reflecting users' fundamental social attributes. Behavioral personas focus on dynamic usage behaviors, including consumption frequency, usage timing, interaction patterns, and preferences, which more directly reflect actual user demands. In segmented market analysis, these two types of personas complement each other and enable a multidimensional understanding of user groups. Segmented market theory views markets as heterogeneous structures composed of users with different demands, behaviors, and value orientations. While traditional segmentation relies mainly on demographic or geographic variables, recent research increasingly incorporates behavioral data to support dynamic and behavior-based segmentation. User personas therefore function as a key technical carrier that transforms abstract segmentation concepts into quantifiable user characteristics. The integration of user persona theory and segmented market theory provides a solid foundation for data-driven demand analysis and supports subsequent persona construction and analysis using tools such as SQL[1].

2.2 Related Research on Data-Driven Demand Analysis

Data-driven demand analysis refers to the identification of user demand structures, change patterns, and driving factors based on user data, behavioral data, and business data through statistical analysis, data mining, and model inference. Its purpose is to support product design, operational strategy formulation, and resource allocation. Research in this field began relatively early in foreign studies, which initially focused on database marketing, customer relationship management, and user lifecycle value. These studies emphasized inferring user preferences from transaction records, access logs, and interaction behaviors to enable differentiated services and precision marketing. With the rapid expansion of internet platforms and mobile applications, user behavior data have become high-frequency, large-scale, and diverse. Consequently, research has shifted from descriptive statistics to more advanced behavioral modeling and predictive analysis, including path analysis to describe usage processes, clustering and association rules to identify behavior patterns, and classification or prediction models to assess conversion and churn risks[2]. This evolution has made data-driven demand analysis increasingly aligned with real business decision-making. Domestic research has also developed rapidly in recent years, particularly in fields such as e-commerce, social media, online education, and internet finance. Many studies focus on user segmentation, preference identification, recommendation optimization, and product iteration, highlighting the role of data analysis in improving user experience and operational efficiency. Methodologically, demand analysis is often divided into stages such as demand identification, classification, prediction, and strategy output, with techniques including statistical analysis, text mining, log analysis, and machine learning applied at different stages. Overall, both domestic and international studies show that data-driven demand analysis can overcome the limitations of traditional surveys in sample size, timeliness, and objectivity, and has formed relatively mature application paths. Despite this progress, notable gaps remain. Many studies rely on big data platforms or complex algorithmic frameworks, paying limited attention to relational database environments common in enterprises[3]. In addition, implementation details—such as data cleaning, integration, and indicator construction—are often insufficiently described. Given that many business systems still rely on SQL, systematic and practical SQL-based demand analysis frameworks are still lacking. Therefore, this study focuses on relational database scenarios and adopts SQL as the core tool to support segmented market demand analysis, aiming to provide a practical, interpretable, and easily deployable approach[4].

3. Data Sources and Research Methods

3.1 Data Sources and Data Structure Description

The data used in this study are obtained from relational databases generated during the routine operation of enterprises. These data include basic user information, business transaction records, and user behavior logs, all of which are automatically collected by operational systems. Such data are characterized by high continuity, strong authenticity, and stable update frequency, enabling them to reflect user behavior patterns and demand changes across different business scenarios in a comprehensive manner. Compared with questionnaire surveys or interview-based data, operational database data reduce subjective bias and are therefore more suitable for behavior-based user persona construction and segmented market demand analysis. From the perspective of data composition, the research data consist of three main categories[5]. The first category is basic user information data, typically stored in a user master table, which records users' registration details and fundamental attributes such as unique user identifiers, registration time, geographic region, and account status. These attributes are relatively stable and

provide the foundation for the basic attribute dimension of user personas. The second category is business transaction data, derived from order or transaction systems. These data record users' actual transaction behaviors, including transaction time, transaction amount, product or service type, and transaction frequency, and directly reflect users' consumption capacity and value characteristics. The third category is user behavior log data, mainly sourced from system access logs or behavior tracking records, which capture users' browsing, clicking, staying, and interaction behaviors on the platform and are used to describe usage habits and potential demands. In terms of data structure design, the different data categories are stored in multiple business tables and linked through a unique user identifier[6]. The user information table functions as the core dimension table, while transaction tables and behavior log tables record events at finer granularities. This structured design enables SQL-based join operations across tables, supporting the integration of user features from multiple behavioral and business dimensions. To ensure analytical validity, this study filters data fields to retain only those relevant to persona construction and demand analysis. By unifying user identifiers, standardizing time formats, and handling missing or abnormal values, a consistent and interpretable data structure is established, providing a reliable foundation for subsequent SQL-based data processing and feature extraction[7].

