Frontiers in Business, Economics and Management
ISSN: 2766-824X | Vol. 16, No. 2, 2024

Research on Automobile User Labeling Library Based on Factor Analysis

Zhao Liu *, Chenyi Xing

China Automotive Technology and Research Center Co., Ltd., Tianjin, 300300, China

* **Corresponding author**: Zhao Liu (Email: Liuzhao@catarc.ac.cn)

Abstract: The purpose of this paper is to explore in-depth the demographic characteristics of automobile users of luxury brands, traditional joint-venture brands, traditional independent brands, new brands of traditional automobile enterprises and new power brands, as well as the differences and commonalities of their attitude on automobile consumpt by means of a questionnaire survey. First, statistical methods were used to analyze and reveal the significant differences in basic characteristics between different brands of automobile users, such as comparing and finding significant differences in income levels between luxury brand users and independent brand users. Secondly, through the factor analysis technique, a few representative factors are extracted from the complicated conceptual descriptors to simplify and clearly present the core dimensions of attitude on automobile consumpt of users of different brands, so as to provide a scientific basis for automobile manufacturers to formulate market segmentation and marketing strategies.

Keywords: Automobile users, Tag pool, Basic demographic characteristics, Significant difference test, Factor analysis.

1. Introduction

1.1. Background and Significance of the Study

1.1.1. Background of the study

In today's increasingly diversified automotive market, brand competition has become more and more intense, not only in terms of technological innovation and product quality, but also in terms of the ability to deeply understand and accurately position the target consumer groups. Luxury brands, traditional joint venture brands, traditional independent brands, new brands of traditional automobile enterprises and new brands of new forces each carry different brand stories, technological deposits and market positioning, attracting consumer groups with different characteristics and preferences. Therefore, an in-depth understanding of the demographic characteristics of the users of these brands and the attitude on automobile consumpt behind them is key for automakers to develop differentiated market strategies, enhance brand loyalty, and expand market share.

1.1.2. Significance of the Study

(1) Theoretical contributions

This study enriches the theoretical system of automobile user behavior research through the comprehensive use of multiple data analysis methods. It not only explores the differences in the demographic characteristics of automobile users of different brands, but also analyzes in depth the attitude on automobile consumpt behind these characteristics, providing new perspectives and empirical evidence for research in related fields.

(2) Practical guidance

Precise Market Positioning: Based on the in-depth understanding of the characteristics of users of different brands of automobiles, automobile manufacturers can more accurately position themselves in the target market and define the core value proposition of the brand, so as to formulate a more targeted marketing strategy.

Differentiated marketing strategies: Understanding the differences in attitude on automobile consumpt among users of different brands helps automakers design differentiated marketing campaigns and communication messages to better reach and engage their target consumer groups.

1.2. Purpose of the Study and Formulation of the Problem

1.2.1. Purpose of the study

This study aims to comprehensively and deeply explore the demographic characteristics of automobile users of luxury brands, joint-venture brands, independent brands, new brands of traditional automobile enterprises and new power brands, as well as the differences in attitude on automobile consumpts reflected behind these characteristics, and to construct a highly efficient user labeling library, through systematic questionnaire surveys and in-depth data analyses.

1.2.2. Problem formulation

In the current context of increasingly fierce competition in the automotive market, automobile manufacturers are faced with the challenge of how to accurately position their target markets, understand user needs, and formulate effective market strategies. In response to this situation, this study asks the following core questions:

What are the significant differences in the demographic characteristics of car users across brands (luxury brands, traditional joint venture brands, traditional autonomous brands, new brands of traditional car companies and new power brands)?

What is the attitude on automobile consumpt of these users? Can representative core dimensions be extracted through factor analysis?

1.3. Overview of the Scope and Methodology of the Study

1.3.1. Scope of the study

This study focuses on five major brand categories in China's automotive market: luxury brands, joint venture brands, independent brands, new brands of traditional car companies and new power brands. The scope of the study

covers the demographic characteristics of the users of these brands, including but not limited to basic information such as gender, age, marital status, occupation, job title, income and education; at the same time, the study explores the users' attitude on automobile consumpt, which cover a variety of dimensions, including the users' recognition of brand values, considerations in the decision-making process of purchasing a car, and their preferences for car performance and design.

