Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2020) 28: 2110 – 2127

doi:10.3906/elk-1911-15

Passenger scoring for free-pass promotion in public transportation

Ufuk DEMİR ALAN[1], **Derya BİRANT**[2,*]
[1]The Graduate School of Natural and Applied Sciences, Dokuz Eylül University, İzmir, Turkey
[2]Department of Computer Engineering, Dokuz Eylül University, İzmir, Turkey

Received: 04.11.2019 • **Accepted/Published Online:** 06.03.2020 • **Final Version:** 29.07.2020

Abstract: The focus of promotions targeted to increase the use of public transportation concentrates on increasing the attractiveness of it, particularly by decreasing transportation fares. To serve that purpose, this paper proposes a novel passenger scoring model, namely RFLT (recency, frequency, loyalty, and time), for offering a free-pass promotion in public transportation. It presents the comparison results of RFLT and wRFLT (weighted version) using a real-world dataset obtained by a near field communication (NFC) mobile payment application. The experimental results show that the w-RFLT model provides a more balanced score distribution than the RFLT model, and the frequency parameter (F), among four metrics (R, F, L, and T), influences the scoring results the most. The results of this study can be used to establish an efficient policy for increasing public transportation ridership.

Key words: Information systems, scoring, public transportation, RFM model, mobile application

1. Introduction

Public transportation (PT) plays an important role in everybody's life, especially in terms of their economic and social life quality. Observation of developed provinces shows clearly how high the importance given to PT is. However, even though PT has gained importance lately, car dependence in a local area has increased steadily over the last decade. This causes a series of problems in terms of human health and the environment [1]. In order to overcome car dependency and encourage the use of PT, some promotions may be planned and offered for the use of everybody. This paper gives an idea about the type of activities to be conducted in order to encourage people to use PT.

The most significant factors affecting the intention to increase the use of PT are personal norm, service quality, and economic factors [2]. All these factors have a positive effect on encouraging the use of PT. The present paper focuses on the third factor (economic) through the offer of free-pass promotion. Socioeconomic factors (i.e. sex, hometown, and education level) do not have a direct influence on the intention to increase the use of PT but have an indirect effect through vehicle ownership [2].

The aim of the present study was to develop a sustainable policy for promoting PT. The "free riding right", which has a great motivating effect on the use of PT [3], has been the subject of study for this purpose. The rewarding of active PT riders is referenced as an incentive method. The active passengers to be given free passes can be determined by a scoring method.

A novel passenger scoring method, namely RFLT (recency, frequency, loyalty, and time) is proposed. The well-known marketing model RFM (recency, frequency, and monetary) [4] was extended by adding two other

*Correspondence: derya@cs.deu.edu.tr

parameters, loyalty (L) and time (T), to form the RFLT model. While loyalty defines the time period between the user's first entry to the system and today, time measures roughly the times of daily public transport usage.

The novelty and main contributions of this paper are three-fold. First, it proposes a novel passenger scoring mechanism (RFLT). Thus, it introduces a new version of the RFM model that differs from previous studies. Second, it demonstrates the applicability of four metrics (recency, frequency, loyalty, and time) on PT for the first time. Third, the proposed approaches (RFLT and its weighted version wRFLT) are compared using a real-world dataset for sensitivity analysis. To the best of our knowledge it is the first time that sensitivity analysis has been applied to determine how variations in these parameters change passenger scores.

In the experimental studies, the proposed scoring technique RFLT was applied to a real-world dataset obtained by a near field communication (NFC) mobile payment application. The RFLT and w-RFLT (weighted version) techniques were compared with each other to investigate the importance of the parameters. The experimental results show that the proposed scoring method is practical and can be used efficiently for promoting PT.

This paper is divided into five sections and organized as follows. Section 2 presents related work on the topic and describes the benefits of free-pass promotion in PT. Section 3 presents basic concepts, notations, novel definitions, and the proposed method. Details on the experiments are presented in Section 4, including a description of the dataset, the experiment design, and the comparative results. Finally, Section 5 concludes the paper with a summary and discusses possible future works.

2. Related work

2.1. Advantages of free-pass promotion in public transportation

Applying small discounts to riding prices has an impact on the use of PT, but a free pass is a very strong tool to change the driving habits of citizens. The study by Thogersen and Moller [5] confirms that a free 1-month transportation card effectively releases the habits of car drivers and increases their use of PT. Galey [6] also presents the benefits of free passes in relation to the city of Tallinn, capital of Estonia. He mentioned that, with free-pass rights, preliminary results indicated an increase in passenger demand of 3% citywide in the 3-month period after implementation. Notably, however, passenger counts increased 10% in a district of the city. Belter et al. [7] explain the advantages of free PT services from four different points of view: aspects for public transport users, for the community, for the local authority, and for the public transport operator. From the point of public transport users, they mentioned that public transport becomes more attractive as it can be used free of charge. The study by Fearnley [8] also shows that a free pass in public transport is necessary because of its positive social, economic, health, and environmental impacts. Decreasing the number of cars on roads reduces congestion, pollution, and parking challenges. The study by Gase et al. [9] also presents the benefits of providing free-pass rights to students in Los Angeles county. According to their research, free transit passes for students enables them to increase school attendance, have more freedom and mobility, strengthen social networks, decrease contact with the justice system, and increase disposable income, while reducing traffic volume and allowing them to participate in sports and cultural events.

