Airline Customers Strategic Decision Analysis and Formulation

Chenge Yan [1, *, #], Zebin Wu [1, #], Haoyu Gan [1, #], Ziwen Zhu [2, #]

[1]School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen, China, 518172

[2]School of Management and Economics, The Chinese University of Hong Kong, Shenzhen, China, 518172

*Corresponding author: 119020063@link.cuhk.edu.cn

[#]These authors contributed equally.

Abstract. In recent years, estimating customer value for airline companies has attracted much scholarly attention. Many researchers have tried to classify customer rankings based on different commercial models [1-2] (e.g., Huang & Liu, 2016; Liu & Du, 2010). However, few studies have quantified specific value rankings and designed tailored strategies. This project aims to fractionate customer value rankings and make corresponding strategies depending on their features. The whole project can be divided into two parts. First, customer values are quantified and six value rankings (Rank A to F) are developed by applying factor analysis method and regression models. Second, researchers use K-means algorithm to find out customer features for each value ranking. Customers are classified into two or three feature groups within each value ranking. Budgets were allocated based on value rankings and tailored strategies were made according to feature groups. The results of this project can assist the airline company in allocating resources and attracting customers.

Keywords: Airline, Customer values, Decision making, Regression, RFM, K-means cluster.

1. Introduction

Nowadays, Chinese airline companies are facing both external and internal challenges. The high-speed railway industry has expanded considerably, which has taken away some market shares that once belonged to the airline industry. Meanwhile, competitions are also growing among the major airlines [3] (Babić et al., 2017). According to the 2019 annual report of China Southern Airline [4], airline companies should adopt customer-oriented service strategies based on customer value. The restructuring of customer rankings has attracted much scholarly attention in recent years and has formed an academic area [5-7] (eg., Kotler, 2017; Verhoef & Lemon, 2013; Beckers, Risselada &Verhoef, 2014). Scholars used different criteria and methods to quantify customer value rankings. Although there have been intensive studies on quantifying customer value and making corresponding strategies, these studies still have some insufficiency. First, irrelevant variables might have been included when building the model. Second, the customer rankings in the previous studies are not specific enough. Third, little attention has been paid to the decision-making process. To bridge these gaps, the present study quantifies customer value, fractionates specific rankings and designs corresponding strategies. Factor analysis techniques are used to integrate correlated variables. Logistic regression is also applied to test whether the factor analysis results are reasonable. Besides, specific customer rankings are classified by quantifying customer value. K-means is utilized to capture customers' characteristics. Furthermore, detailed strategies are made based on the new customer rankings and their features. These strategies can assist the airline company in retaining customer resources.

2. Methodology

2.1 Sample characteristics and variable selection

The experimental subjects in this study were 62988 airline customers aged from 7 to 110. Male constituted 76.42% of the dataset and the rest 23.58% were female. Some observations are invalid because of missing information. In data preprocessing stage, researchers selected 61884 valid observations as the sample. Table 1 shows the details of selected and omitted variables.

Table 1. The details of selected and omitted variables

Variable Name	Selected or Omitted
Eleven important characteristic variables: 1. Membership ranking 2. Number of flights during observation 3. Accumulated membership points 4. Total consumption 5. Total flying distance 6. Period between last flight and end of observation 7. Average time interval between two flights 8. Maximum time interval between two flights 9. Point exchange times 10. Average discount 11. Extra points except flying	Selected for data analysis.
Eighteen redundant variables (e.g. Flying distance in the first year; Flying distance in the second year)	Omitted to avoid perfect collinearity.
Two meaningless variables (Observation time & EP_sum)	Omitted because the sample has zero variation in these two variables.
Thirteen irrelevant variables (e.g. Membership number; First flying date)	Omitted because these variables are irrelevant to customer value and their consumption feature.

2.2 Data analysis and customer value

Factor analysis techniques, logistic regression and Pareto principle were used in data analysis process. First, Kaiser-Meyer-Olkin (KMO) test and Bartlett test will be performed. According to Williams, Onsman and Brown [8] (2010), the KMO test and Bartlett' test should be used to assess if the data is appropriate for factor analysis.

```
Bartlett test of sphericity

Chi-square         =        5.32e+05
Degrees of freedom =              55
p-value            =           0.000
H0: variables are not intercorrelated


Kaiser-Meyer-Olkin Measure of Sampling Adequacy
KMO                =      0.790
```

Figure 1. KMO and Bartlett test

As shown in Figure 1, the data passed Bartlett test (p-value < 0.05) and KMO measure of sampling adequacy equals to 0.79 > 0.7. This indicated that the 11 selected variables were suitable for factor analysis. In the factor analysis procedure, the selected data was analyzed by Stata. This procedure was expected to examine the relationship between variables and compute the factor scores (F) for each customer, which represented quantified customer value.