3.2 SQL-Based Data Processing Methods

In a relational database environment, SQL serves as the core tool connecting raw business data with analytical results. To ensure the scientific validity and reproducibility of user persona construction and demand analysis, this study adopts SQL as the primary data processing method to systematically perform data cleaning, integration, aggregation, and feature extraction. By explicitly expressing data processing logic through structured query statements, the transparency of the analytical process is enhanced, and the results can be more easily reused and maintained in practical business systems[8].

SQL Data Processing Methods

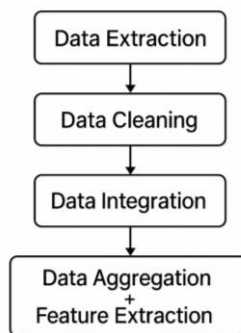


Figure 1. Data Cleaning Process

During the data cleaning stage, the main objective is to improve data quality and eliminate noise that may interfere with analysis results. SQL is used to handle missing values, outliers, and duplicate records. For example, conditional filtering is applied to remove records with missing critical information, reasonable thresholds are used to identify and exclude obviously abnormal transaction data, and potential duplicate user or behavior records are removed through deduplication operations. In addition, time fields are converted into a unified format and categorical fields are standardized to ensure semantic and structural consistency across different data tables, thereby creating favorable conditions for subsequent multi-table joins. In the data integration stage, the unique user identifier is used as the core key to combine the user basic information table, transaction data table, and user behavior log table through SQL join operations. By appropriately combining inner joins and outer joins, a unified analytical view of data from different sources is constructed, allowing each analytical record to simultaneously contain multidimensional information such as user attributes, transaction behaviors, and usage behaviors[9]. This relational structure-based integration approach helps maintain data completeness while avoiding redundant calculations, providing a stable foundation for comprehensive user feature analysis. In the data aggregation stage, SQL grouping operations and aggregation functions are fully utilized to conduct multi-level summary analysis of user behavior and transaction data. By aggregating indicators such as transaction count, consumption amount, and behavior frequency at the user level, high-frequency and discrete event records can be transformed into representative user-level features. At the same time, by incorporating time-based grouping, user behavior patterns across different time periods are extracted, reflecting the stage-specific characteristics of user demand. This transformation from detailed records to aggregated indicators is a critical step in the transition of user personas from the "data layer" to the "feature layer." In the feature extraction stage, aggregated statistical results are further

transformed into persona features based on business understanding and research objectives. Through SQL operations such as conditional statistics, window functions, and derived field calculations, indicators including user activity level, consumption preference, and usage stability are constructed, enabling user characteristics to be expressed in the form of structured labels or numerical values. Compared with feature generation methods based on complex models, this SQL-based feature extraction approach offers stronger interpretability, clearly reflecting the correspondence between indicators and user behaviors, and provides intuitive and reliable inputs for subsequent segmented market user persona construction and demand analysis[10].

4. Construction of Segmented Market User Personas

4.1 Design of the User Persona Indicator System

In segmented market analysis, the design of the user persona indicator system directly determines the completeness and interpretability of the persona results. Based on the previous discussion of user persona theory and data structure, this study follows a combined principle of business relevance, data availability, and indicator computability, and develops a user persona indicator system oriented to segmented market demand analysis. The system is built on data that can be stably obtained from relational databases, and indicators are calculated and updated through SQL, so that the persona results can achieve both sufficient analytical depth and strong practical feasibility. From an overall structural perspective, the indicator system is divided into three layers: the basic attribute dimension, the behavioral characteristic dimension, and the value characteristic dimension. These dimensions complement each other and jointly form a multidimensional description of users. Specifically, the basic attribute dimension mainly describes users' static characteristics and serves as the fundamental basis for market segmentation; the behavioral characteristic dimension focuses on users' dynamic behavior patterns during actual use and is used to reveal explicit and implicit demands; the value characteristic dimension evaluates user importance from the perspective of transactions and contribution and provides a reference for demand prioritization. The overall structure of this indicator system is shown in Figure 2.

