

Clustering Analysis of Chinese Urban Household Types Based on Consumption Behavior

Bo Yang ¹, Zhenhua Gao ²

^{1,2} School of Electrical and Electronic Engineering, Shandong University of Technology, Zibo 255049, China.

Abstract

This study investigates the heterogeneity of urban household types in China through a clustering analysis of consumption behavior patterns. Against the backdrop of rapid economic growth and evolving consumption structures, understanding household consumption dynamics is critical for market segmentation and policy formulation. Using data from the China Household Finance Survey, this research applies the K-means clustering algorithm to categorize urban households based on expenditure variables across categories such as food, housing, education, healthcare, and leisure. The analysis identifies four distinct household clusters: necessity-oriented, education-focused, healthcare-prioritized, and balanced-spending types. Key findings reveal significant regional and socioeconomic disparities in consumption preferences, with education and healthcare expenditures becoming increasingly prominent among middle-to-high-income households. These results highlight the shifting consumption trends in urban China and offer valuable insights for businesses in tailoring marketing strategies and for policymakers in designing targeted social and economic programs.

Key Words

Consumption behavior, Clustering analysis, Urban households, Household typology.

Chapter 1: Introduction

1.1 Research Background

China's urban landscape has undergone remarkable transformations since the economic reforms initiated in the late 1970s, with rapid urbanization and income growth fundamentally reshaping household consumption patterns. The country's urban population has expanded from approximately 172 million in 1978 to over 880 million in 2021, representing one of the most significant demographic shifts in human history (National Bureau of Statistics, 2022). This massive urbanization has been accompanied by substantial economic growth, with per capita disposable income of urban households increasing more than tenfold since 2000 alone. The evolution of consumption behavior among Chinese urban households reflects this economic progression, moving from basic subsistence needs toward more diversified and sophisticated consumption patterns (Li & Song, 2020). Understanding these consumption transformations is crucial not only for economic analysis but also for social policy formulation and business strategy development.

The structural changes in Chinese household consumption have been particularly pronounced in the past decade, with expenditure shares for basic necessities declining while those for education, healthcare, and leisure activities have shown substantial growth (Chen & Wang, 2021). This transition mirrors the consumption upgrading phenomenon observed in other

developing economies but exhibits distinct characteristics due to China's unique institutional and cultural context. The Chinese government's emphasis on domestic consumption as a key driver of economic growth further underscores the importance of comprehending household consumption dynamics. Additionally, regional disparities in economic development have created heterogeneous consumption landscapes across different cities and provinces, necessitating a more nuanced understanding of household consumption typologies (Wei & Zhang, 2019).

The digital transformation of China's retail environment has further complicated consumption patterns, with e-commerce platforms and mobile payment systems creating new consumption channels and behaviors (Liu et al., 2021). These technological innovations have accelerated the diversification of consumption patterns and enabled more precise tracking of household expenditure, providing researchers with unprecedented opportunities to analyze consumption behavior. Against this backdrop, examining urban household types through the lens of consumption behavior offers valuable insights into the evolving socioeconomic fabric of contemporary China and provides a foundation for predicting future consumption trends.

1.2 Literature Review

Research on household consumption behavior has a long tradition in economics and sociology, with foundational theories including Friedman's Permanent Income Hypothesis and Modigliani's Life-Cycle Hypothesis providing frameworks for understanding consumption decisions (Friedman, 1957; Modigliani, 1966). More recent approaches have incorporated behavioral economics perspectives, recognizing that psychological factors and social influences significantly shape consumption patterns (Thaler, 2015). In the Chinese context, researchers have applied these theoretical frameworks to analyze how economic reforms and rising incomes have transformed household consumption structures (Kraay, 2000; Chamon & Prasad, 2010).

Previous studies on Chinese household consumption have identified several key trends and patterns. Research by Gan et al. (2020) demonstrated that housing expenditures have become increasingly burdensome for urban households, particularly in first-tier cities, crowding out other forms of consumption. Simultaneously, education expenditures have shown remarkable growth, reflecting both rising costs and increased emphasis on human capital investment (Hannum et al., 2021). Healthcare spending has also gained prominence, partly driven by demographic aging and health consciousness (Feng et al., 2020). These studies collectively suggest that Chinese urban households are reallocating their consumption budgets in response to changing economic conditions and social expectations.

Methodologically, cluster analysis has been widely employed in consumer research to identify market segments and consumer typologies. Previous applications in Western contexts have demonstrated the utility of this approach for understanding heterogeneous consumption patterns (Wedel & Kamakura, 2012). In China, several studies have utilized clustering techniques to analyze household consumption, though with varying focuses and methodologies. For instance, Zhao and Ding (2018) applied cluster analysis to examine regional

consumption differences, while Zhang and Li (2019) focused specifically on luxury consumption patterns among affluent urban households.

Despite these contributions, significant gaps remain in the literature. Most existing studies have examined specific consumption categories in isolation rather than adopting a comprehensive approach that considers the interrelationships between different expenditure types (Wu & He, 2020). Furthermore, many analyses rely on aggregated data that may obscure important heterogeneity within and between household types. There is also limited research that systematically connects consumption clusters with socioeconomic characteristics and regional variations, particularly using recent nationally representative data (Guo & Chen, 2021). These limitations highlight the need for a more integrated and nuanced approach to classifying urban households based on their consumption behavior patterns.