Next, researchers ran a logistic regression between the original rankings (rank 4, 5 and 6) and the factor scores (F). In this logistic regression, original rankings were the dependent variables and factor scores were the independent variables. This step intended to test whether criteria of factor analysis can efficiently explain the original model.

Multinomial logistic regression		Number of obs	=	61,884
		LR chi2(2)	=	22092.10
		Prob > chi2	=	0.0000
Log likelihood = -9100.1891		Pseudo R2	=	0.5483

ffp_tierme~g	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
4	(base outcome)					
5						
F	5.777929	.0714826	80.83	0.000	5.637826	5.918033
_cons	-4.483214	.0415919	-107.79	0.000	-4.564732	-4.401695
6						
F	6.955264	.085479	81.37	0.000	6.787728	7.1228
_cons	-6.683335	.073831	-90.52	0.000	-6.828041	-6.538629

Figure 2. Results of the logistic regression

Figure 2 displays the pseudo R2 and p-value for this logistic regression. This reveals that criteria in factor analysis can explain the original model. Thus factor scores (F) can accurately represent quantified customer value. Researchers fractionated factor scores (F) into six quantified rankings by Pareto principle. According to Ivančić [9](2014), Pareto principle indicates that 80% of business comes from 20% of customers and this principle can be directly applied in commercial activities. To be specific, top 20% customers of each original ranking were selected as a quantified group and the rest 80% formed another group.

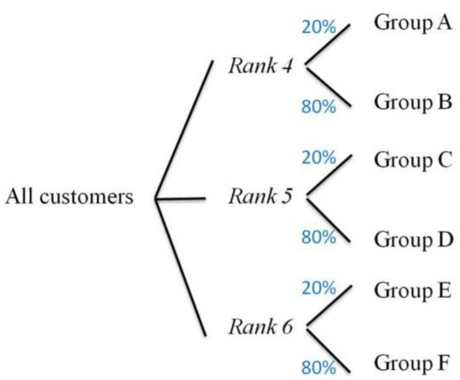

Figure 3. Fractionating customer value rankings by Pareto principle

As Figure 3 shows, all customers were fractionated into six value rankings by Pareto principle (denoted as Group A-F). Customer value decreases from Group A to F. In decision making process, budgets also decrease from Group A to F. To conclude, researchers used factor scores (F) to represent customer value and developed six value rankings.

2.3 Customer features and decision making

In previous data analysis, customers were fractionated into six quantified rankings (Group A-F). Apart from these value rankings, customers still need to be segmented by their features to facilitate strategies making process. Recency-Frequency-Monetary (RFM) model was applied to classify customer features. RFM analysis is a popular customer segmentation and identifiable technique in database marketing. In recent years, many scholars improved traditional RFM model to capture special situation. In the present project, each customer under RFM model is scored based on the following three dimensions:

(1) Recency: It is the period between the last purchase and the last day of observation.

(2) Frequency: It refers to the number of purchases during the observation period.

(3) Monetary: It is the amount of money spent during the observation period.

Under RFM framework, K-means algorithm was applied to classify customers from the same value rankings into several clusters. These clusters were classified by features of customers in the same value ranking. As a result, researchers did not include "membership ranking" as a variable in K-means cluster. The number of clusters was determined by maximizing Silhouette coefficient and Calinski-Harabasz index. The critical point of inertia curve was also found to optimize the number of clusters. In other words, each of the six value rankings was classified into these clusters based on K-means algorithm was performed to generate these clusters.

From previous procedures, customers were classified into six value rankings. Each of them was further classified into several clusters based on their features. Airline companies should allocate more budgets to customers with high value rankings. As illustrated in Sects 4.2, factor scores (F) represent quantified customer value. Different budgets were allocated to different value rankings according to customers' factor scores (F). Budget percentage for each value ranking was calculated as the sum of factor scores after min-max normalization in a certain rank divided by the sum of all customers' factor scores.

Most importantly, strategies were made based on the customer features in different customer value rankings of each cluster. Researchers made differential and tailored strategies to each specific cluster.

3. Results

3.1 Customer value rankings: factor analysis and logistic regression

Factor analysis was applied to find main factors for customer value. Researchers used these main factors to compute factor score (F) for each customer. The following tables and figures show the results of this procedure.

