

Visualization of Consumption Trends and Category Demand Analysis of Prepared Dishes

Shunhua Liu, Xixi Huo, Qingfeng Zhou

College of Artificial Intelligence and Big Data, Henan University of Technology, Zhengzhou, Henan, China

Abstract: To address the industry pain points in the prepared dish market, such as category homogenization, inaccurate consumption trend prediction, and insufficient regional preference mining, this paper designs and implements a set of visualization system for consumption trends and category demand analysis of prepared dishes by combining big data analysis and front-end and back-end development technologies. The system adopts a full-process architecture of "data processing - data analysis - visualization display". The back-end builds a four-layer data warehouse(ODS,DWD,DWS,ADS)based on Hadoop and Hive, realizes ETL task scheduling through Azkaban, and completes cross-database data synchronization with DataX. The front-end uses Vue.js+ECharts to build a responsive visualization interface, supporting interactive analysis in six dimensions including consumption trends, category demand, and crowd portraits. Test results based on more than 1,200 valid prepared dish consumption data show that the system's query response time is ≤ 3 seconds and the visualization chart loading time is ≤ 5 seconds. It can intuitively present market rules and user demand characteristics, provide data-driven decision support for enterprises, and has good practicability and scalability.

Keywords: Prepared Dishes; Data Warehouse; Separation of Front-End and Back-End; Visualization; Multi-dimensional Analysis; Task Scheduling

1. Introduction

With the popularization of fast-paced lifestyles and the upgrading of home cooking demands, the prepared dish market has witnessed explosive growth and become a new growth point in the food industry [1]. Consumers' demand for convenient, efficient, and diversified

catering solutions continues to rise, covering multiple scenarios such as single-person meals, family dinners, and small gatherings, driving the expansion of prepared dish categories toward local specialty dishes, healthy light meals, banquet dishes and other directions [2]. However, the industry still faces multiple challenges: severe category homogenization intensifies market competition; inaccurate prediction of consumption trends by enterprises leads to inventory backlogs; insufficient excavation of regional consumption preference differences results in poor product adaptability; and inadequate insight into user preferences impairs marketing accuracy [3]. Meanwhile, the lack of an integrated data analysis and visualization tool makes it difficult for enterprises to quickly obtain core market information. Against this background, building a system integrating data processing, multi-dimensional analysis, and visualization display is of great practical significance for the high-quality development of the prepared dish industry.

2. Visualization Analysis Process

(1) Clarify Objectives. Visualization analysis of prepared dishes requires multi-dimensional information such as product information, sales transaction data, user portrait data, and scenario-related data. Through five dimensions—consumption trends, crowd portraits, scenario correlation, price sensitivity, and regional differences—it reveals market rules and user demand characteristics, providing data support for enterprises' product research and development (R&D) and marketing strategy formulation.

(2) Data Collection and Preprocessing. The dataset in this study is a CSV file provided by the Model Whale platform, covering core information such as orders, commodities, and users. Data is collected via Flume and stored in HDFS [4]. Data cleaning and field standardization are completed using Hive SQL

to build a four-layer data warehouse (ODS, DWD, DWS, ADS). Then DataX is employed to realize data synchronization between Hive and MySQL, providing a standardized data source for visualization.

(3) Visualization Analysis of Data. Line charts and bar charts are used for consumption trends to intuitively present changes in sales volume and sales revenue of single products and categories across different time periods; bar charts are adopted for crowd portraits to show the consumption proportion and preferences of users by age, gender, and city level; pie charts and bar charts are utilized for consumption scenario correlation to compare category demand differences among "single-person meal/small gathering/family dinner" scenarios; multiple bar charts are applied for price sensitivity to demonstrate the combined effects of different price ranges and preferential methods; bar charts are used for regional differences to visually display the total consumption and category preference distribution of each province.