1.3.2. Overview of the methodology

In order to achieve the research objectives and to address the questions posed, the following research methodology will be used in this study:

Questionnaire method: A scientifically designed questionnaire will be distributed to the target respondents through online or offline channels to collect data on their demographic characteristics and attitude on automobile consumpt. The questionnaire will be pre-tested to ensure its validity and reliability.

Literature review method: through reviewing relevant books, journals, reports and other literature to understand the current status and cutting-edge dynamics of domestic and international research on automobile user behavior, market segmentation and marketing strategy, to provide theoretical support and background information for this study.

Statistical analysis method: descriptive statistics are used to analyze the basic characteristics of the sample, and statistical methods such as the chi-square test are used to reveal the significant differences in the demographic characteristics of car users of different brands.

Factor analysis: Factor analysis was conducted on the descriptors of diversified perceptions involving attitude on automobile consumpt in the questionnaire, and a few representative factors were refined through dimensionality reduction to simplify and clearly present the core dimensions of these perceptions of the users of each brand.

2. Literature Review

2.1. Current Status of Related Research at Home and Abroad

Zexing Wang [1] and his research team skillfully integrated big data analysis techniques to systematically sort out the in-depth mining and application of user profiling, user and product images in the field of new energy vehicles, and deeply analyzed the examples of automobile user and product image construction, which provided valuable insights for the industry. Jian Wang [2] and his team further innovated by constructing a model to predict the risk level of driving behavior of automobile users. The model realizes accurate grading of drivers' risky behaviors. Li Yongpan [3] and other scholars focus on the behavior of electric vehicle users, and through in-depth analysis of heat maps and charging time distribution, they accurately extract the typical behavioral characteristics of electric vehicle users, which provides data support for the formulation of electric vehicle marketing strategies. Qiu Lixi [4] 's team, on the other hand, constructed an evaluation system of automobile consumers' values containing 14 indicators through extensive data collection of real car owners, combined with exploratory and validation factor analysis, and deepened the understanding of consumers' psychology and preferences. Cheng Deng [5] et al. took a different approach by relying on vehicle driving data and innovatively utilizing DBSCAN and K-means hybrid clustering algorithm to deeply analyze user driving data and successfully predict user's residence, which greatly enriched the dimension and precision of the new energy vehicle user profiles.

2.2. Theoretical Foundations

Factor Analysis Theory: Factor analysis is a statistical method used to extract a few representative factors from multiple variables in order to simplify the data structure and explain the intrinsic relationship between variables. In the fields of psychology and market research, factor analysis is often used to refine the core dimensions of complex conceptual systems such as consumers' values and consumption views. In this study, the factor analysis technique will be used to reduce the dimensionality of the diversified conceptual descriptors collected involving attitude on automobile consumpt, and extract the key factors that can represent the conceptual characteristics of users of different brands of automobiles.

2.3. Innovative Points of This Paper

(1) Multi-dimensional Comprehensive Comparative Analysis: Through a questionnaire survey, this paper comprehensively collects the demographic characteristics and multi-dimensional data on attitude on automobile consumpt of automobile users of luxury brands, traditional joint venture brands, traditional independent brands, new brands of traditional automobile enterprises, and new power brands, and realizes a comprehensive comparative analysis of the users among different brands.

(2) In-depth factor analysis and refinement of core dimensions: Using factor analysis technology, a few representative factors are extracted from a large number of conceptual descriptors, clearly presenting the core automotive dimensions of each brand's users, which provides automobile manufacturers with a scientific basis for in-depth understanding of users' psychology and behavior.

3. Research Methodology

3.1. Questionnaire Design and Data Collection

3.1.1. Questionnaire design

This study collects and analyzes multi-dimensional data on respondents' basic information, population characteristics, interests and attitude on automobile consumpt by means of a questionnaire in order to gain an in-depth understanding of the characteristics of the target group and their potential demand for new energy vehicles.

