# Towards Better Understanding for Consumer Personalization in Marketing

Jincheng Wang[1,*,+], Yihui Sun[2,+]

[1]Financial Management, Hainan University, Danzhou, 571737, China
[2]Accounting and Finance, University of Surrey, Guildford, GU2 7X, UK
[+]These authors contributed equally to this work and should be considered co-first authors.

*Corresponding author. Email: 1842089303@qq.com

**Abstract.** The purpose of this study is to explore the role of customer attributes in explaining customer consumption preferences. Through the different personal situation of customers, the consumption choice of customers is classified and discussed. Specifically, the annual household income, the customer's education background, age, marital status and whether there are minors in the family have different degrees of influence on the consumption type of the customer. This can effectively help enterprises classify different ideal customer groups. This study used SPSS auxiliary research data. The practical problems are studied by statistical methods such as multiple linear regression analysis, factor analysis and normal test. Through them to the subject questions to carry out data interpretation and answer. Three conclusions are drawn from our research: 1. The difference between high and low household annual income will affect customers' consumption choices. 2.The differences in education, age and marital status will affect the annual household income of customers and thus affect their consumption. 3.Whether there are minors at the customer's home will affect the expenditure of wine and sweet. This study provides new insights into enterprise personalization analysis. It is helpful for enterprises to formulate marketing strategies and realize precise operation when facing different customers. Carry out different personalized services for different customer groups. The limited resources will be allocated to customers with different personalities to achieve the maximum benefit.

**Keywords:** Customer personalized classification 1, Household annual income 2, Customer attributes 3, Consumption type 4.

## 1 Introduction

Having a better understanding of customer attributes can help enterprises sort out customer groups, which will be subdivided according to certain standards to get customer portraits in the subsequent operations. Customer portrait is a labialized customer model abstracted from social attributes, living habits and consumption behaviors of customers. The core work of constructing customer portrait is to give the customer a highly refined personalized identity obtained through customer information analysis. In the marketing


K. Elbagory et al. (eds.), *Proceedings of the 9th International Conference on Financial Innovation and Economic Development (ICFIED 2024)*, Advances in Economics, Business and Management Research 281,
https://doi.org/10.2991/978-94-6463-408-2_27

campaign, Chandra et al. the process of customization involves creating products and services that appeal to the preferences of the customer [1]. Customers' cognitive load can be reduced by personalized content and products based on their preferences. This process effectively aids businesses in understanding their customers and makes it simpler for them to customize their offerings to the distinct needs, habits, and concerns of various customer types.

Customer Personality Analysis (CRM) with big data enables enterprises to become more active in marketing strategies [2]. Companies can utilize big data to determine which segment of consumers is most likely to purchase a product, then market items just to this segment of customers, as opposed to paying money to advertise new products to every client in the corporate database. Huang and Chia state that by boosting customer connections and customer conversion rates, a successful personalization approach can boost revenue and profitability, as well as also promote customer loyalty [3]. As Montgomery and Michael perceptively states that a crucial element of interactive marketing techniques is personalization [4]. The goal is to modify standardized goods or services to meet the demands of various clients. Profits for producers and value for customers are the objectives. The conventional idea of segmentation works well with this goal. With the advent of the Internet, which offers a useful and suitable setting for interaction, personalization apps have advanced significantly.

This study is based on the data in the data set. Three problems related to customer's personalized analysis are proposed. Using these data, three questions related to customer personalization analysis are set up and answered by statistical methods. Three statistical methods are used: multiple regression analysis, factor analysis and normal test. Each question was eventually answered. Combined with practical experience, these conclusions are summarized and discussed. At the end of the study, new insights about customer personalized analysis are obtained.

## 2 Selection of data and statistical methods

### 2.1 Date sources

In this study, the data set Customer Personality Analysis of Kaggle website was studied. In this study, three questions and multiple hypotheses were established. Through multiple linear regression analysis, normal testing, and factor analysis, the goodness of fit and significance of the variables are further explained, and the important role of customer attributes in the final business decision is concluded.