(4) Summary. Integrate and analyze the visualization results of each dimension, extract the core rules of the prepared dish market, and provide practical suggestions for enterprises to optimize product layout and implement precision marketing.

3. Introduction to Prepared Dish Data

3.1 Data Analysis Technologies

(1) Visualization Analysis of Consumption Trends: Through heat ranking charts of category sales revenue and monthly/quarterly sales volume change trend charts, decompose changes in sales volume and sales revenue of single products and categories at the quarterly/monthly granularity, and identify the iteration rules of category popularity and seasonal consumption characteristics.(2) Visualization Analysis of Crowd Portraits: Through order proportion charts, consumption amount proportion charts, and category preference radar charts of different groups, explore differences in category selection tendencies, price acceptance thresholds, and repurchase behaviors of consumers by age, gender, and city level.(3) Visualization Analysis of Scenario Correlation: Relying on scenario order proportion charts, average customer price comparison charts, and average purchase quantity charts, map the three core scenarios of

"single-person meal/small gathering/family dinner" based on packaging specifications, and analyze the category demand, average customer price level, and purchase frequency characteristics corresponding to each scenario.(4) Visualization Analysis of Price Sensitivity: Through four types of visualization charts—order proportion charts by price range, sales volume distribution charts, sales revenue comparison charts, and repurchase rate trend charts—study the combined effects of different price ranges and preferential methods, and decompose differences in sales revenue proportion, user satisfaction, and premium space.(5) Visualization Analysis of Regional Differences: With the help of sales revenue/sales volume ranking charts of each province and consumption comparison charts of different city levels, compare differences in total consumption, category preferences, and repurchase rates among different provinces, and identify regional market opportunities.

3.2 Data Source

The data is derived from public data on the Model Whale platform, covering core data such as prepared dish product information, category data, and user feedback data, which can intuitively reflect the market performance, user preferences, and category competition pattern of prepared dish products. Flume is used to monitor directories and collect data in real time. The data supports efficient processing of more than 100,000 records per year, and the core field information is shown in Table 1.

3.3 Development Tools and Programming Languages

(1) Flume: Used to monitor local directories, collect raw prepared dish consumption data in real time and write it to HDFS, ensuring the real-time performance and stability of data collection.

(2) Hadoop: Provides distributed storage (HDFS) capability for storing massive raw prepared dish data and data at all layers of the data warehouse. The advantage of Hadoop's distributed architecture makes it the mainstream choice for massive data storage and processing [5], widely applied in various big data analysis scenarios.

(3) Hive: Constructs a four-layer data warehouse architecture (ODS, DWD, DWS, ADS), supporting data cleaning, standardization, aggregation, and multi-dimensional analysis.

(4) DataX: Realizes cross-database data synchronization between Hive and MySQL, synchronizes analysis results to relational databases, and supports visualization queries.

Table 1. Core Field Information of Prepared Dish Data

Field Type	Non-null Count/Records	Data Type
order_id	10000	int
product name	10000	string
product category	10000	string
package spec	10000	string
unit price	10000	double
quantity	10000	int
order_amount	10000	double
order_time	10000	string
receive_province	10000	string
city_level	10000	string
pay_channel	10000	string
discount_type	10000	string
discount amount	10000	double
actual pay	10000	double
stock remaining	10000	int
product score	10000	int
is_repurchase	10000	int
user_age_group	10000	string
user_gender	10000	string

(5) MySQL: Used to store synchronized structured analysis results, providing a stable data source for front-end visualization applications.

(6) Azkaban: Builds workflows, schedules full-process ETL tasks, and realizes automatic execution and status monitoring.

(7) Java: Used for data synchronization script development and back-end auxiliary logic implementation, adapting to the secondary development needs of tools such as DataX.

(8) SQL/HiveQL: Used for data warehouse table creation, data cleaning, conversion, and aggregation analysis, serving as the core language for multi-dimensional data analysis.