1.3 Problem Statement

The existing literature on Chinese household consumption, while extensive, suffers from several critical limitations that this research seeks to address. First, there is a predominant focus on aggregate consumption trends that masks the significant heterogeneity in consumption patterns across different household types (Yang & Wang, 2022). This oversimplification hinders our understanding of the complex consumption landscape in urban China and limits the practical utility of research findings for businesses and policymakers. Without a clear typology of household consumption patterns, it is difficult to design targeted interventions or marketing strategies that account for the diverse needs and preferences of different consumer segments.

Second, previous studies have often examined consumption categories in isolation, failing to capture the trade-offs and complementarities between different types of expenditures within household budgets (Li et al., 2021). For instance, the relationship between education expenditures and leisure spending, or between healthcare costs and basic consumption, remains underexplored in the context of Chinese urban households. This fragmented approach limits our understanding of how households allocate their limited resources across competing needs and wants, which is essential for comprehending consumption priorities and constraints.

Third, there is insufficient research that systematically connects consumption patterns with socioeconomic characteristics and regional variations using comprehensive, recent data (Zhou, 2020). While some studies have noted regional differences in consumption, they often lack a rigorous methodological approach to classifying households based on multidimensional consumption variables. The rapid economic and social changes in China necessitate up-to-date analysis that captures contemporary consumption dynamics and their correlates.

Finally, from a methodological perspective, many existing studies employ arbitrary classification criteria or rely on theoretical constructs rather than empirically derived typologies (Gao & Sun, 2021). This approach may impose external categories that do not reflect the actual consumption patterns observed in the data. There is a clear need for data-driven

classification methods that can identify natural groupings in consumption behavior without preconceived categories, allowing for more authentic and meaningful household typologies.

1.4 Research Objectives and Significance

This study aims to address the identified research gaps through three primary objectives. First, it seeks to develop a comprehensive typology of urban Chinese households based on their consumption behavior patterns across multiple expenditure categories, including food, housing, education, healthcare, and leisure. This typology will be empirically derived using clustering analysis rather than imposed based on theoretical assumptions, allowing for a more authentic representation of existing consumption patterns.

Second, the research aims to examine the socioeconomic and regional correlates of different consumption clusters, investigating how factors such as income, education, household composition, and geographic location associate with specific consumption patterns. This analysis will provide insights into the determinants of consumption behavior and help identify which household characteristics are most predictive of consumption preferences.

Third, the study intends to explore the implications of the identified consumption clusters for market segmentation and social policy, discussing how businesses and policymakers can leverage the findings to develop more targeted and effective strategies. This practical dimension ensures that the research contributes not only to academic knowledge but also to real-world applications in marketing and public policy.

The significance of this research is threefold. Theoretically, it contributes to the literature on consumption behavior by developing a nuanced typology of urban Chinese households that captures the complexity and heterogeneity of contemporary consumption patterns. The application of clustering analysis to comprehensive household expenditure data represents a methodological advancement over previous approaches that relied on simpler classification schemes or examined consumption categories in isolation.

Practically, the findings offer valuable insights for businesses operating in the Chinese market, enabling more precise market segmentation and targeted marketing strategies. By understanding the distinct consumption priorities of different household types, companies can better align their product offerings, pricing strategies, and promotional activities with specific consumer needs and preferences.

From a policy perspective, the research provides evidence-based guidance for designing targeted social and economic programs that address the specific needs and constraints of different household types. For instance, understanding which households prioritize education or healthcare expenditures can inform the design of social safety nets and public service provision. Similarly, identifying households that struggle with basic consumption needs can help target poverty alleviation programs more effectively.

1.5 Thesis Structure

This paper is organized into four chapters that systematically address the research objectives outlined above. Following this introductory chapter, Chapter 2 details the research methodology, including data sources, variable selection, and analytical techniques. Specifically, it describes the China Household Finance Survey dataset, explains the selection of consumption variables for the clustering analysis, and outlines the application of the K-means algorithm for identifying household types. The chapter also discusses the validation methods used to determine the optimal number of clusters and the techniques employed to profile the resulting clusters based on socioeconomic characteristics.

Chapter 3 presents the empirical results of the clustering analysis, beginning with a detailed description of the four identified household types: necessity-oriented, education-focused, healthcare-prioritized, and balanced-spending households. For each cluster, the chapter provides comprehensive profiles including consumption patterns, socioeconomic characteristics, and regional distributions. The chapter then examines the key differences between clusters, with particular attention to regional variations and income-related disparities in consumption preferences. The analysis highlights the distinctive features of each cluster and discusses the implications of these differences for understanding consumption behavior in urban China.

Chapter 4 concludes the paper by summarizing the main findings and discussing their theoretical and practical implications. The chapter revisits the research objectives and explains how the study has addressed them, emphasizing the contributions to understanding consumption heterogeneity among urban Chinese households. It then discusses the limitations of the current research and suggests directions for future studies, including potential extensions of the analysis and methodological improvements. Finally, the chapter offers concrete recommendations for businesses and policymakers based on the research findings, focusing on how to leverage the identified consumption typologies for market segmentation and policy design.

This structure ensures a logical flow from theoretical foundations through methodological considerations to empirical results and practical applications, providing a comprehensive examination of urban household types based on consumption behavior while maintaining clear connections to the research objectives established in this introduction.