### 2.2 Statistical methods

For multiple linear regression, Tranmer and Mark state that multiple linear regression is a popular technique in social science in both theory and practice [5]. firstly, three conclusions can be drawn by using multiple linear regression analysis to explain the original data. The goodness of fit of the data represented by $R^2$ can effectively solve the data and predict the efficiency of linear regression work. Whether the linear relationship between the dependent variable and the independent variable is significant can be accurately judged by the significance test in the contrast analysis, and further inferred whether the initial hypothesis in the study is valid. Based on the non-standardized coefficients in the t-test, the positive and negative correlations between variables can be accurately judged. The precondition for establishing unstandardized coefficients is that at least one term of tolerance and VIF is within a specific range.

The second approach is comparative analysis, including normality tests and chi-square tests. According to Das and Imon, before performing a formal statistical analysis, it is essential to determine whether the data are normally distributed because failing to do so could lead to incorrect inferences and conclusions being made [6]. In contrast to many other non-parametric and some parametric statistics, the calculations needed to produce the Chi-square provide important information on how each of the study's groups performed. The complexity of the details that allow the researcher to understand the results make this statistic more complete than many others [7]. Through comparative analysis, such as normality test and chi-square test, whether there is a significant difference between two combinations of a variable can be effectively judged.

Third, this study uses factor analysis which contains four specific statistical analysis methods. As Yong and Pearce point out that the basic goal of factor analysis is to reduce data in a way that makes correlations and patterns easy to see and understand. Rearranging variables into a few clusters based on shared variance is common [8]. Therefore, it is useful to segregate concepts and constructions. KMO and Bartlett test can determine whether the variables in the data are suitable for factor analysis. The rotated composition matrix then helps the composition of each factor extracted by the factor analysis in the previous step.

### 2.3 Data words substitution

Due to the inconvenient processing of some data in the original data, this study used numbers to replace some data words: in the education section, set Basic as 0, graduation as 1, 2n Cycle as 2, Master as 3, PhD as 4. The marriage section is set to Single and Alone, with 0 widows, 1 Divorced, 2 Together and 3 Married.

## 3 Data analysis and research hypothesis

### 3.1 The influence of annual household income on customers' choice of consumption type

Based on the original data in the table, this part of the study set the annual household income of customers as the independent variable of this hypothesis to study the specific consumption conditions of wine, fruit, meat, fish, sweet and gold respectively, and further explore the differences between the types of commodities consumed by high-income and low-income consumer groups.

### 3.1.1 Multiple linear regression

Statistical methods: SPSS Statistics is used for multiple linear regression analysis of the data, and the significance level of detection is 0.05 (bilateral).

Test results: Among the six groups of data, the significance of influencing factors of household annual income in the comprehensive test of multiple linear regression model coefficient is all less than 0.05, so the regression model rejects the null hypothesis at the significance level of 0.05, indicating that the regression model has statistical explanatory ability, so the col-linearity statistics of the six groups of data are valid. $R^2$ of wine, fruit, meat, fish, sweet and gold are 0.335, 0.186, 0.342, 0.193, 0.194, 0.106, respectively. Since the sample content is a large sample, even if the $R^2$ in the measurement model is low, the fitting degree is good, and the measurement detection can be carried out. In the standardized coefficient of T-test, the constants are listed in order of (largest to smallest) gold, fruit, sweet, fish, wine, meat. Since the average value of household annual income is much higher than the value of all kinds of expenditure, the mean value of independent variables is a small positive value, which has little influence on the total amount.

Table 1 shows data that derived from multiple linear regression analysis, and the six variables are compared with income respectively. The data obtained can answer the goodness of fit, significance and correlation between the six variables and income.