3.1.2. Data collection

This study uses online questionnaire as the main data collection tool, aiming to gain an in-depth understanding of car purchasing owners in a specific time period (from May 1, 2023 to April 30, 2024). We collected data on demographic characteristics, including gender, age, marital status, occupation, job title, income, education, etc., as well as diversified conceptual descriptors related to automobile concepts, and chose 70 automobile brands that have been active in the market in recent one year with a high degree of online popularity as our research targets, which cover luxury brands These brands cover luxury brands, traditional joint venture brands, traditional independent brands, new brands of traditional automobile enterprises and new power brands, etc. to ensure the breadth and representativeness of the sample.

3.2. Data Processing and Analysis Methods

3.2.1. Data processing

(1) Missing value processing: first check all collected questionnaire data, and for options with missing values, especially for key variables (e.g., city, age, gender, etc.), try to fill them in through callbacks, automatic coding rules (e.g., extrapolation based on known information), or the use of appropriate statistical methods (e.g., interpolation of means, interpolation of plurals). Missing data that cannot be reasonably filled in are marked or deleted.

(2) Outliers handling: Check for outliers in the data, such as age filled in too large or too small, income filled in not in line with the norm, etc., conduct manual review and decide to retain, correct or delete.

(3) Data formatting: Ensure that all data are formatted in a uniform way, e.g., date, time, numbers, etc., to facilitate subsequent analysis.

3.2.2. Methods of analysis

(1) Descriptive statistical analysis: calculating basic statistics such as frequency distributions for each variable to describe the overall demographic characteristics of a sample.

(2) Significant difference test: use chi-square tests statistical method to compare whether there are significant differences in demographic characteristics, attitude on automobile consumpt labels among different brands of automobile users.

(3) Factor analysis: For descriptions of perceptions involving multiple variables (e.g.attitude on automobile consumpt), factor analysis was conducted to extract potential common factors to simplify the data structure and reveal the intrinsic relationships between variables, and to explain the meaning of each factor.

4. Basic Characterization and Significant Difference Tests for Different Brands of Automobile Users

4.1. Income Status

H8. What is your annual household income? [Single choice]
* Brand type Cross-tabulation

Table 1. Percentage of total

		Brand Type					(grand) total
		luxury brand	Heritage Joint Venture Brands	Traditional independent brands	New brands for traditional car companies	new power brand	
H8. What is your annual household income? [Single choice	Less than 50,000	0.3%	0.9%	0.7%	0.1%		2.0%
	50,000-100,000 (excl.)	0.3%	1.3%	4.0%	0.2%	0.3%	6.0%
	100,000-150,000 (excl.)	1.1%	5.4%	5.5%	1.0%	0.2%	13.1%
	150,000-200,000 (excl.)	1.2%	4.3%	5.2%	1.2%	0.9%	12.7%
	200,000-250,000 (excl.)	0.9%	7.7%	4.0%	1.5%	1.7%	15.8%
	250,000-300,000 (excl.)	1.5%	5.5%	3.6%	1.2%	1.8%	13.6%
	300,000-350,000 (excl.)	2.8%	3.1%	2.5%	1.2%	2.0%	11.6%
	350,000-400,000 (excl.)	3.5%	1.4%	1.5%	1.0%	1.5%	8.9%
	400-450,000 (excl.)	2.1%	1.1%	2.5%	0.7%	1.3%	7.7%
	450,000-500,000 (excl.)	2.9%	1.1%	0.3%	0.4%	1.0%	5.5%
	500,000-1,000,000 (excl.)	1.3%	0.5%	0.0%	0.4%	0.6%	2.8%
	1 million and above	0.2%		0.0%			0.2%
(grand) total		18.1%	32.1%	29.8%	8.7%	11.3%	100.0%

Table 2. Chi-square (math.) test

	(be) worth	(number of) degrees of freedom (physics)	Asymptotic significance (both ways)
Pearson's chi-square (math.)	2664.899a	44	.000
Likelihood ratio (L)	2705.752	44	.000
linear correlation	20.487	1	.000
Number of active cases	9996		
a. 3 cells (5.0%) have expected counts less than 5. The minimum expected count is 1.57.			