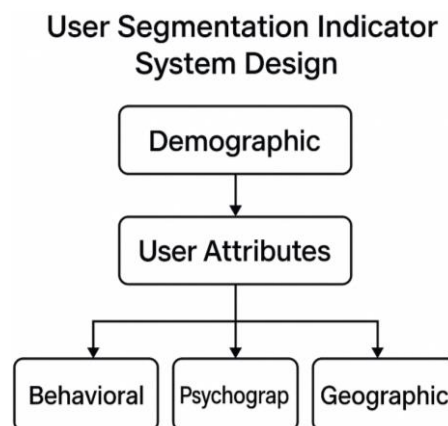


Figure 2. User Persona Indicator Design

As shown in Figure 2, the basic attribute dimension typically includes relatively stable information such as user registration time, geographic region, and account status. These indicators help identify structural differences among segmented markets and provide background conditions for subsequent behavioral analysis. The behavioral characteristic dimension is constructed based on user behavior logs and operational records, and it mainly reflects usage frequency, activity level, and behavioral preferences. This dimension is often the most explanatory part in demand analysis. By statistically summarizing and generalizing user behaviors, it is possible to identify common patterns across different segmented markets in terms of feature usage, access rhythm, and content preferences. The value characteristic dimension relies primarily on transaction data and measures the actual business value of users through indicators such as purchase frequency, cumulative spending, or contribution level, providing a quantitative basis for distinguishing high-value users from potential users. During indicator design, this study particularly emphasizes the correspondence between indicators and SQL implementation. Each type of persona indicator is defined with clear data sources and calculable logic, avoiding reliance on black-box models that are difficult to reproduce. For example, behavioral activity can be obtained by counting events within a specific time window, and consumption preference can be computed through the proportion of transactions across different product or service categories. This design approach, centered on SQL-computable indicators, makes the user persona system

not only theoretically reasonable, but also practically implementable, which is helpful for continuous iteration and application in real business systems. Overall, the user persona indicator system constructed in this study considers both static attributes and dynamic behaviors in its structure, and it makes full use of SQL-based data processing capabilities in its implementation path. It therefore provides a clear and stable indicator foundation for subsequent persona modeling and persona-based demand analysis in segmented markets.

4.2 SQL-Based Persona Modeling and Implementation

The user persona modeling in this study aims to build a user-level feature wide table, where basic user information, transaction records, and behavior logs are aggregated under the primary key `user_id`. The implementation idea is to first use JOIN operations to merge multi-source data into a unified analytical view, and then apply GROUP BY `user_id` to compress detailed events into user-level indicators (such as order count, cumulative spending, and the number of events in the last 30 days). At the same time, CASE WHEN is used for conditional statistics to generate interpretable labels (such as activity level and value level). For time-related features that require ordering, such as the most recent behavior time, a window function (e.g., ROW_NUMBER() or MAX()) is used to obtain the latest record, which complements persona information on the time-series dimension. The final output is an updatable persona table that contains both quantitative indicators and business-friendly label fields, and it can directly support subsequent segmentation and demand analysis. A simplified SQL pseudocode example is provided below (for illustration only):

```
SELECT
  u.user_id,
  u.region,
  COUNT(DISTINCT o.order_id) AS orders_cnt,
  SUM(o.amount) AS total_spend,
  SUM(CASE WHEN l.event_time >= NOW()-INTERVAL '30d' THEN 1 ELSE 0 END) AS events_30d,
  MAX(l.event_time) AS last_event_time,
  CASE WHEN SUM(o.amount) >= 10000 THEN 'High_Value' ELSE 'Normal' END AS value_tag,
  CASE WHEN SUM(CASE WHEN l.event_time >= NOW()-INTERVAL '30d' THEN 1 ELSE 0 END) >=
50
    THEN 'High_Active' ELSE 'Normal' END AS activity_tag
FROM user_table u
LEFT JOIN order_table o ON u.user_id = o.user_id
LEFT JOIN log_table l ON u.user_id = l.user_id
GROUP BY u.user_id, u.region;
```