**Table 1.** Data of Multiple linear regression [Owner-draw]

| Category | R Square | ANOVA Sig. | B Constant | Tolerance | VIF |
|---|---|---|---|---|---|
| Wine | .335 | .000 | -100.039 | 1.000 | 1.000 |
| Fruit | .186 | .000 | -9.229 | 1.000 | 1.000 |
| Meat | .342 | .000 | -105.154 | 1.000 | 1.000 |
| Fish | .193 | .000 | -12.235 | 1.000 | 1.000 |
| Sweet | .194 | .000 | -10.543 | 1.000 | 1.000 |
| Gold | .106 | .000 | 8.915 | 1.000 | 1.000 |

### 3.1.2 Factor analysis

Statistical methods: SPSS Statistics is used for factor analysis of the data. Test results: KMO is greater than 0.5, Bartlett's Test less than 0.05, so suitable for factor analysis. According to the Scree Plot and Total Variance Explained, four factors are selected for analysis. According to the Communalities, the research can see that all seven variables have good explanatory ability. According to the Rotated Component Matrix, the composition combination of the four factors can be known.

Table 2 to Table 5 show data that derived from factor analysis, which combines all the variables needed in the question. The data obtained can answer whether these variables are suitable for factor analysis, the interpretability of each variable, and the need to classify and select several factors for study.

**Table 2.** KMO and Bartlett's Test [Owner-draw]

| KMO Measure of Sampling Adequacy | Sig in Bartlett's Test of Sphericity |
|---|---|
| .881 | .000 |

**Table 3.** Communalities [Owner-draw]

| Category | Initial | Extraction |
|---|---|---|
| Income | 1.000 | .762 |
| Wine | 1.000 | .797 |
| Fruit | 1.000 | .807 |
| Meat | 1.000 | .759 |
| Fish | 1.000 | .744 |
| Sweet | 1.000 | .983 |
| Gold | 1.000 | .980 |

**Table 4.** Total Variance Explained [Owner-draw]

| Component | Total | % of Variance | Cumulative % |
|---|---|---|---|
| 1 | 3.846 | 54.946 | 54.946 |
| 2 | .832 | 11.886 | 66.832 |
| 3 | .722 | 10.318 | 77.150 |
| 4 | .431 | 6.614 | 83.314 |
| 5 | .411 | 5.877 | - |
| 6 | .407 | 5.816 | - |
| 7 | .350 | 4.993 | - |

**Table 5.** Rotated Component Matrix [Owner-draw]

| Category | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Income | .811 | .220 | - | .229 |
| Wine | .837 | .146 | .263 | - |
| Fruit | .201 | .839 | .161 | .190 |
| Meat | .626 | .579 | - | .159 |
| Fish | .220 | .739 | .227 | .312 |
| Sweet | .243 | .368 | .135 | .878 |
| Gold | .195 | .223 | .937 | .118 |

### 3.2 The impact of a customers' personal attribute on annual household income

In this part of the study, the original data is used as the basis, and the customer's education background, birth year and marital status are set as independent variables to study whether the above three independent variables will have a corresponding impact on the customer's annual household income. It is hoped to explore whether the differences in the client's education level, age and marital status have different effects on the client's annual household income.

### 3.2.1 Multiple linear regression

Statistical methods: SPSS Statistics is used for multiple linear regression analysis of the data, and the significance level of detection is 0.05 (bilateral).

Test results: In the three groups of data, the $R^2$ of the customer's birth year and education background are 0.026 and 0.032, respectively, and the significance of the comprehensive test of multiple linear regression model coefficient is less than 0.05. Therefore, the regression model rejects the null hypothesis at the significance level of 0.05, which indicates that the regression model has statistical explanatory ability. Therefore, col-linearity statistics of the two groups of data are valid. It can be seen from the T-test that the year of birth of the customer is negatively correlated with the annual family income. On the contrary, the degree of the customer is positively correlated with the annual family income. During the test, the customer's marital status is excluded from the analysis because it had too little correlation with annual household income.

Table 6 shows data that derived from multiple linear regression analysis, and the three variables of the customer's education background, birth year and marital status are compared and analyzed with their income. Goodness of fit, significance and correlation between variables can be obtained from the data.