The income status of the largest number of people who prefer luxury brands is 350,000-400,000, the income status of the largest number of people who prefer traditional joint venture brands is 200,000-250,000, the income status of the largest number of people who prefer traditional autonomous brands is 100,000-150,000, the income status of the largest number of people who prefer new brands from traditional automobile companies is 200,000-25,000, and the income status of the largest number of people who prefer new brands of new forces is 300,000-350,000.

The degree of freedom is 108 and the Pearson's chi-square value is 1193.908 and the significance value is 0.000 less than 0.05, which is a significant difference.

5. Factor Analysis of Attitude on Automobile Consumpt

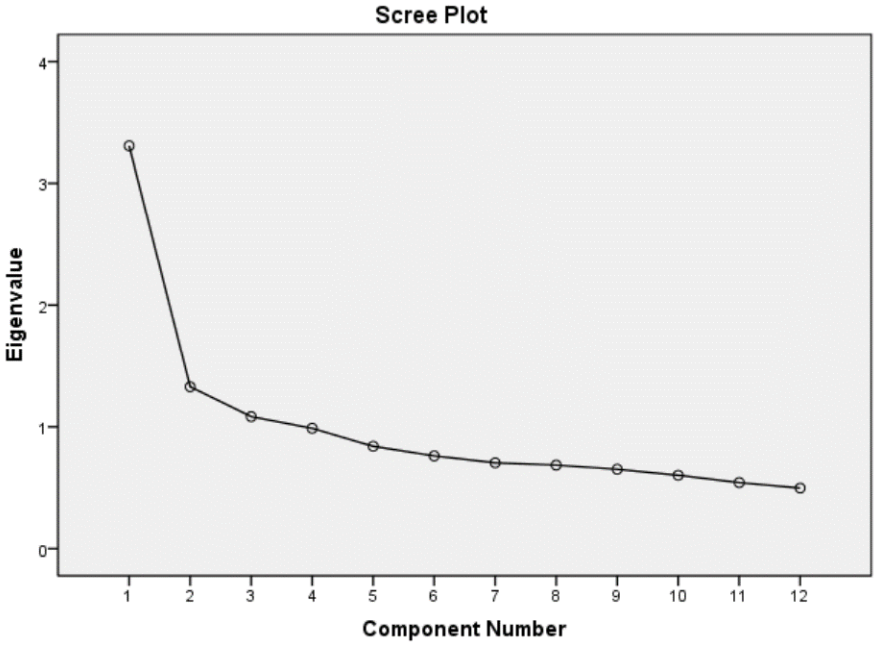

Figure 1. Analysis of the eigenvalues of the factors

As can be seen from the figure, the eigenvalue (variance contribution) of the first factor is very high, which contributes the most to explaining the original question items; the eigenvalues of the factors from the fourth onwards are small, and the contribution to explaining the original question items is very small, which can be ignored, so it is more appropriate to extract three factors.

Table 3. Rotated component matrix[a]

	subassemblies			
	1	2	3	4
H12. Cars are just transportation, as long as they are practical.				.906
H12. Cars as a symbol of face and social status	.761			
H12. Car is not only a transportation tool, but also a pleasure of life		.791		
H12. Owning a car makes me feel independent and free		.691		
H12. the car is a personal space to escape the stresses of life				
H12. don't care about the brand if the vehicle is good			.792	
H12. Brands matter when buying a car			-.729	
H12. I want my car to have a lot of personality	.618			
H12. I like to transform my car to my liking				
H12. I want my car to drive by and passersby to be very envious of it	.774			
H12. Driving a car is an unparalleled pleasure.		.521		
H12. price is secondary as long as the model is to my liking	.577			
Extraction method: principal component analysis. Rotation method: Kaiser standardized maximum variance method.				
a. The rotation has converged after 5 iterations.				