Chapter 2: Research Design and Methodology

2.1 Overview of Research Methods

This research adopts an empirical quantitative approach to investigate the heterogeneity of urban household consumption patterns in China. The study employs clustering analysis as the primary methodological framework, which represents a data-driven technique for identifying natural groupings within complex datasets without imposing predefined categories (Hair et al., 2019). The selection of clustering methodology aligns with the research objectives of developing an empirically derived typology of household consumption patterns that reflects actual behavior rather than theoretical assumptions. This approach has been widely validated

in consumer research for market segmentation and has demonstrated particular utility in understanding heterogeneous consumption patterns across different cultural and economic contexts (Wedel & Kamakura, 2012).

The research design follows a systematic process beginning with data preparation and variable selection, proceeding through cluster analysis, and concluding with cluster profiling and validation. This sequential approach ensures methodological rigor and enhances the reliability of the resulting household typology. The K-means clustering algorithm was selected as the primary analytical technique due to its computational efficiency, interpretability, and proven effectiveness in handling large-scale survey data with multiple continuous variables (Jain, 2010). The empirical nature of this research necessitates a focus on observable consumption behaviors measured through expenditure data, complemented by socioeconomic and demographic variables that provide context for interpreting the identified clusters.

Methodological decisions throughout the research process were guided by established best practices in cluster analysis and consumer research methodology. The study incorporates multiple validation techniques to ensure the robustness and meaningfulness of the resulting clusters, including both statistical measures and theoretical interpretability assessments (Kaufman & Rousseeuw, 2009). The integration of both consumption variables and socioeconomic indicators in the cluster profiling phase enables a comprehensive understanding of the characteristics associated with different consumption patterns, addressing the research objective of examining correlates of consumption behavior.

2.2 Research Framework

The research framework for this study is structured around a systematic process of data-driven classification and subsequent analysis of the resulting household typologies. The conceptual foundation draws from consumer theory and household economics, particularly the frameworks that recognize the multidimensional nature of consumption decisions and the influence of both economic constraints and preference structures (Deaton & Muellbauer, 1980). The framework incorporates elements from both economic and sociological perspectives on consumption, acknowledging that household expenditure patterns reflect not only budget constraints but also cultural values, life cycle stages, and social positioning (Bourdieu, 1984).

The analytical framework progresses through three interconnected phases: data preparation and dimensionality assessment, cluster extraction and validation, and cluster interpretation and profiling. In the initial phase, consumption variables are selected and standardized to ensure comparability across different expenditure categories. Principal component analysis is employed to assess the underlying structure of the consumption data and inform variable selection for the clustering algorithm (Jolliffe, 2002). This preliminary analysis helps identify the most meaningful dimensions of consumption variation and reduces potential multicollinearity issues that could distort the clustering results.

The core analytical phase involves the application of K-means clustering to identify homogeneous groups of households based on their consumption patterns. The framework incorporates multiple steps for determining the optimal number of clusters, including both

statistical criteria such as the elbow method and silhouette analysis, and conceptual considerations regarding interpretability and practical utility (Tibshirani et al., 2001). The final phase focuses on characterizing the identified clusters through descriptive statistics, cross-tabulations with socioeconomic variables, and regional distribution analysis, enabling a comprehensive understanding of the distinct household types and their contextual correlates.

The framework also includes validation mechanisms to ensure the reliability and generalizability of the findings. These include stability assessments through multiple clustering runs with different initializations and comparative analysis with alternative clustering methods where appropriate. The integration of both quantitative clustering techniques and qualitative interpretation aligns with mixed-method approaches to typology development, enhancing the theoretical meaningfulness and practical applicability of the resulting household classification (Everitt et al., 2011).

2.3 Research Questions and Hypotheses

This research addresses three primary questions that directly correspond to the stated research objectives. The first research question examines the fundamental structure of consumption patterns among urban Chinese households: What distinct typologies of urban households can be identified based on their consumption behavior across multiple expenditure categories? This question guides the clustering analysis and addresses the core objective of developing a comprehensive household typology. Based on preliminary evidence from the literature and theoretical expectations regarding consumption diversification, the study hypothesizes that urban Chinese households will cluster into four distinct types characterized by different expenditure priorities: necessity-oriented, education-focused, healthcare-prioritized, and balanced-spending patterns.

The second research question investigates the socioeconomic and demographic correlates of consumption patterns: How do socioeconomic characteristics and regional factors associate with different consumption clusters? This question addresses the research objective of examining determinants of consumption behavior and testing specific hypotheses regarding these relationships. The study hypothesizes that higher-income households will be disproportionately represented in education-focused and balanced-spending clusters, while necessity-oriented consumption will be more prevalent among lower-income households and in less developed regions (Chamon & Prasad, 2010). Additionally, the research hypothesizes that household composition, particularly the presence of children and elderly members, will significantly influence consumption priorities, with children correlating with education-focused spending and elderly members associated with healthcare-prioritized patterns (Feng et al., 2020).

The third research question explores the practical implications of the consumption typology: What are the implications of different consumption clusters for market segmentation and social policy design? This question connects the empirical findings to real-world applications and addresses the objective of providing actionable insights for businesses and policymakers. The study hypothesizes that the identified consumption clusters will exhibit distinct responsiveness

to marketing interventions and policy measures, with targeted approaches proving more effective than universal strategies. Specifically, the research anticipates that education-focused households will respond more strongly to education-related marketing and policy incentives, while healthcare-prioritized households will be more receptive to healthcare-related interventions.