**Table 6.** Data of Multiple linear regression [Owner-draw]

| Category | R Square | ANOVA Sig. | B Constant | Tolerance | VIF |
|---|---|---|---|---|---|
| Birth Year | .026 | .000 | -342.085 | 1.000 | 1.000 |
| Education | .032 | .000 | 1469.373 | .977 | 1.024 |

### 3.2.2 Discrete variable test

Statistical methods: SPSS Statistics is used to test the discrete variables of data. Test results: When the customer's educational background is 0 (Basic), the comparison between the two is more than 0.05, so the difference is not significant. However, the correlation between marital status and birth year of clients and annual family income is less than 0.05, so the differences between these two conditions are significant respectively.

Table 7 shows data that derived from the normal test, and the key variable income is compared with three independent variables. Through the data, we can draw the conclusion whether the normal hypothesis is rejected between variables, so as to study whether the problem is valid.

**Table 7.** Tests of Normality [Owner-draw]

| Category | Category | Sig in Shapiro-Wilk |
|---|---|---|
| Income with Education | 0 | .206 |
| | 1 | .000 |
| | 2 | .000 |
| | 3 | .000 |
| | 4 | .000 |
| Income with Marital Status | 0 | .000 |
| | 1 | .001 |
| | 2 | .000 |
| | 3 | .000 |

### 3.2.3 Factor analysis

Statistical methods: SPSS Statistics is used for factor analysis of the data. Test results: KMO is greater than 0.5, Bartlett's Test is less than 0.05, so it is suitable for factor analysis. According to the Scree Plot and Total Variance Explained, three factors can be selected for analysis. According to the Communalities, the research can see that all the four variables have good explanatory ability. According to the Rotated Component Matrix, the composition combination of the three factors can be determined.

Table 8 to Table 11 show data that derived from factor analysis, which combines all the variables needed in the question. The data obtained can answer whether these variables are suitable for factor analysis, the interpretability of each variable, and the need to classify and select several factors for study.

**Table 8.** KMO and Bartlett's Test [Owner-draw]

| KMO Measure of Sampling Adequacy | Sig in Bartlett's Test of Sphericity |
|---|---|
| .555 | .000 |

**Table 9.** Communalities [Owner-draw]

| Category | Initial | Extraction |
|---|---|---|
| Birth Year | 1.000 | .505 |
| Education | 1.000 | .870 |
| Marital Status | 1.000 | .999 |
| Income | 1.000 | .807 |

**Table 10.** Total Variance Explained [Owner-draw]

| Component | Total | % of Variance | Cumulative % |
|---|---|---|---|
| 1 | 1.277 | 31.937 | 31.937 |
| 2 | 1.001 | 25.020 | 56.957 |
| 3 | .902 | 22.552Component | 79.509 |
| 4 | .820 | 20.491 | - |

**Table 11.** Rotated Component Matrix [Owner-draw]

| Category | 1 | 2 | 3 |
|---|---|---|---|
| Income | .895 | - | - |
| Birth Year | -.567 | -.427 | - |
| Education | - | .932 | - |
| Marital Status | - | - | .999 |

## 3.3 The influence of minors in the family on wine and sweet expenditures

Based on the original data in the table, the annual income of whether there are children or teenagers in the customer family is taken as an independent variable to study the expenditure of the customer family on wine and sweet. This paper wishes to conclude that there are differences in the consumption of wine and sweet products in households with or without minors.

### 3.3.1 Multiple linear regression

Test results: R2 are 0.246 and 0.168, respectively, with significance less than 0.05. Therefore, the regression model rejected the null hypothesis at the significance level of 0.05, indicating that the regression model had statistical explanatory power, so col-linearity statistics are valid. Among them, the T-test of wine and teenagers is significant, so this combination is not suitable for analysis. In the standardized coefficient of T-test of wine and children, it can be clearly seen that there is a negative correlation between them, and the correlation coefficient is large. There is also a negative correlation between sweet and children or teenagers.

Table 12 shows data that based on multiple linear regression analysis. Two independent variables, children and teenagers, and two dependent variables, wine and sweet, are compared respectively. The goodness of fit, significance and correlation among them can be obtained from the data.