5. Demand Analysis Based on User Personas

5.1 User Characteristics Across Different Segmented Markets

After constructing segmented market user personas, this study further takes the persona results as the analysis object and conducts a systematic comparison of common characteristics and differentiated characteristics across different segmented markets. The purpose is to reveal the underlying patterns of user behavior and preference structures. By comparing persona indicators, it can be observed that segmented markets show clear divergence in user activity, consumption behavior, and usage preferences, and such divergence directly reflects differences in user demands. From the perspective of overall common characteristics, users across segmented markets show a certain level of behavioral concentration. In other words, core behaviors and major consumption activities often revolve around a small number of functional modules or product categories. This indicates that user demands are not completely scattered even when the market is segmented, but instead there are relatively stable “main demand scenarios.” In terms of behavioral rhythm, many users show periodic characteristics over time. For example, activity levels rise significantly during certain periods or cycles, suggesting that user demand is influenced jointly by business scenarios and usage habits. These common features provide a unified direction for enterprises in basic function planning and service assurance. Regarding differentiated characteristics, different segmented markets show significant differences in consumption behavior and usage frequency. High-value user groups usually have higher purchase frequency and higher cumulative spending. Their behavior pattern is more stable, with higher repurchase rates, and they show stronger dependence on core products or services. In contrast, low-value or potential user groups more often display low-frequency use and intermittent consumption, and their behaviors are more clearly driven by promotions or short-term needs. From the perspective of usage frequency, highly active users tend to generate behavioral records continuously across multiple time windows, and their usage paths are relatively fixed. Less active users show a more scattered distribution of behaviors, and their demands are less stable. In terms of preference characteristics, users in different segmented markets show structural differences in

selecting functional modules or product categories. Some groups demonstrate a strong single preference in their personas, where both behaviors and spending are highly concentrated in a specific category, reflecting a clear and focused demand orientation. Other groups show diversified preferences, with behaviors and spending distributed across multiple categories, indicating that their demands are exploratory or comprehensive. Such differences in preference structure not only reflect different demand types, but also provide a basis for subsequent demand stratification and product portfolio design. In summary, persona-based segmented market analysis can reveal both common behavioral rules and differentiated characteristics, allowing enterprises to understand user demand structures at both the overall and group levels. Through comparative analysis across segmented markets, this study establishes an analytical foundation for extracting core demands and developing differentiated strategies.

5.2 Demand Identification and Business Implications

Based on the above analysis of user characteristics across segmented markets, the behavior patterns and preference structures reflected in user personas can be further transformed into identifiable core user demands. By comprehensively judging differences among user groups in activity level, consumption behavior, and preference characteristics, this study suggests that user demands in segmented markets show clear hierarchy and direction, which provides a practical entry point for business decision-making and optimization. For user groups with high value and high activity, persona features usually show high usage frequency, stable consumption behaviors, and concentrated preferences. Their core demands are no longer limited to basic functions or price incentives, but place more emphasis on service stability, functional completeness, and continuous improvement of the user experience. Therefore, at the business level, it is necessary to prioritize the reliability of core functions and enhance user stickiness through personalized recommendations, exclusive services, or high-quality content supply, thereby reducing the risk of churn caused by experience fluctuations. At the same time, deeper functions or value-added services can be designed around their high-frequency scenarios to further release long-term value. For users with medium activity or medium value, their behavior characteristics often indicate clear demands but limited usage intensity, and they are relatively sensitive to price, convenience, and usage cost. Their core demand is to lower the usage threshold and obtain clear value feedback. In terms of optimization, enterprises can simplify operational processes, improve user guidance, and introduce stage-based incentive mechanisms to guide users toward higher activity and value. For example, function prompts or bundled recommendations based on preference features can help users complete tasks more efficiently and thus strengthen continued usage intention. For low-activity or potential user groups, persona features typically show scattered behaviors, unstable consumption, and diversified preferences. Their demands are often in an exploration stage and have not yet formed stable usage patterns. Accordingly, the business focus should be on demand activation and user guidance. Through basic feature demonstration, scenario-based recommendations, or trial mechanisms, users can be helped to quickly understand product value. Meanwhile, overly complex function designs should be avoided, because they may increase cognitive burden and reduce the likelihood of continued use. Overall, persona-based demand identification can translate abstract market segmentation into specific and actionable business strategies. By applying differentiated optimization plans according to demand characteristics in different segments, enterprises can improve resource allocation efficiency and achieve better coordination between business outcomes and user experience while meeting diversified user needs.