To encourage the use of PT and, in this way, to provide the benefits aforementioned, a novel scoring mechanism is proposed here to determine the free-pass rights of the inhabitants of a province in a fair way. The proposed scoring method can be implemented by the public transport operators and can be effectively used by the municipalities as a campaign tool within the scope of free rides.

2.2. RFM and its versions

RFM is a scoring method for analyzing customer behavior using the three following variables: *recency* (R) - time elapsed since the last purchase, *frequency* (F) - the number of purchases, and *monetary* (M) - the value of the purchases. The main purpose of this method is to find customer segments and identify target customers for marketing campaigns and events. The RFM model was first proposed by Hughes [4] in 1994. This model has proven very effective [10] when applied to marketing databases.

RFM analysis is a marketing technique generally used for customer selection, cross-selling, and customer churn prediction. Moreover, several studies based on the RFM model have also been proposed for various areas such as manufacturing [11], computer security [12], government [13], tourism [14], social media [15], the automobile industry [16], education [17], banking [18], health care [19], dental care [20], the hairdressing industry [21], telecommunications [22], and the electronics industry [23]. The importance of our work comes from being the first study in which the RFM concept has been applied in the field of PT.

Table 1 presents different versions of the RFM model [16, 17, 20, 22, 24–35] that have been proposed by changing or adding some parameters. The reasons for these modifications are to adapt it to different fields or to achieve better results than the basic RFM model.

Table 1. Different versions of the RFM model.

Author	Year	Method	Description	Area
Babaiyan and Sarfarazi [22]	2019	LRFM	Length, recency, frequency, and monetary	Telecomm.
Ravasan and Mansouri [16]	2018	WRFM	Weighted RFM	Automobile
Huang et al. [24]	2018	RFMS	Recency, frequency, monetary, and standard deviation	Banking
Peker et al. [25]	2017	LRFMP	Length, recency, frequency, monetary, and periodicity	Marketing
Zhang et al. [26]	2015	RFMC	Recency, frequency, monetary, and clumpiness	Marketing
Chen et al. [27]	2015	RFL	Recency, frequency, and length	Television
Cho et al. [28]	2015	FRAT	Frequency, regency, amount, and type of service	Marketing
Liu and Chen [29]	2014	RFAT	Recency, frequency, average monetary, and trend	Marketing
Al-Shayea and Al-Shayea [30]	2014	RFMT	Recency, frequency, monetary, and time	Marketing
Golsefid et al. [31]	2012	RFMCT	Recency, frequency, and Monetary, continuity, trend	Marketing
Wei et al. [20]	2012	LRFM	Length, recency, frequency, and monetary	Dental care
Chang and Tsai [32]	2011	GRFM	Group RFM	Marketing
Bizhani and Tarokh [33]	2011	RF*M*	Recency, new definition of frequency and monetary	Banking
Yan and Chen [34]	2011	RFD	Recency, frequency, and duration	Recommend
Chang [17]	2010	EL-RFM	RFM model for E-learning	Education
Yeh et al. [35]	2009	RFMTC	Recency, frequency, monetary, the first purchase (T), and churn probability (C)	Marketing

In the transportation industry, the RFM and RFM-based models have also been used to determine the values of passengers [36–44] (Table 2). The common characteristic of these studies is that all of them consider the "frequency" parameter. This situation indicates the soundness of the effect of frequency on encouragement in transportation. In some articles [36, 38, 40], simple RFM models have been utilized for passenger segmentation. However, determining passenger value in each type of transportation study has its own characteristics, which might not be fully satisfied by the same model. Hence, different RFM-based models were generally constructed in order to better reflect the behavior of passengers. For example, in the airline industry, the measures of the LDcFR model [39] include length, distance, frequency, and recency, whereas the FSLC model [41] focuses on frequency, season, locations of traveling, and cancellation times; the FMCN model [43] contains frequency, monetary, cancellation times, and the number of family members measures. Our proposed model (RFLT) differs from these previous studies in two respects. First, unlike previous studies that focus on airline, highway, and intercity transfers, our model is proposed for PT (i.e. urban bus and metro transfers). Second, our model considers different parameters (loyalty and time) to meet the needs of local public transport, because some parameters used in the previous studies such as monetary, cancellation times, and the number of family members are not suitable for local public transport. Carmona [45] presents a conceptual model for the implementation of relationship marketing in urban public transport operations, which points to the promotion of loyalty approach for building competitive advantages.

Table 2. Comparison of our model (RFLT) with other RFM-based models in transportation.