As can be seen from the table, "the car is a symbol of face and social status" "I want my car to be very personalized" "I want my car to be driven over, passers-by will be very envious" "As long as the car model I like, the price is secondary" is located on the first factor has higher loadings, the first factor mainly explains these items, interpreted as personality show off; "A car is not only a means of transportation, but also a pleasure of life" "Owning a car makes me feel independent and free" "Driving a car is a kind of unrivaled enjoyment" is located in the second factor with high loadings, the second factor mainly explains these items, which are interpreted as enjoying the freedom of driving;" If the vehicle is good, I don't care about the brand" "Brand is very important when buying a car" is located on the third factor with high loadings, and the third factor mainly explains these items, which are explained as privacy; "A car is just a means of transportation, it's practical as long as it's useful. "Cars are just a means of transportation, just practical" has a high loading on the fourth factor, which mainly explains these items and is interpreted as practical and heavy items.

Table 4. Chi-square (math.) test

	(be) worth	(number of) degrees of freedom (physics)	Asymptotic significance (both ways)
Pearson's chi-square (math.)	69.111a	12	.000
Likelihood ratio (L)	69.915	12	.000
linear correlation	1.445	1	.229
Number of active cases	10001		
a. 0 cells (0.0%) have an expected count of less than 5. The minimum expected count is 206.86.			

From the above table, the chi-square test (cross-tabulation analysis) was used to investigate the relationship between the

differences in the car view labels on the type of preferred brands, as can be seen from the above table, the different car view labels show significance ($p<0.05$) for the type of preferred brands, meaning that the different car view labels show significant differences for the type of preferred brands.

6. Conclusion

Through a comprehensive questionnaire survey and fine-grained data analysis, we deeply explored the demographic characteristics of automobile users of luxury brands, traditional joint venture brands, traditional independent brands, new brands of traditional automobile enterprises and new power brands, as well as the differences in their respective attitude on automobile consumpt. Our study first collects data on demographic characteristics covering multiple dimensions such as gender, age, marital status, occupation, position, income, and education, etc., and through statistical analysis methods, we reveal the significant differences in these basic characteristics between auto users of different brands, such as the clear distinction between the income levels of luxury brand users and those of autonomous brands. Further, we utilize factor analysis techniques to extract "Individuality", "Driving Freedom", "Private Space", and "Practicality and Quality" in terms of attitude on automobile consumpt. The core and representative factors of "practicality and quality" accurately summarize the core dimensions of each brand's user values, consumption and attitude on automobile consumpt. This process not only simplifies the presentation of data, but also helps us to better understand the intrinsic needs and preferences of different brands' user groups. Through the results of these analyses, we arrive at a multi-dimensional user labeling system that can effectively distinguish and describe the differences between different brands of automobile users. Ultimately, these user labels, provide us with a clear and effective tool to differentiate and study the consumer differences between different brands. Based on these user labels, an efficient user label library can be constructed, which can not only help enterprises accurately identify target customer groups, but also provide powerful support in product development, marketing, customer service and other aspects. By continuously optimizing and improving the user tag library, automobile manufacturers can more accurately grasp the market dynamics and meet the diversified needs of consumers, thus occupying a favorable position in the fierce market competition.

References

[1] WANG Zexing, HAN Boyang, LIN Huiguang, et al. Research progress of new energy vehicle portrait based on big data analysis [J]. Automotive Practical Technology, 2023, 48(19):194-200. DOI:10.16638/j.cnki.1671-7988.2023.019.038.

[2] WANG Jian, QUO Lili, PEI Chunqin, et al. Research on car user behavior analysis method based on improved K-mean clustering algorithm [J]. Journal of Yanshan University, 2023, 47(03):229-235+245.

[3] LI Yongpan, HUANG Bing, XIE Da. Visual analysis of electric vehicle user behavioral features based on k-means clustering [J]. Electrical Automation, 2019, 41(01):12-15+81.

[4] Qiu, L. H.. Establishing an evaluation model of automobile consumer values [J]. China Automobile, 2022, (10):55-59.

[5] CHENG Deng, ZHANG Liang, ZHAO Xiaoyu, et al. New energy vehicle user residence prediction based on clustering algorithm [J]. Automobile Practical Technology, 2021, 46(10):11-13. DOI:10.16638/j.cnki.1671-7988.2021.010.003.