Table 2. Result of factor analysis

Factor	Eigenvalue	Difference	Proportion	Cumulative
Factor1	4.90660	3.10227	0.4461	0.4461
Factor2	1.80433	0.76596	0.1640	0.6101
Factor3	1.03838	0.05394	0.0944	0.7045
Factor4	0.98444	0.35591	0.0895	0.7940
Factor5	0.62853	0.03196	0.0571	0.8511
Factor6	0.59657	0.08975	0.0542	0.9054
Factor7	0.50681	0.27325	0.0461	0.9514
Factor8	0.23356	0.03924	0.0212	0.9727
Factor9	0.19432	0.11941	0.0177	0.9903
Factor10	0.07491	0.04336	0.0068	0.9971
Factor11	0.03155	.	0.0029	1.0000

Table 3. Three selected factors

Variable	Factor1	Factor2	Factor3
ffp_tierme~g	0.13938	0.02502	0.21345
flight_cou~t	0.18412	0.01226	-0.08682
bp_sumaccu~t	0.19132	0.03362	0.09520
sumtotalco~n	0.19362	0.02855	0.06096
seg_km_sum~t	0.18833	0.02730	-0.10310
last_to_en~f	-0.07790	-0.32468	0.15788
avg_interv~b	-0.07355	0.44016	0.07658
max_interv~v	-0.04956	0.49846	0.03048
exchange_c~s	0.12928	0.03564	-0.19545
avg_discou~t	0.05590	0.01821	0.84563
eli_add_po~p	0.08452	0.03916	-0.31862

Table 2 and Table 3 display all factors generated by factor analysis. The first three factors (F1, F2, F3) can cumulatively explain 70.45% of customer value. The other 8 factors are not as significant as F1, F2 and F3. Therefore, researchers compute factor score (F) for each customer by using main factors F1, F2 and F3. According to Patro and Sahu [10] (2015), min-max normalization technique was also applied to facilitate decision making process.

Table 4. Summary of factor scores (F) for all customers

Variable	Obs	Mean	Std. Dev.	Min	Max
F	61,884	1.57e-11	.4845746	-.9602487	11.14838

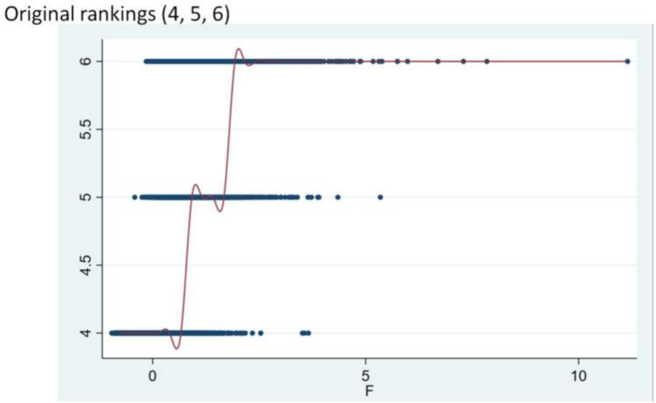

Figure 4. Plot of the logistic regression

Table 4 shows the summary of factor scores (F) for 61884 customers. All factor scores (F) fall in range [-0.9602, 11.1484]. Figure 4 is the plot of logistic regression. These factor scores (F) are consistent with the original model and can accurately represent customer value. Customers are fractionated into six value rankings based on Pareto principle.

Table 5. Customer value rankings

New ranking	Propotion	Number of People	Round-off Number
A	0.004851012	300.2	300
B	0.019404046	1200.8	1201
C	0.010985069	679.8	680
D	0.043940275	2719.2	2719
E	0.18416392	11396.8	11397
F	0.736655678	45587.2	45587

As shown in Table 5, groups A-F are the six new quantified value rankings. Generally, customer value decreases from Rank A to Rank F. Number of customers in different rankings are different. Rank A constitutes 0.48% of the sample size while Rank F constitutes 73.67%. The other four rankings generally contain 1% to 18% customers. Airline companies' budget for each value ranking also differs, which will be discussed in Sects.3.3.

3.2 Customer feature groups: K-means cluster under RFM model

K-means method is applied to each value ranking (Rank A-F), which determines customer features in each value ranking. The number of feature groups in each value ranking is determined by Silhouette coefficient, Calinski-Harabasz index and inertia curve.