Ref.	Year	Method	R	F	M	Other parameters	Type of transportation
[36]	2019	RFM	√	√	√	-	High-speed train
[37]	2019	FMA		√	√	A: Average number of group travelers	Airline industry
[38]	2018	RFM	√	√	√	-	Electronic toll collection in highway transportation
[39]	2018	LDcFR	√	√		L: Length Dc: Distance	Airline industry
[40]	2017	CV-RFM	√	√	√	CV: Customer values	Weekend city traveling
[41]	2017	FSLC		√		S: Season L: Locations of traveling C: Cancellation times	Airline industry
[42]	2017	CFMY		√	√	C: Recency/avg. ticket buying time Y: Total days of tickets booked/ ♯ tickets purchased	Highway passenger transport
[43]	2014	FMCN		√	√	C: Cancellation times N: Number of family members	Airline industry
[44]	2012	FPDN		√		P: Price discount D: Destination N: No-show	Airline industry
RFLT (proposed in this paper)			√	√		L: Loyalty T: Time (peak or off-peak hours)	Public transportation

3. RFLT model

RFLT, which is a modified version of the RFM method, is a novel scoring technique proposed in the present study for the PT field. The RFLT model brings a framework to the table for objectively measuring the following four ideas on a numerical scale:

- Recency - How recently did the customers use PT?
- Frequency - How often do they use it?
- Loyalty - How long do they use it?
- Time - What time do they generally use it?

To demonstrate RFLT scoring, an example dataset about passenger transactions is given in Table 3. The first column, public transportation card number (PTCN), is a unique number to identify passengers. The other columns are recency, frequency, loyalty, and time values.

Recency is the number of days since the passenger's last boarding. It measures the interval between the most recent transaction date and the analyzing date. For example, if the last transaction date of the passenger is 16.12.2016 and the analyzing date is 30.12.2016, the recency value will be 14 days. For instance, passenger *P2* given in Table 3 has a recency value of 1, which means that passenger *P2* used the PT yesterday, if the baseline is today. Recency is measured in days, that is, the lower the number of days, the higher the score of recency. According to RFLT, the passenger who has recently used PT is rewarded with a higher score, as well as more free-pass rights. Hence, citizens may prefer to use PT more than they do presently to keep short the interval between his/her last boarding date and the analyzing date. In addition, it is a valuable measure to identify passengers at risk of churn (when recency is low). To calculate the recency parameter, it is necessary to find the last riding date value of the related passenger. Recency is calculated by the last boarding date subtracted from the analyzing date, as given in Equation (1).

Table 3. A sample dataset about passenger transactions.

PTCN	Recency (days)	Frequency (number)	Loyalty (days)	Time (number)
P1	6	720	1000	$288/720 = 0.40$
P2	1	840	620	$720/840 = 0.86$
P3	24	512	90	$412/512 = 0.80$
P4	2	53	27	$10/53 = 0.19$
P5	5	464	58	$455/464 = 0.98$
P6	7	57	222	$30/57 = 0.53$
P7	35	75	150	$30/75 = 0.40$
P8	3	892	45	$695/892 = 0.78$
P9	150	834	450	$576/834 = 0.69$
P10	10	576	556	$46/576 = 0.08$
P11	14	345	360	$87/345 = 0.25$
P12	34	30	185	$20/30 = 0.67$
P13	87	854	859	$700/854 = 0.82$
P14	8	90	156	$54/90 = 0.60$
P15	38	5	40	$1/5 = 0.20$

Definition 1. (*Recency*). Recency is the number of elapsed days between the last date of the period being analyzed and the last date the PT was boarded. The recency value (R) for the passenger P_i in days is calculated as

$$R_{(P_i)} = ED - \max_{SD \leq t \leq ED} TD_{(P_i)}^{(t)}, \tag{1}$$

where SD and ED are the start and end dates of the analyzing period t, respectively, $TD_{(P_i)}^{(t)}$ includes the transaction dates recorded in the period t for the i^{th} passenger, $i = 1, 2, ..., m$, and m is the number of passengers.

Frequency is a magic parameter to identify active and inactive passengers. It is the number of rides performed by the related passenger in a given time period. The frequency value is increased each time a passenger boards a public transport vehicle even though the boarding may be the result of a transfer from another route to complete the same one-way journey (i.e. a transfer between vehicles is counted as two unlinked trips). For instance, passenger *P2* given in Table 3 has a frequency value of 840, which means that passenger *P2* used PT 840 times during the period between January 2014 and September 2016. Frequency is measured in terms of ride counts. Some passengers might transit several times during the same day. However, the number of boardings is counted independently from the day, and so each boarding of the public transport vehicle is considered a separate trip even through occurring in the same day. Over time, frequency can be used to predict the future likelihood. According to RFLT, the passenger who has frequently used PT should be rewarded with more scores and so more free-pass rights as well.

Definition 2. (*Frequency*). Frequency is the number of boarding transactions made by a passenger during the analysis period. The frequency value (F) for the i^{th} passenger is calculated as

$$F_{(P_i)} = \sum_{t=SD}^{ED} \left| B_{(P_i)}^{(t)} \right|, \tag{2}$$

where SD and ED are the start and end dates of the analyzing period t, respectively, and $|B_{(P_i)}^{(t)}|$ is the number of boarding transactions recorded in the period t related to the passenger P_i.