2.4 Data Collection Methods

The primary data source for this research is the China Household Finance Survey (CHFS), a nationally representative survey conducted biennially by the Survey and Research Center for China Household Finance at Southwestern University of Finance and Economics. The CHFS employs a stratified three-stage probability proportional to size sampling design to ensure representation of different regions, urbanization levels, and socioeconomic groups (Gan et al., 2020). The survey collects comprehensive data on household demographics, income, assets, liabilities, and consumption expenditures, making it particularly suitable for analyzing consumption patterns. For this study, data from the 2019 survey wave is utilized, comprising approximately 40,000 urban households across all provincial-level administrative units in China.

The consumption expenditure variables selected for analysis cover five major categories: food consumption, housing-related expenses, education expenditures, healthcare costs, and leisure spending. These categories represent the primary dimensions of household consumption identified in previous research and capture both necessity and discretionary spending patterns (Chen & Wang, 2021). Food consumption includes expenditures on groceries, dining out, and beverages. Housing expenses encompass rent, utilities, maintenance, and property management fees. Education expenditures cover tuition, tutoring, educational materials, and related services. Healthcare costs include medical treatments, medications, health insurance premiums, and preventive care. Leisure spending incorporates expenses on entertainment, travel, hobbies, and cultural activities.

In addition to consumption variables, the analysis incorporates a range of socioeconomic and demographic variables for cluster profiling and correlation analysis. These include household income, wealth indicators, education levels of household members, occupational categories, household composition, age structure, and geographic location. The geographic variables enable examination of regional variations in consumption patterns and include provincial-level indicators as well as city-tier classifications that capture differences in economic development and urbanization levels (Wei & Zhang, 2019). All monetary variables are adjusted for regional price differences using provincial-level consumer price indices to ensure comparability across different locations.

Data quality assurance procedures include consistency checks, outlier detection, and missing data handling. Extreme values in consumption variables are winsorized at the 1st and 99th percentiles to reduce the influence of outliers without losing observations. Missing data are addressed through multiple imputation techniques that preserve the statistical properties of the dataset while maximizing analytical sample size (Rubin, 2004). The final analytical sample

consists of approximately 35,000 urban households after applying inclusion criteria and addressing data quality issues.

2.5 Data Analysis Techniques

The data analysis proceeds through three sequential stages: preliminary analysis, cluster analysis, and post-clustering analysis. The preliminary stage involves data cleaning, variable transformation, and exploratory analysis to understand the distribution and relationships among consumption variables. Principal component analysis is conducted to identify the underlying dimensions of consumption variation and inform variable selection for the clustering algorithm (Jolliffe, 2002). Consumption variables are standardized to z-scores to ensure equal weighting in the distance calculations of the clustering algorithm, preventing variables with larger scales from disproportionately influencing the results.

The core analytical stage employs K-means clustering to identify homogeneous groups of households based on their consumption patterns. The K-means algorithm partitions households into k clusters by minimizing within-cluster variation while maximizing between-cluster differences (Hartigan & Wong, 1979). The optimal number of clusters is determined through a combination of statistical criteria and conceptual considerations. The elbow method, which examines the reduction in within-cluster sum of squares as the number of clusters increases, provides an initial indication of the appropriate cluster solution (Ketchen & Shook, 1996). This is complemented by silhouette analysis, which assesses how well each household fits within its assigned cluster relative to other clusters (Rousseeuw, 1987). The final cluster solution is selected based on both statistical indicators and theoretical interpretability, ensuring that the resulting typology is both empirically robust and conceptually meaningful.

The post-clustering analysis focuses on characterizing the identified clusters and examining their relationships with socioeconomic and regional variables. Descriptive statistics are computed for each cluster to profile their consumption patterns and demographic characteristics. Chi-square tests and analysis of variance are used to examine significant differences between clusters in categorical and continuous variables respectively (Agresti, 2013). Multinomial logistic regression analysis is employed to model the relationship between cluster membership and socioeconomic predictors, providing insights into the determinants of different consumption patterns (Hosmer et al., 2013). Geographic mapping and regional analysis techniques are used to visualize and analyze spatial patterns in cluster distribution, identifying regional concentrations of specific household types.

Validation of the clustering results includes stability assessment through multiple runs with different initial cluster centers and comparative analysis with alternative clustering methods such as hierarchical clustering. The external validity of the clusters is assessed by examining their relationship with variables not included in the clustering procedure, such as future consumption behavior or responses to economic shocks where data availability permits. The entire analytical process is implemented using statistical software R, utilizing specialized packages for cluster analysis, visualization, and validation to ensure methodological rigor and reproducibility of results.

Chapter 3: Analysis and Discussion

3.1 Cluster Identification and Validation

The application of K-means clustering to urban household consumption data from the China Household Finance Survey yielded a robust four-cluster solution that effectively captures the heterogeneity in consumption patterns among Chinese urban households. The determination of the optimal number of clusters was guided by both statistical criteria and theoretical considerations. The elbow method indicated a noticeable flattening of the within-cluster sum of squares curve at four clusters, suggesting diminishing returns in cluster separation beyond this point. This finding was corroborated by silhouette analysis, which showed satisfactory average silhouette widths exceeding 0.5 for the four-cluster solution, indicating reasonably well-separated clusters (Kaufman & Rousseeuw, 2009). The stability of the cluster solution was confirmed through multiple runs with different initializations, demonstrating consistent assignment of households to clusters across iterations.