(9) Shell: Used for writing task scheduling scripts and cluster environment operation and maintenance, supporting script execution and environment management of Azkaban workflows.

4. System Design

4.1 Overall Architecture Design

The system adopts a full-process architecture of "data collection - data processing - data analysis

- visualization display", divided into four core layers: data layer, processing layer, analysis layer, and application layer, as detailed below:

(1) Data Layer: Responsible for raw data storage, including structured data such as e-commerce platform sales data and enterprise operation data. The HDFS distributed storage scheme is adopted to ensure data access efficiency.

(2) Processing Layer: Realizes data cleaning, integration, and modeling based on Hadoop and Hive, builds a four-layer data warehouse (ODS, DWD, DWS, ADS), and completes field standardization and indicator calculation. The four-layer data warehouse architecture [6] refers to mature data warehouse design specifications in the industry, which can effectively improve the hierarchy and efficiency of data processing.

(3) Analysis Layer: Carries out special analysis in six dimensions, including consumption trends, category demand, crowd portraits, scenario correlation, price sensitivity, and regional differences, and outputs structured analysis results.

(4) Application Layer: Divided into front-end visualization display and back-end service support. The back-end provides data via API interfaces, the front-end realizes chart display and interactive functions, and integrates task scheduling tools to implement process monitoring [7].

4.2 Data Cleaning

This paper constructs a temporary table tmp_user_order_clean via Hive SQL to complete the full-process cleaning and standardization of prepared dish data, laying a foundation for subsequent data warehouse modeling and multi-dimensional analysis. First, integrity verification is performed on the raw data to confirm that the data covers core fields such as orders, commodities, and users without missing key information; then abnormal data is filtered, invalid records with order amount ≤ 0 , purchase quantity ≤ 0 , and abnormal order time format are eliminated, and non-numeric characters in order IDs are removed to ensure the uniformity of core field formats. For duplicate values, duplicate order records are identified and eliminated through data deduplication logic to guarantee data uniqueness. On this basis, classification fields are standardized, unifying classification standards such as city levels (new

first-tier/first-tier/second-tier/third-tier/fourth-tier/fifth-tier), user age groups (18-24/25-34/35-44/45-54/55+), and user gender (male/female/unknown) to avoid the impact of inconsistent calibers on analysis results. Finally, derived fields are generated through calculation, including converting the original 0-50-point score into a standardized 5-point score, extracting packaging weight (g) from packaging specifications, dividing score levels (1-5 levels), packaging specification levels (1-3 levels), and order amount levels (1-4 levels) by intervals, providing structured and standardized data support for subsequent special analysis in five dimensions.

4.3 Data Warehouse Design

A four-layer data warehouse architecture is adopted to ensure the standardization and efficiency of data processing:

- (1) ODS Layer (Operational Data Store): Stores raw data collected by Flume, retains the original form of data, with CSV as the file format, generated in a rolling manner every minute.
- (2) DWD Layer (Data Warehouse Detail): Generates core detail tables through data cleaning, field standardization, and derived field calculation.
- (3) DWS Layer (Data Warehouse Summary): Builds 5 summary tables according to analysis themes, corresponding to the five analysis dimensions respectively, realizing data aggregation and preliminary indicator calculation.
- (4) ADS Layer (Application Data Service): Generates 5 application reports, including commodity sales trend reports, user portrait reports, price sensitivity reports, etc., directly supporting visualization queries.