Loyalty is the time period of membership. Loyalty is measured in terms of days, and is calculated by the first boarding date subtracted from the analyzing date. For instance, passenger *P2* given in Table 3 has a loyalty value of 620, which means that passenger *P2* has used PT for the first time 620 days before the analyzing date. The higher the number of days means the higher the score of loyalty. In recent years, loyalty has gained importance in PT [46, 47]. Loyalty is important since the cost of getting new passengers is noticeably higher than the cost of retaining the existing ones when we target increasing the use of PT. Furthermore, it assures continuity of the passenger life cycle. Loyalty is important, since it focuses on the value of the passenger in terms of PT ridership. Increased loyalty prevents citizens from transferring from PT to private transportation. It can be an easy and good strategy to encourage former passengers, mainly because of old habits. Hence, loyalty is a key metric for many organizations and an inherently supported capability within most intelligent software systems. Generally, organizations tend to reward former members. From this point of view, former users are more valuable and deserve to be rewarded in the RFLT method.

Definition 3. (*Loyalty*). Loyalty is the number of elapsed days to date from the first boarding transaction date of the passenger. The loyalty value (L) for the i^{th} passenger P_i is calculated as

$$L_{(P_i)} = ED - \min_{SD \le t \le ED} TD_{(P_i)}^{(t)}, \tag{3}$$

where SD and ED are the start and end dates of the analyzing period t, respectively, and $TD_{(P_i)}^{(t)}$ includes the transaction dates recorded in the period t for the passenger.

The *time* parameter in the RFLT model is the average off-peak hour travel times. In other words, it is the ratio of the number of trips the passenger took on the transportation service during the off-peak period to the total ride counts (frequency) of the passenger. For instance, passenger *P2* given in Table 3 has a time value of 0.86, because he used PT 720 times at the off-peak period and 840 times in total (frequency value of *P2*) and so the ratio is $720/840 \sim 0.86$. Peak hours can change from province to province but generally there are two peak times in a day: from 7 am to 9 am and from 4 pm to 6 pm. The time parameter is important in order to provide a higher service quality, especially availability, in PT by encouraging riders to use it outside of peak hours [48]. Traveling at an out-of-peak hour is more valuable to reduce passenger volume (i.e. overcrowded boarding) at peak hours. For this reason, in the RFLT model, the passengers who use PT at off-peak hours are rewarded with higher scores. Thus, free-pass opportunities will encourage passengers who can choose their departure time to make the trip at off-peak hours. In this way, the PT system can reduce the burden by peak spreading [8].

Definition 4. (*Time*). The time parameter in the RFLT model is the ratio of off-peak hour ride counts to total ride counts of the passenger. The time value (T) for the passenger P_i is calculated as

$$T_{(P_i)} = \left|E^{(t)}_{(P_i)}\right| / F_{(P_i)}, \tag{4}$$

where $\left|E^{(t)}_{(P_i)}\right|$ is the number of transactions recorded at out-of-peak hours in the analyzing period t for the i^{th} passenger, $i = 1, 2, ..., m$, and m is the number of passengers.

The monetary parameter in the original version of the RFM is discarded because in urban PT ticket prices are fixed; thus cost information is not an accurate parameter for segmenting passengers. The monetary value of a passenger is directly related to frequency, which already exists in the model.

In principle, RFM analysis can be performed by following one of two common scaling and scoring methods: (i) quintiles (equal-frequency binning) and (ii) simple fixed ranges (threshold-based binning) [49]. In the quintiles method, the records are sorted and divided into five equal groups, and then the top 20% values are given a score of 5, the next 20% values are given a score of 4, and so forth. In other words, grouping is done based on the number of passengers but passengers with equal values should be in the same group. In the second method (simple fixed ranges), the records are also sorted and then scored based on the range of numerical values. For instance, recency can be divided into five intervals according to the following ranges: 0–2 months, 3–6 months, 7–11 months, 12–24 months, and 25+ months, which are scored as 5, 4, 3, 2, and 1, respectively. Different businesses may use different methods for ranking the RFM values on the scale of 1 to 5. Each approach has some advantages and some drawbacks compared to the other. The quantiles method yields equal numbers of customers in each segment, but it gives little insight about relationships. In the fixed ranges method, it is difficult to decide on ideal boundaries and update them periodically as the data grow. Several segments can be very small and so it may not be worth targeting the segment if it contains a very small set of customers [49].

In the present study, we chose the quintiles method for the RFLT model for the public transport industry because of four reasons. First, dividing passengers into equal segments is an ideal way for giving free-pass rights to passengers. In the range-based method, the bins vary significantly in the number of passengers. Giving too many free-pass rights is not acceptable for the municipalities or transport authorities because this type of promotion leads to the loss of thousands of dollars each day. Second, there are challenges with fixed range calculation for RFLT scores since the users should decide what range they consider ideal for recency, frequency, loyalty, and time values. However, the quintiles method is an easy way to manage since ranges are picked from the data themselves. Third, the quintiles method is more appropriate if segmentation schemes are generated periodically (i.e. free-pass rights of the passengers may be calculated at the end of each month). However, in the range-based method, the boundaries may need frequent adjustment when the data grow. Fourth, the quintiles method is easy to understand for public transport passengers. Complex and dynamic range calculations are not easily explained to passengers, especially to old and low-educated passengers.

Table 4 shows the main steps of the RFLT model involving scaling of passengers based on each R, F, L, and T factor separately. The process starts with sorting passengers in ascending order based on recency, i.e. period since last boarding. The passengers are then divided into five equal parts and the top 20% given a recency score of 5, the next 20% a score of 4, and so on. Passengers are then sorted by frequency in decreasing order and then scored similarly. This process is then undertaken for the loyalty and time columns as well. Lastly, all passengers are ranked by concatenating R, F, L, and M values and scored by the average of them.