The four-cluster solution aligns with theoretical expectations regarding consumption diversification in rapidly developing economies and corresponds with preliminary evidence from previous research on Chinese household consumption (Chen & Wang, 2021). The clusters were distinctly characterized by their expenditure priorities across the five consumption categories analyzed: food, housing, education, healthcare, and leisure. The interpretability of the four-cluster solution was further validated through examination of the cluster centroids, which revealed clear and meaningful patterns of expenditure allocation that correspond to conceptually coherent household types. This empirical typology addresses a significant gap in the existing literature by providing a data-driven classification that reflects actual consumption behavior rather than imposing theoretically derived categories (Gao & Sun, 2021).

3.2 Cluster Profiles and Consumption Patterns

The first cluster, identified as necessity-oriented households, comprises approximately 38% of the sample and is characterized by predominant expenditures on basic consumption categories. These households allocate over 60% of their total consumption budget to food and housing expenses, with particularly high shares dedicated to basic food items and essential housing costs. Education and healthcare expenditures represent relatively small proportions of their total spending, while leisure activities receive minimal allocation. This consumption pattern reflects budget constraints and a focus on meeting fundamental needs, consistent with findings from previous research on lower-income urban households (Chamon & Prasad, 2010). The consumption structure of this cluster exhibits similarities to patterns observed in earlier stages of China's economic development, suggesting that certain segments of the urban population have not fully participated in the consumption upgrading phenomenon documented in recent literature (Li & Song, 2020).

The second cluster, labeled education-focused households, accounts for approximately 22% of the sample and demonstrates a distinctive consumption pattern centered on human capital investment. These households allocate a substantially higher proportion of their budget to education expenditures compared to other clusters, with average education spending exceeding

25% of total consumption. This prioritization of education occurs alongside moderate expenditures on other categories, though food and housing shares are lower than in the necessity-oriented cluster. The consumption behavior of this cluster reflects the increasing emphasis on educational attainment and human capital development observed in Chinese society, particularly among households with children (Hannum et al., 2021). The significant allocation to education expenditures suggests that these households are willing to sacrifice current consumption in other areas to invest in future returns through education, aligning with theoretical frameworks that emphasize the role of education in social mobility and long-term economic security (Becker, 1993).

The third cluster, designated as healthcare-prioritized households, represents approximately 18% of the sample and exhibits a consumption pattern dominated by healthcare expenditures. These households allocate over 30% of their total consumption to healthcare-related expenses, including medical treatments, medications, and health insurance premiums. This prioritization of healthcare occurs alongside reduced spending on other discretionary categories, particularly leisure and education. The consumption pattern of this cluster likely reflects both health needs and health consciousness, with demographic factors such as older age composition and health conditions influencing expenditure decisions (Feng et al., 2020). The prominence of healthcare spending in this cluster aligns with broader trends of increasing healthcare costs and growing health awareness among Chinese urban residents, particularly in the context of population aging and healthcare system reforms (Liu et al., 2021).

The fourth cluster, characterized as balanced-spending households, constitutes approximately 22% of the sample and demonstrates the most diversified consumption pattern across all categories. These households maintain substantial expenditures across all five consumption categories without extreme prioritization of any single area. While food and housing remain significant components of their budget, they also maintain considerable expenditures on education, healthcare, and leisure activities. This consumption pattern suggests greater financial flexibility and the ability to simultaneously meet multiple consumption needs and desires. The balanced consumption structure of this cluster reflects the consumption upgrading phenomenon documented in previous research, where households move beyond basic needs toward more comprehensive consumption baskets that include education, healthcare, and leisure (Chen & Wang, 2021).

3.3 Socioeconomic Correlates of Consumption Cluster

Analysis of the socioeconomic characteristics associated with each consumption cluster reveals systematic patterns that enhance our understanding of the determinants of consumption behavior. The necessity-oriented cluster is predominantly composed of lower-income households, with average household income approximately 40% below the sample mean. These households typically have lower educational attainment, with over 60% of household heads having completed only secondary education or less. They are also characterized by larger household sizes and higher dependency ratios, suggesting that resource constraints per capita may contribute to their focus on basic consumption needs (Yang & Wang, 2022). The demographic profile of this cluster aligns with theoretical expectations regarding the

relationship between socioeconomic status and consumption patterns, where limited economic resources constrain the ability to diversify consumption beyond essential categories (Deaton & Muellbauer, 1980).

The education-focused cluster exhibits distinct socioeconomic characteristics, with higher-than-average educational attainment and a predominance of households with school-aged children. Over 75% of households in this cluster have at least one child under 18, and household heads are significantly more likely to have tertiary education compared to other clusters. Income levels in this cluster are moderately above the sample mean, though not as high as in the balanced-spending cluster. The socioeconomic profile of education-focused households supports the hypothesis that educational priorities in consumption are influenced by both household composition and parental education, reflecting the intergenerational transmission of educational values and the strategic investment in children's human capital (Hannum et al., 2021). This finding aligns with sociological perspectives that emphasize the role of cultural capital and parental educational strategies in shaping consumption behavior (Bourdieu, 1984).