5. Visualization Analysis

5.1 Visualization Analysis of Consumption Trends

The consumption trends of prepared dishes in 2024 are comprehensively presented through multiple charts, clearly showing the changes in annual sales volume and category popularity. The total annual sales volume reached 30,955 units, the market size exceeded 1.03 million yuan, and the total number of orders was 10,000. The overall consumption demand was stable without significant fluctuations:

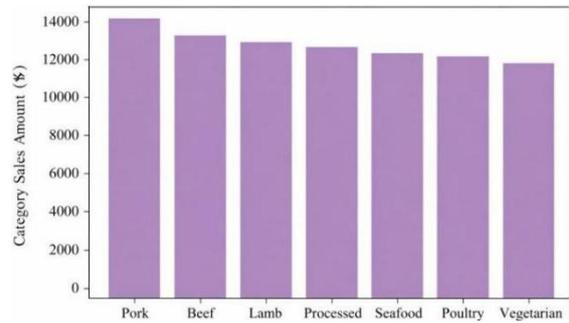


Figure 1. Heat Ranking Chart of Category Sales Revenue

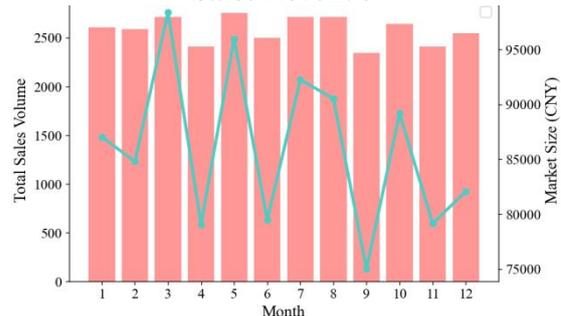


Figure 2. Monthly Sales Volume Change Trend Chart

From the data in Figure 1, in terms of categories, pork products ranked first in total sales revenue, followed by beef products, vegetable products, bean products, and soup products with small gaps in sales revenue, forming a balanced echelon. Figure 2 shows that the total sales volume was the highest in May 2024 and the lowest in September, with a small difference and a narrow overall fluctuation range, without obvious holiday "peak" characteristics—consumption peaks around the Spring Festival and National Day in traditional cognition did not appear significantly; instead, sales volume remained high in March, May, and July-August, which is speculated to be related to the demand for spring outings, picnics, and convenient diet in high summer temperatures. Quarterly data indicates that Q1 had the highest sales volume, Q4 the lowest, and the market size of the four quarters was stably around 250,000 yuan. It is recommended that enterprises do not over-focus on holiday stock preparation in production and inventory planning, but tilt resources toward months with relatively high sales volume such as March, May, and July-August; strengthen the supply chain advantages of pork and beef products, and sign long-term cooperation agreements with high-quality suppliers; meanwhile, launch summer-exclusive packages of "room temperature storage + rapid heating" according

to the characteristics of high summer temperatures to further expand the advantages of seasonal consumption.

5.2 Visualization Analysis of Crowd Portraits

The distribution of prepared dishes among different consumer groups presents the consumption preferences of various groups through diverse charts.