Table 4. The steps of the RFLT analysis.

PTCN	Recency	R	PTCN	Frequency	F	PTCN	Loyalty	L	PTCN	Time	T	PTCN	RFLT	Score
P2	1	5	P8	892	5	P1	1000	5	P5	0.98	5	P2	5555	5
P4	2	5	P13	854	5	P13	859	5	P2	0.86	5	P13	1555	4
P8	3	5	P2	840	5	P2	620	5	P13	0.82	5	P8	5514	4
P5	5	4	P9	834	4	P10	556	4	P3	0.80	4	P1	4452	4
P1	6	4	P1	720	4	P9	450	4	P8	0.78	4	P5	4325	4
P6	7	4	P10	576	4	P11	360	4	P9	0.69	4	P10	3441	4
P14	8	3	P3	512	3	P6	222	3	P12	0.67	3	P9	1444	3
P10	10	3	P5	464	3	P12	185	3	P14	0.60	3	P11	3342	3
P11	14	3	P11	345	3	P14	156	3	P6	0.53	3	P6	4233	3
P3	24	2	P14	90	2	P7	150	2	P1	0.40	2	P14	3233	3
P12	34	2	P7	75	2	P3	90	2	P7	0.40	2	P3	2324	3
P7	35	2	P6	57	2	P5	58	2	P11	0.25	2	P12	2133	2
P15	38	1	P4	53	1	P8	45	1	P15	0.20	1	P4	5111	2
P13	87	1	P12	30	1	P15	40	1	P4	0.19	1	P7	2222	2
P9	150	1	P15	5	1	P4	27	1	P10	0.08	1	P15	1111	1

There is an inverse relationship between recency and passenger score. This means that the recent usage has a positive effect on the method; thus, if recency is small, the score of the passenger is increased. For this reason, the recency column is sorted in ascending order, while the others are sorted in descending order. Frequent, former, and off-peak hour travelers are more valuable than the infrequent, newer, and peak hour travelers.

The RFLT analysis assigns a quadruple value to each passenger on the basis of his/her past behavior. To deal with 625 different score combinations, it is possible to homogeneously group RFLT scores into segments that have similar characteristics, such as best, good, valuable, ideal, medium-level, uncertain, new, nonvaluable, and churned passengers. As shown in Table 5, the *best passengers* are in quintile 5 for each factor (5555) who have used PT most recently, most frequently in the off-peak period. While passengers with score 5111 are *new passengers*, 1111's are *nonvaluable*. The passengers that have the RFLT pattern x51x are *valuable* ones because they have used PT frequently in a short period. The passengers with score xxx5 can be evaluated as *ideal passengers*, because they generally use PT at out-of-peak hours. Transportation operators or governments can use RFLT values in developing different strategies for different passenger segments. For example, they can identify *churned passengers* who have 1555 value and may contact them to get them to return.

Table 5. The passenger segments according to RFLT scores.

Segment	Example RFLT patterns	Example passengers
Best passengers	5555, 5554, 5545, 5455, 4555	P2
Good passengers	4444, 4445, 4454, 4544, 5444	P1
Valuable passengers	5514, 4514, 4515, 5415, 5414	P8
Ideal passengers	4325, 4425, 3425, 4235, 3345	P5
Medium-level passengers	3333, 3334, 3324, 3343, 3433	P11
Uncertain passengers	2222, 2223, 2232, 2322, 3222	P7
New passengers	5111, 5112, 5113, 5114, 5115	P4
Nonvaluable passengers	1111, 1112, 1211, 2111, 1122	P15
Churned passengers	1555, 1554, 1553, 1455, 1454	P13

3.1. Passenger scoring

Scoring in PT must be performed in order to determine how many free-pass rights will be given as a promotion to passengers in a particular month. This study proposes the new scoring methods: RFLT and wRFLT (weighted version of RFLT). In the RFLT method, each parameter has the same effect on the result, $W_R = W_F = W_L = W_T = 1$; while in the wRFLT model, each parameter has its own weight equal to or higher than 1. In these methods, the score of the passenger P_i is computed by Equation (5).

$$wRFLT_{(P_i)} = \frac{W_R * R_{(P_i)} + W_F * F_{(P_i)} + W_L * L_{(P_i)} + W_T * T_{(P_i)}}{W_R + W_F + W_L + W_T} + \alpha + \beta, \tag{5}$$

where $R_{(P_i)}$, $F_{(P_i)}$, $L_{(P_i)}$, and $T_{(P_i)}$ represent the recency, frequency, loyalty, and time values of passenger P_i and W_R,W_F, W_L, and W_T represent the weights of each variable, respectively, according to their importance. The average RFLT score can vary between 1 and 5. The parameter α is used to express "special day effect". In PT, municipalities may want to give extra rides to citizens on special days such as birthdays, the first day of school, international women's day, and teachers' day. Hence, citizens may prefer to use PT rather than private transportation, since it is more economical on these special days. This parameter can be tailored according to the decisions of the municipality where scoring is done.