6. Conclusion

This study focuses on the practical problem of demand analysis in segmented markets and proposes and validates an SQL-based approach for user persona construction and application. By systematically integrating user basic information, transaction data, and behavior logs in relational databases, a multidimensional user persona system is established, covering demographic attributes, consumption behaviors, usage frequency, and preference characteristics. On this basis, the differences in user characteristics and core demands across segmented markets are further analyzed. The results indicate that the SQL-based user persona method has clear advantages in terms of implementation cost, interpretability, and engineering feasibility, and it can effectively support user demand identification and business decision optimization in segmented markets. This study provides a feasible path for enterprises to conduct data-driven user analysis within existing database environments, and it also lays a foundation for introducing more advanced analytical methods and expanding application scenarios in future work.

References

- [1] Chandan, A. (2024). Unveiling customer personalities using segmentation and exploratory data analysis. *Health Leadership and Quality of Life*, 3(420), 1. <https://doi.org/10.56294/hl2024.420>
- [2] Ikeh, C. O. (2024). AI-powered market segmentation and personalization strategies for enhancing digital product lifecycle management. *International Journal of Research Publication and Reviews*, 6(6), 21107.

<https://doi.org/10.55248/gengpi.6.0625.21107>

- [3] Griva, A., et al. (2024). A two-stage business analytics approach to perform behavioural and geographic customer segmentation using e-commerce delivery data. *Journal of Decision Systems*, 33(1), 1–29. <https://doi.org/10.1080/12460125.2022.2151071>
- [4] Khamaj, A., & Ali, A. M. (2024). Adapting user experience with reinforcement learning: Personalizing interfaces based on user behavior analysis in real-time. *Alexandria Engineering Journal*, 95, 164–173. <https://doi.org/10.1016/j.aej.2024.03.045>
- [5] Zamil, M. H., Jyoti, S. N., & Ikbal, M. Z. (2023). Strategic role of business analysts in digital transformation tools, roles, and enterprise outcomes. *American Journal of Scholarly Research and Innovation*, 2(02), 246–273. <https://doi.org/10.63125/rc45z918>
- [6] Nookala, G. (2025). Integration of AI and machine learning with Snowflake: How Snowflake is enabling advanced analytics and AI-driven insights in cloud environments. *International Journal of Artificial Intelligence, Data Science, and Machine Learning*, 6(1), 147–155. <https://doi.org/10.63282/3050-9262.IJAIDSML-V6I1P116>
- [7] Saidur, M. J. I. (2025). AI-enhanced business intelligence dashboards for predictive market strategy in US enterprises. *International Journal of Business and Economics Insights*, 5(3), 603–648. <https://doi.org/10.63125/8cvgn369>
- [8] Furmakiewicz, M., et al. (2025). So, you want to build a Copilot? In *The design of human-centered artificial intelligence for the workplace* (p. 125). Springer. https://doi.org/10.1007/978-3-031-83512-4_8
- [9] Gabelaia, I. (2025). The relevance of AI-driven marketing through design thinking for a strategic approach to resilient problem-solving and communication for SMEs. *Sustainability*, 17(20), 8994. <https://doi.org/10.3390/su17208994>
- [10] Adamyk, B., et al. (2025). Risk management in DeFi: Analyses of the innovative tools and platforms for tracking DeFi transactions. *Journal of Risk and Financial Management*, 18(1), 38. <https://doi.org/10.3390/jrfm18010038>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).