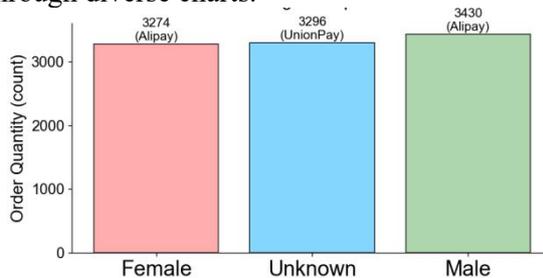


Figure 3. Order Volume and Preferred Payment Channels by Gender

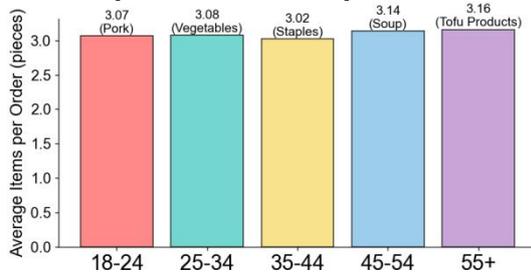


Figure 4. Average Purchase Quantity and Preferred Categories by Age Group

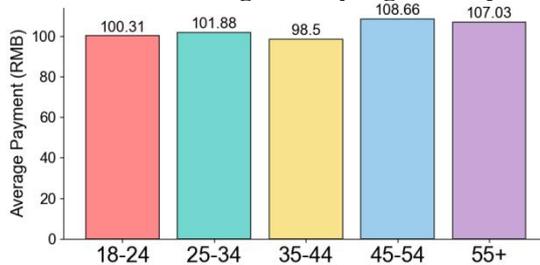


Figure 5. Comparison of Average Payment Amount by Age Group

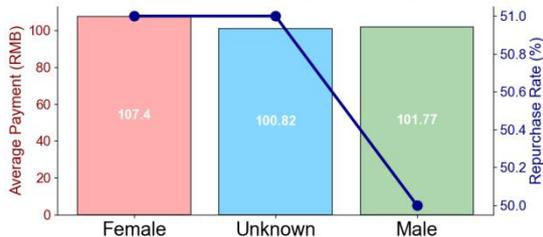


Figure 6. Comparison of Average Payment Amount & Repurchase Rate by Gender

As shown in Figures 3 and 4, the order proportion of prepared dishes is evenly distributed: 34.30% for males, 32.74% for females, and 32.96% for unknown gender. Among them, females have the highest average payment amount (107.40 yuan) and repurchase

rate (51.00%), while males have the largest number of orders; both prefer Alipay as the payment channel. In terms of age, as shown in Figures 5 and 6, the order volume of each age group is evenly distributed (1,935-2,097 orders). The 25-34 age group is the core consumer group with an order proportion of 20.97%; the 55+ age group has the highest average payment amount and prefers bean products; the 45-54 age group has a strong demand for soup products; and the 18-24 young group favors pork products. Overall, there is a trend that the older the age, the higher the average payment amount and average purchase quantity, and the middle-aged and elderly groups outperform young groups in terms of consumption intensity and purchase volume.

5.3 Visualization Analysis of Consumption Scenario Correlation

Focusing on three core scenarios, the scenario-based demand characteristics are revealed through scenario order proportion charts, scenario average customer price comparison charts, and scenario average purchase quantity charts.

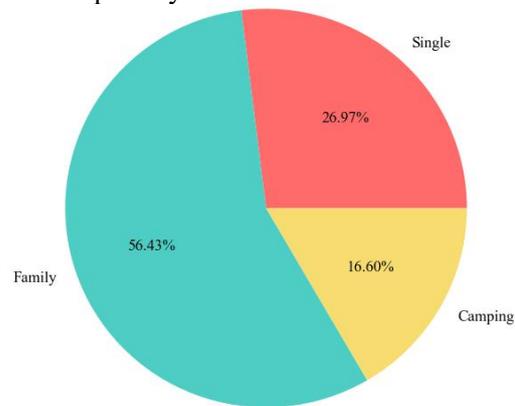


Figure 7. Order Proportion of Each Consumption Scenario

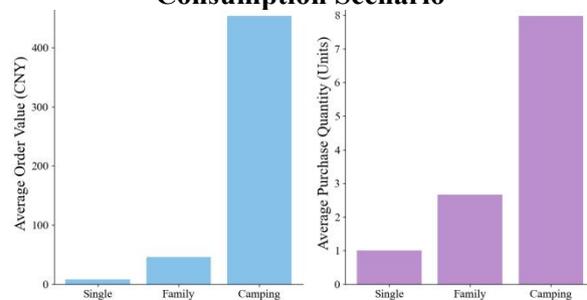


Figure 8. Comparison of Each Consumption Scenario

The consumption scenarios of prepared dishes present a "one main and two auxiliary" pattern. As shown in Figures 7 and 8, family dinners are

the absolute core scenario with an order proportion of 56.43%, with an average customer price of 45.87 yuan and an average purchase of 2.66 units, balancing frequency and basic consumption scale; single-person meals account for 26.97% of scenarios, with an average customer price of 8.05 yuan and an average purchase of 1 unit, focusing on convenient demand with high frequency and low customer price; festival banquets/outdoor camping account for 16.60% of scenarios, with a high average customer price of 453.30 yuan and an average purchase of 7.98 units, serving as a high-value incremental scenario. The three form a distinct gradient in order proportion, average customer price, and purchase quantity, covering the entire dimension of daily rigid demand and high-end consumption [8].