A free-pass promotion may be given to infrequent users to encourage them to use PT. Frequent passengers do not need much encouragement as they are already active users; on the other hand, infrequent passengers can be expected to travel more if they are encouraged with free PT tickets. Likewise, PT agencies or institutions trying to encourage irregular car users to shift to PT can try to identify nonrecent and nonloyal passengers and offer them free PT tickets. In our model, these motivations are provided by the last parameter (β), which is used to encourage nonrecent, infrequent, nonloyal, and peak-hour (rush-hour) passengers. This parameter has an inverse relationship with the R, F, L, and T values such that $\beta = x/(R_{(P_i)} + F_{(P_i)} + L_{(P_i)} + T_{(P_i)})$, where x is an integer determined by the user; thus, lower R, F, L, and T values will lead to an increase in free-pass promotions. Table 6 lists the meanings of symbols used.

The coding of recency, frequency, loyalty, and time values is arbitrary. Quintiles (5-recency, 5-frequency, 5-loyalty, and 5-time value divisions) are frequently used and hence 625 different score combinations are assumed.

However, as given in Table 5, RFLT codes can be grouped into several segments, similar to the study by Olson and Chae [50]. In addition, depending on the application, RFLT coding can be changed with the parameter k, i.e. quartile (256 cells) or thirds (81 cells), in the case of an insufficient number of passengers. Which records will be included in each segment is depicted by Equation (6).

$$D_t = \bigcup_{i=0}^{k-1} (|D_t|)/(k*(i)) - |D_t|/(k*(i+1)), \tag{6}$$

where $D_t = \{d_1, d_2, ..., d_n\}$ is the dataset that consists of transactions recorded in the period t and k is the number of partitions. The top and bottom limits of each segment, recording number, and order are determined with this equation, e.g., for the dataset including 40 records, where $k = 4$, the resulting groups will be as follows: [0–10], [10–20], [20–30], and [30–40]. The value of k will be defined by the organization, i.e. local government or PT operator.

Table 6. Notation table.

Symbol	Description
t	Time period
D_t	Dataset (transactions recorded in the period t)
$\|D_t\|$	The number of transactions in D_t
k	The number of partitions
m	The number of passengers
a	The number of attributes, i.e. $a = 4$ for RFLT (recency, frequency, loyalty, time)
P_i	i^{th} passenger, where $i = 1, 2, ..., m$
$R_{(P_i)}$	The recency value for the passenger P_i in days, where $R_{(P_i)} = ED - \max_{SD \leq t \leq ED} TD^{(t)}_{(P_i)}$
$F_{(P_i)}$	The frequency value for the passenger P_i, where $F_{(P_i)} = \sum_{t=SD}^{ED} \left\| B^{(t)}_{(P_i)} \right\|$
$L_{(P_i)}$	The number of days up to now from the first ride of P_i, where $L_{(P_i)} = ED - \min_{SD \leq t \leq ED} TD^{(t)}_{(P_i)}$
$T_{(P_i)}$	The ratio of off-peak hour usage to total usage, where $T_{(P_i)} = \left\| E^{(t)}_{(P_i)} \right\| / F_{(P_i)}$
B	The boarding transactions recorded in a day, where $B \subseteq D_t$
E	The transactions recorded at out-of-peak hours, where $E \subseteq D_t$
W_R	Weight for recency R
W_F	Weight for frequency F
W_L	Weight for loyalty L
W_T	Weight for time T
α	Special day effect such as birthday and international women's day
β	Parameter to encourage inactive (nonrecent, infrequent, nonloyal) and peak-hour passengers

In the RFLT model, the free-pass right of a passenger is calculated as the weighted average of R, F, L, and T values of the passenger as given in Equation (5), and so 625 cells are reduced into five scores only, between 1 and 5. For instance, while passenger P2 with 5555 RFLT value in Table 4 is scored as 5 on average, passengers P13, P8, P1, P5, and P10 are scored as 4, and so forth.

3.2. Algorithm description

The RFLT algorithm is applied by defining the scales of R, F, L, and T attributes. This process consists of four steps as follows:
Step 1: Sort the data according to R attribute.
Step 2: Partition the dataset into k equal parts and the records in each part are assigned as *k, k-1, ... ,1* score referring to passenger contribution to the use of public transportation.
Step 3: Repeat the previous two steps for other attributes F, L, and T individually.
Step 4: Calculate free-pass rights by dividing the sum of multiplied R, F, L, and T values with their weights by total weight value.

The time complexity of the RFLT algorithm is composed of sorting, scaling, and merge operations as follows: $O(nlogn * a + n * a + n)$, where n is the number of transactions in the dataset and a is the number of attributes.

4. Experimental results

In the present study, the proposed model RFLT and its weighted version wRFLT were applied to the real-world dataset to demonstrate their applicability in PT. The number of free-pass rights to be given to passengers for the following month in order to encourage the use of PT was calculated by the RFLT scoring technique.

4.1. Dataset description

In the present study, real-world data obtained through the use of NFC-enabled mobile phones in PT were used. These particular data were collected between January 2014 and September 2016. The dataset is composed of 35,230 transactions belonging to 520 different PT cards. Therefore, the values of the input parameters are $t = 33$ months, $k = 5$, $a = 4$, $m = 520$, and $|D_t| = 35,230$.