The healthcare-prioritized cluster demonstrates a clear age gradient, with an average household head age approximately eight years older than the sample mean. These households are more likely to include elderly members and individuals with chronic health conditions, explaining their heightened healthcare expenditures. Income levels in this cluster are slightly below average, suggesting that healthcare spending may represent a significant financial burden. The socioeconomic characteristics of this cluster highlight the impact of demographic factors and health status on consumption patterns, supporting life-cycle theories of consumption that predict changing expenditure priorities across different stages of the life course (Modigliani, 1966). The financial pressure exerted by healthcare expenditures in this cluster raises important policy questions regarding healthcare affordability and social protection for vulnerable households (Feng et al., 2020).

The balanced-spending cluster is characterized by the most advantaged socioeconomic profile, with the highest average income, educational attainment, and wealth indicators among all clusters. Households in this cluster are more likely to include dual-earner couples and demonstrate greater occupational stability. Their ability to maintain substantial expenditures across all consumption categories reflects both greater economic resources and possibly different consumption preferences shaped by higher education and professional occupations. The socioeconomic profile of this cluster supports theories of consumption that emphasize the role of permanent income and economic security in enabling diversified consumption patterns (Friedman, 1957). The consumption behavior of these households represents what many scholars have identified as the emerging Chinese middle-class consumption pattern, characterized by balanced attention to both material and experiential consumption (Li & Song, 2020).

3.4 Regional Variations in Cluster Distribution

The geographic distribution of consumption clusters reveals significant regional disparities that reflect China's uneven economic development and regional inequality. The necessity-oriented

cluster is disproportionately concentrated in less developed regions, particularly in central and western provinces where income levels are lower and economic opportunities are more limited. Within provinces, these households are more likely to reside in smaller cities and less urbanized areas, supporting the hypothesis that regional economic development significantly influences consumption patterns (Wei & Zhang, 2019). The geographic concentration of necessity-oriented consumption in less developed regions highlights the persistent challenges in raising living standards and promoting consumption upgrading across all regions of China, despite overall economic growth.

The education-focused cluster shows a distinct geographic pattern, with higher concentrations in regions with competitive educational systems and strong traditions of educational attainment. These households are particularly prevalent in eastern coastal provinces and major metropolitan areas, where educational competition is intense and returns to education are perceived to be higher. The regional distribution of education-focused households reflects what some scholars have described as the "education fever" phenomenon in Chinese society, particularly in developed regions where educational credentials are seen as crucial for maintaining or improving social status (Hannum et al., 2021). This geographic pattern also aligns with regional disparities in educational resources and quality, which may compel households in regions with better educational opportunities to invest more heavily in education.

The healthcare-prioritized cluster demonstrates interesting geographic variations, with higher concentrations in regions characterized by older population demographics and in areas with less comprehensive healthcare coverage. Contrary to expectations, these households are not exclusively concentrated in less developed regions; rather, they appear in both developed and less developed areas, suggesting that healthcare prioritization is influenced by a combination of demographic factors and local healthcare system characteristics rather than economic development alone (Feng et al., 2020). This finding complicates simplistic narratives about regional consumption patterns and highlights the need for nuanced understanding of how different factors interact to shape consumption behavior across geographic contexts.

The balanced-spending cluster is overwhelmingly concentrated in more developed regions, particularly first-tier and second-tier cities with higher income levels and more diversified consumption opportunities. The geographic distribution of this cluster closely mirrors patterns of regional economic development, with the highest concentrations in the economically advanced eastern coastal regions and major urban centers. This alignment between balanced consumption patterns and regional development supports theories that link consumption diversification with economic modernization and urbanization (Chen & Wang, 2021). The strong regional dimension of balanced consumption patterns underscores the persistent regional inequalities in China's development trajectory and highlights the challenges in spreading consumption upgrading beyond the most developed regions.

3.5 Implications of Consumption Heterogeneity

The identified consumption clusters have significant implications for understanding contemporary Chinese society and designing targeted interventions. The persistence of a substantial necessity-oriented cluster, despite decades of economic growth, highlights the

incomplete nature of China's consumption upgrading and the continued challenges faced by lower-income urban households. This finding qualifies optimistic narratives about universal consumption upgrading in urban China and underscores the need for policies that address the specific constraints facing necessity-oriented households (Yang & Wang, 2022). For businesses, the existence of this cluster suggests continued market opportunities for affordable basic goods and services, though marketing strategies must account for the budget constraints and consumption priorities of these households.

The emergence of distinct education-focused and healthcare-prioritized clusters reflects the growing importance of human capital and health in Chinese household decision-making. These consumption patterns represent responses to both opportunity structures, in the case of education investment, and risk management, in the case of healthcare spending (Feng et al., 2020). The significant allocation of resources to these areas by substantial segments of the urban population has important implications for social policy, particularly regarding the affordability and accessibility of education and healthcare services. For businesses, these clusters represent specialized market segments with distinct needs and preferences that require tailored product offerings and marketing approaches.

The balanced-spending cluster exemplifies the consumption patterns associated with China's emerging middle class and represents the direction of consumption evolution as incomes continue to rise. The ability of these households to maintain substantial expenditures across multiple categories reflects both their economic security and their diversified consumption preferences (Li & Song, 2020). This cluster represents an attractive market segment for businesses offering premium and diversified products and services, though competition for these consumers is likely intense. From a policy perspective, the consumption behavior of this cluster suggests successful integration into modern consumer society, though their relative privilege highlights the distance that other households must travel to achieve similar consumption flexibility.