5.4 Visualization Analysis of Price Sensitivity

With the help of four types of visualization charts—order proportion charts by price range, sales volume distribution charts, sales revenue comparison charts, and repurchase rate charts—this paper decomposes the differences in market performance and user acceptance of low-priced, mid-priced, and high-priced prepared dishes from four core dimensions: order structure, consumption scale, profit contribution, and user loyalty, providing multi-dimensional data support for pricing strategy optimization and promotion plan design.

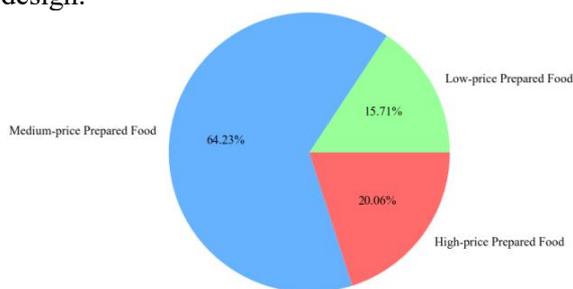


Figure 9. Order Proportion by Different Price Ranges



Figure 10. Sales Volume Distribution by Different Price Ranges

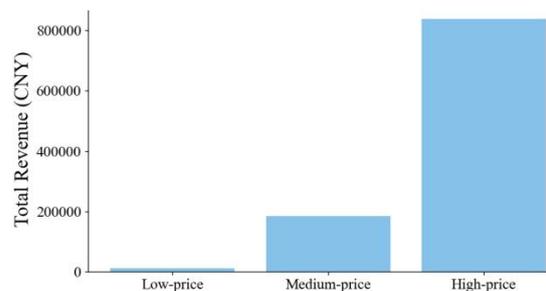


Figure 11. Sales Revenue Comparison by Different Price Ranges

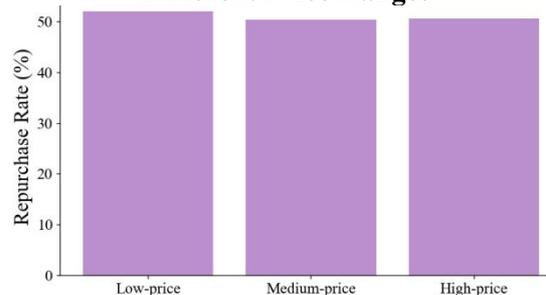


Figure 12. Repurchase Rate by Different Price Ranges

Consumption across different price ranges presents distinct "three-dimensional differentiation" characteristics: in terms of order structure, as shown in Figure 9, mid-priced prepared dishes have become the market mainstream with a proportion of 64.23%, forming a strong consumption foundation, which confirms the public's core demand for "cost-effective" prepared dishes; in terms of sales volume and profit contribution, high-priced prepared dishes show the dual advantages of "high sales volume + high output value"—as shown in Figures 10 and 11, their total sales volume reaches 14,980 units, contributing 81.10% of the total sales revenue, becoming the core pillar of profits, reflecting the willingness of high-end consumer groups to pay for high-quality prepared dishes; in terms of user loyalty, as shown in Figure 12, low-priced prepared dishes lead with a repurchase rate of 51.94%, and the repurchase rates of mid-priced and high-priced products are similar (about 50.4%), indicating that low-priced products have unique value in maintaining user stickiness, while mid-to-high-priced products achieve user retention through stable quality. The three form a complementary pattern of "mid-priced products stabilize the foundation, high-priced products generate profits, and low-priced products lock in users", jointly constituting the price ecology of the prepared dish market [9].