Table 6. Result of K-means cluster

Value Ranking	Feature Group	Number of people	Round-off Percentage
A	A1	50	0.08080%
	A2	250	0.40398%
B	B1	499	0.80635%
	B2	700	1.13115%
	B3	2	0.00323%
C	C1	364	0.58820%
	C2	315	0.50902%
	C3	1	0.00162%
D	D1	1481	2.39319%
	D2	1238	2.00052%
E	E1	5590	9.03303%
	E2	5807	9.38369%
F	F1	12730	20.5708%
	F2	32857	53.0945%

As can be seen in Table 6, Rank A, D, E and F are classified into 2 groups based on customer features. Some feature groups are relatively small, like Group A1, which only contains 50 customers. The largest Group F2 has 32857 customers and constitutes 53.0945% of the total sample size. Similarly, Rank B and C are classified into 3 groups. B3 and C3 are two special groups, which have very few customers.

3.3 Decision making: Budget allocation and strategies establishing

This section provides expected budget allocation and tailored strategies. Total budget of the airline company is considered as 100%. The following table shows the result of budget allocation.

Table 7. Budget allocation for different value ranking

Value Ranking	Budget Allocation for Different Value Ranking (%)	Budget per Capita (%)
A	2.13	7.08×10^{-3}
B	5.40	4.49×10^{-3}
C	2.44	3.59×10^{-3}
D	8.04	2.96×10^{-3}
E	23.79	2.09×10^{-3}
F	58.22	1.27×10^{-3}

As can be seen in Table 7, budget allocation for each value ranking differs. Rank F is allocated 58.22% of the total budget while Rank A was only allocated 2.13%. Contrary to total budget, budget per capita increases from Rank F to Rank A. Higher value rankings correspond with higher budget per capita. Rank A enjoys the highest budget per capita, which constitutes 7.08×10-3 % of the total budget. In comparison, budget per capita of Rank A is 5.6 times of Rank F.

Based on value rankings and feature groups, the following tables show the tailored strategies for each group. All decisions are made from two directions: monetary strategies and service improvement.

Table 8. Strategies for Rank A

Value Ranking	A	
Feature Group	A1	A2
Monetary strategies	1. Give discount to point exchange 2. Give extra point per unit of distance	1. Give discount for long-distance flight
Service Improvement	1. Distribute questionnaire 2. Send notification frequently 3. Send airline souvenir 4. Offer additional service	1. Recommend long-distance sightseeing journey 2. Recommend duty-free shops frequently

In Rank A of Table 8, both Group A1 and A2 are high value customers. Group A1 has a relatively higher customer attrition rate than Group A2. To stimulate consumption for Group A1, extra membership point per unit of distance is given. Airline companies should send some questionnaires and notifications to A1 customers for their suggestions. Group A2 has a high-flying frequency but short flying distance. Therefore, discount for long-distance flight is given to A2 customers. Meanwhile, the airline company should also recommend long-distance sightseeing journey to these customers.

Table 9. Strategies for Rank B

Value Ranking	B		
Feature Group	B1	B2	B3
Monetary Strategies	1. Give discount to point exchange 2. Give extra point per unit of distance	1. Give large discount in duty-free shop	1. Give small discount in duty-free shop
Service Improvement	1. Introduce high quality service of the first class 2. Send notification 3. Advertise special service	1. Give large discount in duty-free shop	1. Advertise duty-free items frequently

Group B1 of Table 9 has similar features as Group A1. More discount and membership points are given to lower customer attrition rate. Additionally, the airline company should introduce the high-quality service of the first class to B1 customers because they have potential purchasing ability. Group B2 and B3 have relatively low consumption in duty-free shop. Corresponding discounts are given to these two groups. According to Gao and Chen [11] (2015), large discount is one of the most useful way to stimulate customers' purchasing behavior.

Table 10. Strategies for Rank C

Value Ranking	C		
Feature Group	C1	C2	C3
Monetary Strategies	1. Give discount for the first class 2. Give discount in duty-free shop	1. Give discount for ticket price	1. Give discount for ticket price
Service Improvement	1. Advertise duty-free items 2. Transportation bonus	1. Offer traveling information and suggestions frequently 2. Advertise duty-free items	1. Offer traveling information and suggestions frequently 2. Advertise duty-free items

Similar to Rank B, Rank C also consists of three feature groups. Group C1 seldom purchase tickets of the first class. Some discounts are given for high level classes. Group C2 and C3 Table 10 share similar features. In addition to ticket price discount, some traveling information are offered, which are expected to increase their flying frequency. The airline company should advertise duty-free items to stimulate their consumption in the airport.