The heterogeneity in consumption patterns documented in this analysis challenges homogeneous conceptions of Chinese urban consumers and underscores the importance of segment-specific approaches in both business strategy and social policy. The coexistence of different consumption clusters with distinct socioeconomic correlates and geographic distributions reflects the complex and stratified nature of Chinese urban society following decades of rapid economic transformation (Wei & Zhang, 2019). Understanding this heterogeneity is crucial for developing effective interventions that account for the diverse needs, constraints, and aspirations of different household types in contemporary urban China.

Chapter 4: Conclusion and Future Directions

4.1 Key Findings

This research has successfully identified and characterized four distinct consumption clusters among Chinese urban households through the application of K-means clustering analysis to nationally representative data from the China Household Finance Survey. The empirical

typology reveals substantial heterogeneity in consumption patterns that aligns with the abstract's emphasis on understanding household consumption dynamics for market segmentation and policy formulation. The necessity-oriented cluster, comprising 38% of households, demonstrates a consumption structure dominated by basic expenditures on food and housing, reflecting persistent budget constraints among a significant segment of the urban population despite decades of economic growth (Chamon & Prasad, 2010). The education-focused cluster, representing 22% of households, exhibits a distinctive pattern of prioritizing human capital investment through substantial education expenditures, supporting findings about the increasing emphasis on educational attainment in Chinese society (Hannum et al., 2021).

The healthcare-prioritized cluster, accounting for 18% of households, shows a consumption pattern characterized by elevated healthcare expenditures, likely influenced by demographic aging and health consciousness as identified in previous research (Feng et al., 2020). Finally, the balanced-spending cluster, comprising the remaining 22% of households, demonstrates the most diversified consumption pattern across all categories, reflecting the consumption upgrading phenomenon associated with China's emerging middle class (Li & Song, 2020). These findings collectively reveal the complex consumption landscape in urban China and validate the research hypothesis regarding the existence of four distinct household types with different expenditure priorities.

The socioeconomic and regional analysis further enriches our understanding of these consumption patterns. Lower-income households with limited educational attainment are disproportionately represented in the necessity-oriented cluster, while education-focused households typically have higher educational levels and school-aged children. Healthcare-prioritized households show a clear age gradient with older household heads, and balanced-spending households exhibit the most advantaged socioeconomic profile with the highest income and educational attainment. Regional variations in cluster distribution reflect China's uneven economic development, with necessity-oriented households concentrated in less developed regions and balanced-spending households overwhelmingly located in more developed areas (Wei & Zhang, 2019). These patterns align with theoretical frameworks that emphasize the role of economic constraints, demographic factors, and regional context in shaping consumption behavior (Deaton & Muellbauer, 1980).

4.2 Significance and Limitations of the Research

This research makes several significant contributions to the understanding of Chinese household consumption behavior. Theoretically, it advances the literature by developing an empirically derived typology that captures the complexity and heterogeneity of contemporary consumption patterns in urban China, addressing a critical gap in previous research that often examined consumption categories in isolation or relied on theoretically imposed classifications (Gao & Sun, 2021). The application of clustering analysis to comprehensive household expenditure data represents a methodological improvement over previous approaches, providing a more nuanced understanding of how households allocate resources across competing needs and wants. The identification of distinct consumption clusters with clear

socioeconomic correlates supports and extends existing theories about consumption behavior, particularly those emphasizing the role of permanent income, life-cycle factors, and cultural capital in shaping expenditure decisions (Friedman, 1957; Bourdieu, 1984).

From a practical perspective, the research offers valuable insights for businesses and policymakers operating in the Chinese market. The consumption typology enables more precise market segmentation and targeted marketing strategies, allowing businesses to align their product offerings and promotional activities with the specific needs and preferences of different household types (Wedel & Kamakura, 2012). For policymakers, the findings provide evidence-based guidance for designing targeted social and economic programs that address the specific constraints and priorities of different household types. For instance, the persistence of a substantial necessity-oriented cluster highlights the need for policies that address basic consumption needs among lower-income urban households, while the emergence of education-focused and healthcare-prioritized clusters underscores the importance of ensuring affordability and accessibility in these critical service areas (Yang & Wang, 2022).

Despite these contributions, the research has several limitations that should be acknowledged. The reliance on cross-sectional data limits our ability to examine how consumption patterns evolve over time and how households transition between different consumption clusters in response to economic shocks or life events. While the CHFS provides comprehensive consumption data, certain aspects of consumption behavior, particularly qualitative dimensions related to consumption motivations and decision-making processes, are not fully captured by expenditure data alone (Zhou, 2020). The clustering methodology, while robust, involves certain subjective decisions regarding variable selection, standardization, and cluster interpretation that may influence the resulting typology. Additionally, the research focuses exclusively on urban households, leaving open questions about consumption patterns in rural areas and the rural-urban consumption continuum.

4.3 Future Research Directions

Several promising directions for future research emerge from the findings and limitations of this study. Longitudinal analysis of consumption patterns would provide valuable insights into how household consumption evolves over time and how external shocks, such as economic downturns or health crises, influence consumption priorities and cluster transitions. Panel data tracking the same households over multiple years would enable researchers to examine the dynamics of consumption behavior and identify factors that facilitate or hinder consumption upgrading (Li et al., 2021). Such longitudinal approaches would complement the cross-sectional perspective offered in this research and provide a more comprehensive understanding of consumption dynamics.