5.5 Visualization Analysis of Regional

Differences

With the help of four types of visualization charts—sales revenue ranking charts of each province, sales volume ranking charts of each province, sales revenue comparison charts of different city levels, and average customer price comparison charts of different city levels—this paper systematically decomposes the regional consumption differences of prepared dishes from three core dimensions: provincial consumption scale, category preferences, and urban hierarchical consumption capacity.

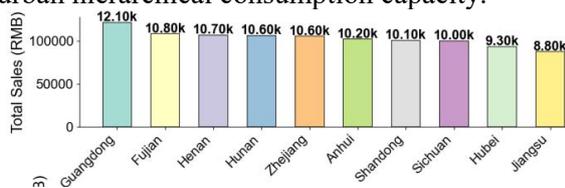


Figure 13. Sales Revenue Ranking Chart of Prepared Dishes by Province

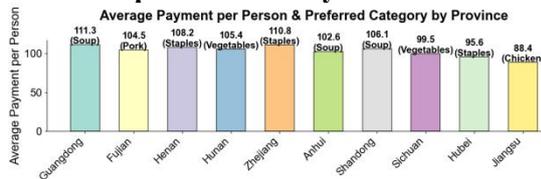


Figure 14. Average Payment Amount and Preferred Categories by Province

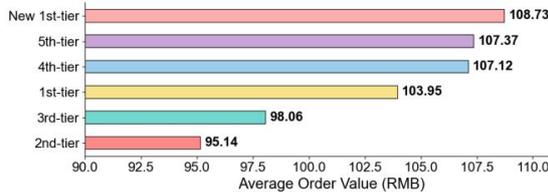


Figure 15. Comparison of Average Customer Price by City Level

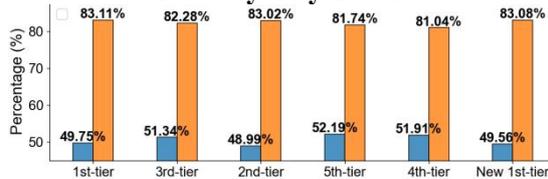


Figure 16. Comparison of Repurchase Rate & Dependence by City Level

At the provincial level, there is a significant "core agglomeration" characteristic. As shown in Figures 12 and 13, Guangdong, Fujian, Henan, and Hunan provinces are the main consumers. Among them, Guangdong Province ranks first with a total sales volume of 3,450 units and a total sales revenue of 121,469.0 yuan, with an average payment amount of 111.34 yuan and a preference for soup products; Fujian Province follows closely, with a core preference for pork products; as can be seen from Figures 14 and 15, the top 10 provinces all take

mid-priced prepared dishes as the consumption mainstream, with convergent regional price sensitivity, forming a stable mass consumption foundation[10].

At the urban level, consumption performance presents the characteristics of "strong sinking market": the sales revenue of fourth-tier and fifth-tier cities reaches 179,107.0 yuan and 178,767.0 yuan respectively, leading all levels; the average customer price presents a "new first-tier leading" pattern: 108.73 yuan for new first-tier cities, 107.37 yuan for fifth-tier cities, and 103.95 yuan for first-tier cities. As further revealed in Figure 16, the repurchase rate and consumption dependence of sinking markets (especially fourth- and fifth-tier cities) also show strong resilience, which indicates that the consumption capacity of the sinking market is no less than that of first and second-tier cities [11]. The sinking market has become a new growth pole in the prepared dish industry, and the excavation of its consumption demand is of great significance for enterprise layout [12].