The dataset covers five attributes drawn from different tables in the database by queries. They are PTCN, last boarding, membership, total usage count, and total off-peak hour usage count. Table 7 shows the attribute names, their meanings, and min./max. values of the dataset. Peak hours can change from province to province but generally there are two peak hour times in a day: from 7 am to 9 am and from 4 pm to 6 pm.

Table 7. The attributes of the dataset with minimum and maximum values.

Attribute name	Description	Min. value	Max. value
PTCN	Public transport card number is a unique virtual card number to identify passenger.	16 digit numeric value.	
Last boarding	Date and time value of the last transaction before analyzing time (dd.mm.yyyy hh:mm).	20.10.2014 06:35 pm	09.09.2016 11:34 am
Membership	Date and time value of the first transaction (dd.mm.yyyy hh:mm).	20.01.2014 06:32 pm	06.09.2016 09:25 am
Total usage count	Total transaction count of the user (in the period t).	1	1473
Total off-peak hour usage count	Total transaction count of the user in off-peak hours (in the period t).	0	1095

4.2. Comparison of RFLT and wRFLT

The RFLT and w-RFLT techniques were compared with each other to investigate the importance of parameters. First, using the attributes shown in Table 7, R, F, L, and T values were calculated with the proposed algorithm. Afterwards, using the formula given in Equation 5, RFLT score values and consequently the "free-pass right" numbers were calculated. The number of free-pass rights is minimum 1 and maximum 5. The extra free-pass right (α) to be given on special days and the promotion strategy (β) to be given to encourage inactive passengers were considered 0 in the present study. However, they can be easily added to the scoring results upon demand since the model supports them.

Weight values are determined as 1 ($W_R = W_F = W_L = W_T = 1$) initially, and afterwards passenger scores are calculated with different weight values. The weight values in the present study were calculated according to priority values of the variables following discussions with the municipality. For example, R4F3L2T1 means that RFLT index weights are $W_R = 4$, $W_F = 3$, $W_L = 2$, and $W_T = 1$.

Table 8 and Figure 1 compare RFLT and wRFLT techniques and show how many passengers have won how many free-pass rights, varying from 1 to 5. In all cases, the passengers were awarded 3 free rides at most, while they were generally assigned 1 free ride at least. According to the RFLT results, 108 people won 2 free rides, 254 people 3 free rides, 145 people 4 free rides, and only 13 people 5 free rides. Score distribution in the RFLT technique is observed to be imbalanced. There is nobody winning just 1 free-pass right while almost 50% of the users win 3 riding rights. It is possible to make score distributions more balanced by changing the weight values in the wRFLT technique. It is observed that the most balanced situation arises when the frequency parameter is emphasized more, i.e. with R2F4L1T1 weight values. The R4F3L2T1 model constructed a form quite similar to the bell curve. Actually, it is an expected situation for a good scoring process. However, the skewness and kurtosis of the curve can be changed by using different weight parameters. As a result, use of the weighted RFLT technique is more suitable for this dataset. In the experiments, the weight values were limited to the range [1, 4] in order to avoid strong effects of the parameters. However, the wider ranges can also be determined to give stronger effects on some parameters of wRFLT.

Table 8. Comparison of the RFLT and wRFLT techniques.

		Number of passengers				
	Score	1	2	3	4	5
	RFLT	0	108	254	145	13
wRFLT	R2F3L1T3	18	113	228	152	9
wRFLT	R4F3L2T1	2	155	205	137	21
wRFLT	R1F1L4T1	10	171	186	135	18
wRFLT	R2F4L1T1	9	159	172	150	30

4.3. Sensitivity analysis of the RFLT model

Sensitivity analysis of the RFLT model was performed to find out how sensitive passenger scores are to any change in R, F, L, and T weights. This study is applied to determine how different weight values impact free-pass promotion in PT. Four types of analysis were performed for this purpose. In each analysis, we changed one weight while keeping the other weights constant.

Figure 2 gives the changing curve for each riding right (from 1 to 5) obtained by different weight values

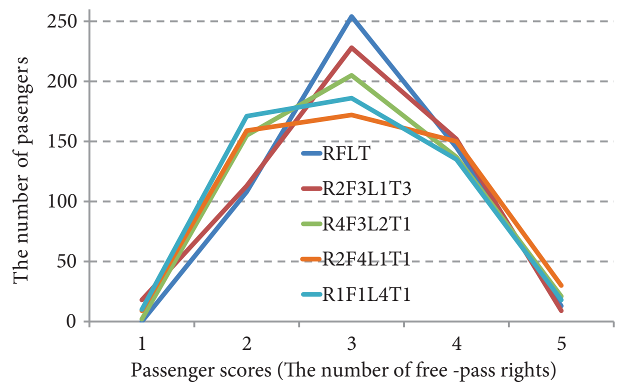

Figure 1. Comparison of RFLT and wRFLT with different weight values.

in the wRFLT method. The maximum weight value was limited to 3 times in order to avoid strong effects of the parameters. In all cases, the passengers were awarded 3 free rides at most, while they were assigned 1 free ride at least. For example, according to R3F1L1T1 results, 144 people won 2 free rides, 201 people 3 free rides, and 154 people 4 free rides, while only 19 and 2 people won 5 and 1 free rides, respectively. All models produced nonzero free-pass rights, except the RTLT model, in which nobody won 1 free-pass right. When weight values differ, the number of passengers winning 1 and 5 pass rights does not quite change, while those winning 2, 3, and 4 pass rights are considerably influenced. The minimum decrease in the number of those winning 3 pass rights occurs when the loyalty weight value is increased. The most balanced score distribution (4 free rides = 157, 3 free rides = 172, 2 free rides = 161) is realized when the frequency weight value is increased. The fact that loyalty, recency, time, and frequency attributes affect passenger scoring results is displayed with this study.