**Table 12.** Data of Multiple linear regression [Owner-draw]

| Category | R Square | ANOVA Sig. | B Constant | Tolerance | VIF |
|---|---|---|---|---|---|
| Kid and Wine | .246 | .000 | -310.572 | .999 | 1.001 |
| Teen and Wine | | | -8.100 | .999 | 1.001 |
| Kid and Sweet | .168 | .000 | -28.908 | .999 | 1.001 |
| Teen and Sweet | | | -13.350 | .999 | 1.001 |

### 3.3.2 Discrete variable test

Statistical methods: SPSS Statistics is used to test the Continuous variable of data. Test results: Four combinations (sweet and teen, sweet and children, wine and teen, wine and children) are significant. Only the combination of sweet and teenagers has a significance greater than 0.05, so the difference of this combination is not significant. The significance of the other three combinations is less than 0.05, so the other three combinations (sweet and children, wine and teen, wine and children) are significantly different.

Table 13 shows data that derived from the Chi-square test. Two independent variables and two dependent variables are combined respectively and analyzed. Through the data, the significance between variables can be obtained, so as to study whether the problem is valid.

**Table 13.** Chi-Square Summary [Owner-draw]

| Category | Asymptotic Sig in Pearson Chi-Square |
|---|---|
| Kid and Wine | .006 |
| Teen and Wine | .000 |
| Kid and Sweet | .000 |
| Teen and Sweet | .214 |

### 3.3.3 Factor analysis

Statistical methods: SPSS Statistics is used for factor analysis of the data. Test results: KMO is greater than 0.5, Bartlett's Test is less than 0.05, so it is suitable for factor analysis. According to the Scree Plot and Total Variance Explained, three factors can be selected for analysis. According to the Communalities, the research can see that all the four variables have good explanatory ability. According to the Rotated Component Matrix, the composition combination of the three factors can be determined.

Table 14 to Table 17 show data that derived from factor analysis, which combines all the variables needed in the question. The data obtained can answer whether these variables are suitable for factor analysis, the interpretability of each variable, and the need to classify and select several factors for study.

**Table 14.** KMO and Bartlett's Test [Owner-draw]

| KMO Measure of Sampling Adequacy | Sig in Bartlett's Test of Sphericity |
|---|---|
| .646 | .000 |

**Table 15.** Communalities [Owner-draw]

| Category | Initial | Extraction |
|---|---|---|
| Kid | 1.000 | .782 |
| Teen | 1.000 | .999 |
| Wine | 1.000 | .725 |
| Sweet | 1.000 | .994 |

**Table 16.** Total Variance Explained [Owner-draw]

| Component | Total | % of Variance | Cumulative % |
|---|---|---|---|
| 1 | 1.826 | 45.644 | 45.644 |
| 2 | 1.004 | 25.111 | 70.755 |
| 3 | .670 | 16.747 | 87.502 |
| 4 | .500 | 12.498 | - |

**Table 17.** Rotated Component Matrix [Owner-draw]

| Category | 1 | 2 | 3 |
|---|---|---|---|
| Kid | -.853 | - | -.159 |
| Teen | - | .996 | - |
| Wine | .841 | - | .190 |
| Sweet | .250 | - | .964 |

## 4 Discussion

For the first question, according to the multiple linear regression and factor analysis of this part, customers with higher annual household income spend more on gold commodities, followed by fruit, sweet and fish, while customers with lower annual household income spend more on meat and wine commodities.

For the second question, according to the research, the following three conclusions are drawn: Firstly, the annual household income of customers with higher education is generally greater than that of customers with lower education. Secondly the annual household income of customers with older age is generally greater than that of customers with younger age. Thirdly, the marital status of customers has no significant impact on their annual household income.

For the third question, the findings of this section are as follows: Firstly, for the customer group with children in the family, the decrease of family consumption in wine expenditure is very obvious. Secondly, for the customer group with teenagers in the family, there is no obvious connection with the wine's expenditure in the family, so the relationship between them cannot be defined. Thirdly, both the customer groups with children in the family and the customer groups with teenagers in the family consume less sweet.