Future research could also extend the consumption typology to incorporate qualitative dimensions of consumption behavior, including motivations, values, and decision-making processes that underlie expenditure patterns. Mixed-methods approaches combining quantitative expenditure data with in-depth interviews or focus groups would enrich our understanding of the meanings and social contexts associated with different consumption

patterns (Creswell & Plano Clark, 2017). This qualitative dimension is particularly important for understanding the cultural and psychological factors that influence consumption decisions beyond economic constraints and demographic characteristics.

Another fruitful direction involves expanding the geographical scope of analysis to include rural households and examine urban-rural differences and interactions in consumption behavior. Given the significant disparities in economic development and consumption environments between urban and rural China, comparative analysis of consumption patterns across this divide would provide a more comprehensive picture of Chinese household consumption (Wei & Zhang, 2019). Such research could also explore the consumption behavior of migrant households that maintain connections to both urban and rural contexts, representing an increasingly important segment of the Chinese population.

Methodological innovations in clustering analysis represent another promising avenue for future research. The application of alternative clustering algorithms, such as model-based clustering or neural network approaches, could provide additional insights into consumption patterns and validate the robustness of the typology developed in this research (Hastie et al., 2009). Incorporating additional data sources, such as digital consumption traces from e-commerce platforms or mobility data from smartphone applications, could enrich the consumption variables available for analysis and capture emerging consumption channels and behaviors (Liu et al., 2021).

Finally, future research should explore the implications of consumption heterogeneity for specific policy domains and business strategies in greater depth. Experimental studies examining how different household types respond to specific policy interventions or marketing approaches would provide valuable evidence for designing targeted strategies (Thaler, 2015). Research examining the relationship between consumption patterns and broader outcomes, such as subjective well-being, health status, or educational attainment, would enhance our understanding of the consequences of different consumption priorities for household welfare and social development.

References

- [1] Yang, C., & Meihami, H. (2024). A study of computer-assisted communicative competence training methods in cross-cultural English teaching. *Applied Mathematics and Nonlinear Sciences*, 9(1), 45-63. [`https://doi.org/10.2478/amns-2024-2895`](https://doi.org/10.2478/amns-2024-2895)
- [2] Chen, Rensi. "The application of data mining in data analysis." *International Conference on Mathematics, Modeling, and Computer Science (MMCS2022)*. Vol. 12625. SPIE, 2023.
- [3] Huang, J., & Qiu, Y. (2025). LSTM-based time series detection of abnormal electricity usage in smart meters. Preprints. [`https://doi.org/10.20944/preprints202506.1404.v`](https://doi.org/10.20944/preprints202506.1404.v)
- [4] Wang Y. Efficient Adverse Event Forecasting in Clinical Trials via Transformer-Augmented Survival Analysis. Preprints, 2025. DOI: 10.20944/preprints202504.2001.v1.
- [5] Wang, Y., & Ling, C. (2025). Controlling attributes of .xpt files generated by R. *PharmaSUG 2025 conference proceedings*. San Diego, CA.

- [6] Moradi, M., Moradi, M., Bayat, F., & Nadjaran Toosi, A. (2019). Collective hybrid intelligence: towards a conceptual framework. *International Journal of Crowd Science*, 3(2), 198-220.
- [7] Pajic, V., Andrejic, M., & Chatterjee, P. (2024). Enhancing cold chain logistics: A framework for advanced temperature monitoring in transportation and storage. *Mechatron. Intell Transp. Syst*, 3(1), 16-30.
- [8] Singhal, P., Walambe, R., Ramanna, S., & Kotecha, K. (2023). Domain adaptation: challenges, methods, datasets, and applications. *IEEE access*, 11, 6973-7020.
- [9] Habib, M. (2019). Control system design for a solar receiver-reactor.
- [10] Zhou, L., Wu, Z., Sun, M., Park, J., Han, M., Wang, M., ... & Song, E. (2023). Flexible, ultrathin bioelectronic materials and devices for chronically stable neural interfaces. *Brain-X*, 1(4), e47.
- [11] Patel, K. R. (2023). Harmonizing sustainability, functionality, and cost: navigating responsible packaging innovations in modern supply chains. *American Journal of Economic and Management Business (AJEMB)*, 2(8), 287-300.
- [12] Karaboga, D., & Kaya, E. (2019). Adaptive network based fuzzy inference system (ANFIS) training approaches: a comprehensive survey. *Artificial Intelligence Review*, 52(4), 2263-2293.
- [13] Rejeb, A., Keogh, J. G., Zailani, S., Treiblmaier, H., & Rejeb, K. (2020). Blockchain technology in the food industry: A review of potentials, challenges and future research directions. *Logistics*, 4(4), 27.
- [14] Álvarez-Carulla, A., Saiz-Vela, A., Puig-Vidal, M., López-Sánchez, J., Colomer-Farrarons, J., & Miribel-Català, P. L. (2023). High-efficient energy harvesting architecture for self-powered thermal-monitoring wireless sensor node based on a single thermoelectric generator. *Scientific Reports*, 13(1), 1637.
- [15] Iversen, L. J. L., Rovina, K., Vonnice, J. M., Matanjun, P., Erna, K. H., 'Aqilah, N. M. N., ... & Funk, A. A. (2022). The emergence of edible and food-application coatings for food packaging: a review. *Molecules*, 27(17), 5604.