Table 11. Strategies for Rank D

Value Ranking	D	
Feature Group	D1	D2
Monetary strategies	1. Give discount for the first class 2. Offer cheaper class-upgrade opportunities	
Service Improvement	1. Transportation coupon	1. Offer traveling information and suggestions 2. Hotel recommendation

Group D1 of Table 11 has potentially large purchasing power. Some discounts for high level classes and cheaper class-upgrade opportunities are offered to them. Besides, they also enjoy transportation coupon like a free taxi drive. Group D2 has a relatively lower flying frequency. Thanks to the budget limit, the airline company should focus on offering traveling information and recommending accommodation.

Table 12. Strategies for Rank E

Value Ranking	E	
Feature Group	E1	E2
Monetary strategies		1. Give discount for the first class
Service Improvement	1. Bonus reminder 2. Introduce membership priority frequently	

Rank E of Table 12 customers have less consumption than Rank A to D. Strategies for Group E1 focus on introducing membership priority and bonus reminder, which motivate them to become higher ranking members. More specifically, bonus reminder can inform customers once the bonus has been accumulated to a certain level. Group E2 has higher purchasing ability than E1. Discount for high level classes are offered.

Table 13. Strategies for Rank F

Value Ranking	F	
Feature Group	F1	F2
Monetary strategies		1. Give small discount in various service
Service Improvement	1. Introduce membership priority	

Rank F of Table 13 constitutes over 70% of total customers but they have the lowest value. Similar to Group E1, membership priority should be introduced clearly to Group F1. Group F2 has relatively higher purchasing ability. Thus, small discounts in various services are offered.

4. Conclusions

This project has quantified customer values, fractionated their rankings and made tailored strategies based on customers' features for an airline company. It has been shown that customers were fractionated into six value rankings (Rank A to F). There are significant differences between different ranks. Within each value ranking, customers were classified into two or three feature groups. Finally, budgets were allocated based on value rankings and tailored strategies were made according to feature groups. This study is important in at least three aspects. First, it specifies the value rankings by using factor analysis techniques. This detailed classification can benefit the decision-making process. Second, this project provides specific strategies for the airline company. Since decision making has been paid little attention in the previous studies, the present project can fill this gap by providing these tailored strategies. Third, this study outlines a framework of quantifying customer values. This framework can be sustainably applied to future customer relationship management. However, this study still has some insufficiency and limitations. First, macroeconomic factors, such as government policy and economic fluctuation are not considered in this project. They can also influence customer values and features considerably. Second, Pareto principle is used when fractionating customer value rankings. This principle is an empirical model rather than a quantitative model. As a result, there might be some inaccuracy of customer value rankings. Therefore, future studies should take external factors into consideration and use more quantitative methods to fractionate customer value rankings.

References

[1] Huang, W., Chen, S., & Liu, Z. (2016, May). Big data applications in business analysis. In 27th Annual Conference on Production and Operations Management Society (POMS), Orlando, the United States.

[2] Liu, J., & Du, H. (2010, June). Study on airline customer value evaluation based on RFM model. Papers presented at 2010 International Conference on Computer Design and Applications, Qinhuangdao, China.

[3] Babić, R. Š., Tatalović, M., & Bajić, J. (2017). Air transport competition challenges. International Journal for Traffic & Transport Engineering, 7(2), 144-163.

[4] China Southern Airline. (2019). The annual financial of China Southern Airline. Driver, J. (1999). Development in airline marketing practice. Journal of marketing practice: Applied marketing science, 5(5), 134-150.

[5] Kotler, P. (2017). Customer value management. Journal of creating value, 3(2), 170-172

[6] Verhoef, P. C., & Lemon, K. N. (2013). Successful customer value management: Key lessons and emerging trends. European Management Journal, 31(1), 1-15.

[7] Beckers, S. F., Risselada, H., & Verhoef, P. C. (2014). Customer engagement: A new frontier in customer value management. In Handbook of service marketing research. Edward Elgar Publishing.

[8] Williams, B., Onsman, A., & Brown, T. (2010). Exploratory factor analysis: A five-step guide for novices. Australasian journal of paramedicine, 8(3).

[9] Ivančić, V. (2014). Improving the decision-making process trought the Pareto principle application. Ekonomska Misao i Praksa, 23, 633-656.

[10] Patro, S., & Sahu, K. (2015). Normalization: A preprocessing stage. arXiv preprint arXiv:1503.06462.

[11] Gao, F., & Chen, J. (2015). The role of discount vouchers in market with customer valuation uncertainty. Production and Operations Management, 24(4), 665-679.