In summary, the demands of customers are various. In particular, different requirements apply to different persons, the same person has different needs in various circumstances, and there may even be variances between different needs in the same person in the same circumstance. The requirements of customers can also be influenced by their living conditions, cultural background, religious beliefs, and the culture, customs, and habits of the nation and area in which they reside. Therefore, marketers can only more successfully assist consumers in flowing economic value to themselves rather than rivals by understanding the peculiarities of their demands and creating the prospective demand market. Due to recent Internet developments and the explosion of e-commerce, businesses may now more affordably and in real time get information on specific clients online. According to Cheung et al. (2003), the efficacy of direct marketing in cyberspace may be considerably increased by using the information acquired to create extensive consumer profiles that support one-to-one marketing and other customized services [9].

Based on the above research, this study proposes the following corporate recommendations: the customer's family income level, the educational background and age of customers, and whether there are minors in the family are all factors that affect the customer's consumption in the paper. However, in real household consumption, the factors that affect customers' personalized consumption are far more than the variables studied in this paper. The point of Riemer and Carsten is that with the development of Internet technology, there have been several chances to build individualized, cost-efficient partnerships with clients [10]. It aspires to provide individualized, customer-focused information and goods. Utilizing the benefits of big data analysis and Internet information interaction is crucial for businesses to analyze a range of consumption-influencing factors. By doing so, it will be relatively simple to offer customized services to various customer groups based on data analysis [11].

## 5 Conclusion

A significant usage and contribution of new management seems necessary in order to accomplish this aim. Enterprises want to implement various methods for various clients to achieve accurate operation and the highest possible conversion rate while creating operation and marketing plans for customers. client relationship management is the foundation of precise operation, and client classification is its essential component. Customer classification allows for the division of customer groups, the identification of low value and high value customers, the provision of various personalized services to different customer groups, the reasonable distribution of scarce resources among customers with various values, and the maximization of benefits.

## Acknowledgments

Jincheng Wang and Yihui Sun contributed equally to this work and should be considered co-first authors.

## References

1. Chandra, S., Verma, S., Lim, W. M., Kumar, S., & Donthu, N. (2022) Personalization in personalized marketing: Trends and ways forward. Psychology & Marketing., 39: 1529-1562.
2. Anshari, M., Almunawar, M. N., Lim, S. A., & Al-Mudimigh, A. (2019) Customer relationship management and big data enabled: Personalization & customization of services. Applied Computing and Informatics., 15: 94-101.
3. Huang, E. Y., & Lin, C. Y. (2005) Customer-oriented financial service personalization. Industrial Management & Data Systems., 105: 26-44.
4. Montgomery, A. L., & Smith, M. D. (2009) Prospects for Personalization on the Internet. Journal of Interactive Marketing., 23: 130-137.
5. Tranmer, M., & Elliot, M. (2008) Multiple linear regression. The Cathie Marsh Centre for Census and Survey Research (CCSR)., 5: 1-5.
6. Das, K. R., & Imon, A. H. M. R. (2016) A brief review of tests for normality. American Journal of Theoretical and Applied Statistics., 5: 5-12.
7. McHugh, M. L. (2013) The chi-square test of independence. Biochemia medica., 23: 143-149.

8. Yong, A. G., & Pearce, S. (2013) A beginner's guide to factor analysis: Focusing on exploratory factor analysis. Tutorials in quantitative methods for psychology., 9: 79-94.
9. Cheung, K. W., Kwok, J. T., Law, M. H., & Tsui, K. C. (2003) Mining customer product ratings for personalized marketing. Decision Support Systems., 35:231-243.
10. Riemer, K., & Totz, C. (2003) The many faces of personalization: An integrative economic overview of mass customization and personalization. The customer centric enterprise: advances in mass customization and personalization., 35-50.
11. Pierański, B., & Strykowski, S. (2017) Towards a personalized virtual customer experience. Advanced Topics in Intelligent Information and Database Systems., 9: 185-195.