6. Conclusion

Aiming at the industry pain points in the prepared dish market such as category homogenization and inaccurate consumption trend prediction, this study successfully designs and implements a visualization system for consumption trends and category demand analysis of prepared dishes. The system adopts a full-process architecture of "data processing-data analysis-visualization display", constructs a four-layer data warehouse via Hadoop and Hive, realizes efficient data processing and synchronization with Azkaban and DataX, and builds a responsive visualization interface based on Vue.js+ECharts at the front-end, featuring good technical feasibility and stability. Test results show that the system's query response time is ≤ 3 seconds and the visualization chart loading time is ≤ 5 seconds, which can meet enterprises' demands for efficient data analysis. Based on the multi-dimensional visualization analysis of the system, the core rules of the prepared dish market are clarified: in terms of consumption trends, the annual demand is stable, with March, May, and July-August as peak sales periods, and pork and beef products as core categories; in terms of crowd portraits, the 25-34 age group is the core consumer group, the middle-aged and elderly groups have stronger consumption

intensity, and females have the highest repurchase rate; in terms of scenario distribution, family dinners are the core scenario, and festival banquets/outdoor camping are high-value incremental scenarios; in terms of price ecology, a three-dimensional differentiation pattern of "mid-priced products stabilize the foundation, high-priced products generate profits, and low-priced products lock in users" is formed; in terms of regional characteristics, four provinces including Guangdong are the main consumers, and the sinking market has outstanding consumption capacity. These conclusions accurately reveal the core characteristics and demand rules of the prepared dish consumption market, providing clear data support for enterprises' product R&D, inventory planning, and marketing strategy formulation. The application value of this system lies in realizing the visualization and refined analysis of prepared dish consumption data, filling the gap of integrated data analysis tools in the industry, and featuring good practicability and scalability.

References

- [1] Weiwei Zhang, Qiang Wang. Analysis on the Development Status and Trends of China's Prepared Dishes Industry. *Science and Technology of Food Industry*, 2021, 42(22):1-7.
- [2] Dongmei Li, Beiwei Zhu. Inheritance and Innovation of Chinese Prepared Dish Industry. *Journal of Chinese Institute of Food Science and Technology*, 2022, 22(10): 1-8.
- [3] Jianguo Wang, Min Liu. Homogenization Competition in the Prepared Dish Market and Its Breakthrough Paths. *Journal of Commercial Economics*, 2021(18): 135-137.
- [4] Yunlong Ma, Xiaofeng Zhang. Design and Implementation of a Massive Data Governance Framework Based on Hadoop. *Application Research of Computers*, 2019, 36(8): 2386-2389.
- [5] Xiaodong Lin, Gang Zhao. Research on the Application of Hadoop Distributed Storage in Big Data Analysis. *Computer Science*, 2020, 47(S1): 456-459.
- [6] Na Li, Qiang Zhang. Design of a Prepared Dish Consumption Demand Analysis System Based on Big Data Visualization. *Computer Engineering and Applications*, 2023, 59(18): 234-241.
- [7] Peng Wang, Jie Liu. Design and Practice of a Four-Layer Enterprise-Level Data Warehouse Architecture. *Data Analysis and Knowledge Discovery*, 2021, 5(3): 102-109.
- [8] Jing Zhao, Jun Li. Scenario-Based Segmentation Trends and Marketing Strategies in the Consumer Goods Industry. *Commercial Research*, 2020(10): 89-96.
- [9] Ming Sun, Li Zhou. Research on the Matching Degree Between Price Hierarchy Distribution and Consumption Demand of Fast-Moving Consumer Goods. *Price Theory and Practice*, 2021(6): 145-148.
- [10] Junwen An, Ziyang Fang. Development Status, Influencing Factors and Trends of China's Prepared Dish Industry. *Journal of Food Science and Technology*, 2023, 41(3): 132-140.
- [11] Lei Gao, Min Wu. Development Opportunities of the Prepared Dish Industry Under the Background of Consumption Upgrade in the Sinking Market. *Journal of Commercial Economics*, 2022(12): 141-144.
- [12] Min Zhang, Wei Liu. Research on Prepared Dish Consumer Portraits and Category Preferences—Based on 2023 E-Commerce Data. *Journal of Commercial Economics*, 2023(15): 156-160.