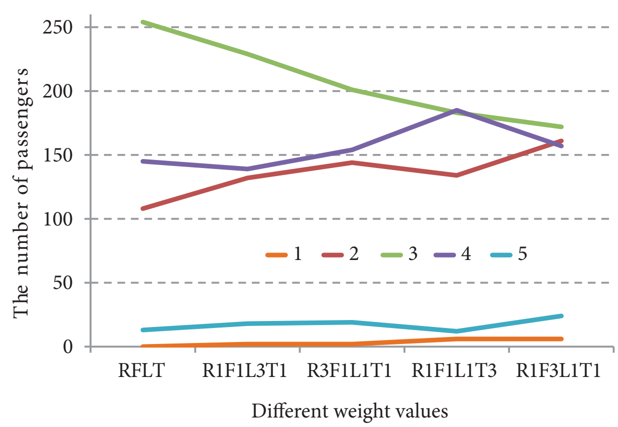

Figure 2. Sensitivity analysis of the RFLT model with different weight values.

4.4. Discussion

The strategies aiming to change behavior regarding PT can be broadly classified into experiments aimed at encouraging the use of PT [2, 3, 5, 6, 51] and those aimed at reducing peak congestion [48, 52]. The proposed scoring model in the present study, named RFLT, aims to achieve these two objectives, increasing the ridership by analyzing the recency (R), frequency (F), and loyalty (L) parameters and reducing travel demand in rush hours by considering the time (T) parameter.

Different promotional strategies can be designed to persuade or remind target clients about PT ridership. The promotions can be designed to identify potential clients, keep loyal clients, increase the use of a product/service, get people to try products/services, teach clients about potential services, provide information, and even create awareness [53]. Any PT agency or institution may seek to achieve one or more of these goals. For example, it can focus on infrequent passengers since they need encouragement or it can focus on frequent passengers to keep loyal clients.

Mechanisms involving incentives usually offer monetary incentives, such as free PT tickets, to encourage irregular car users to try alternative behavior such as using PT. Hence, infrequent, nonrecent, and/or nonloyal passengers of PT can become permanent if they are encouraged with free PT tickets. Once a number of participants adopting the behavior use free PT tickets, then they can begin to use PT frequently by changing their habitual behavior. In order to serve this purpose, a scoring mechanism can be defined to identify target passengers and to give them a free-pass promotion with the purpose to shift them from private transportation to PT.

Previous research on fare-free PT has shown that different policies can be designed to enhance the attractiveness of riding PT and has emphasized a critical need to "clearly identify the objectives addressed by the free-pass policy" [8, 51]. Based on this motivation, our scoring method is designed to be able to promote recent, frequent, loyal, and off-peak hour passengers through R, F, L, and T values with their weight parameters W_R, W_F, W_L, and W_T, respectively, and nonrecent, infrequent, nonloyal, and peak-hour passengers through the β parameter. These parameters (actually target passengers) can be adjusted according to the decisions of the PT agency where scoring is done. The findings of the experiments performed with different parameter settings will be useful to policymakers evaluating free-pass PT and considering how best to target and promote relevant policy. Our model can therefore be classified as a combination of a commitment strategy (supporting willingness to commute by PT as usual) and an incentive strategy (differentiating potential users, who need encouragement).

The time parameter in our RFLT model is important in order to provide a higher service quality, especially availability, in PT by encouraging riders to use PT at off-peak hours. Traveling at out-of-peak hours is more valuable to reduce passenger volume (i.e. overcrowded boarding) at peak hours. The policy that provides peak-hour travelers with easier access to an unfamiliar option through a monetary incentive (free PT tickets) may encourage them to use the PT option at off-peak hours on an occasional or even regular basis. Hence, in our model, a time-based free-pass promotion strategy is also designed to route passengers out of peak time.

A number of experiments [48, 52] have been conducted whereby a free PT pass or discount is given to participants for traveling within the off-peak period. One example [48] showed that 23% of rail passengers in Melbourne, Australia had shifted their travels away from the morning peak time (before 7:00 a.m.) with an average time shift of 42 min. Another example [52] employed several off-peak incentives to especially shift PT demand from the peak hour to less congested times of day (before or after the peak). Their results showed that off-peak strategies have a positive impact on travel demand (reducing the peak ridership). All these experiments have proved that the promotions related to peak and off-peak hours serve as an incentive for passengers to shift their travel times to an earlier or later period. Based on this motivation, our model is designed to provide free-pass opportunities with the purpose to encourage passengers, who tend to shift their arrival/departure times to their jobs/home, to make their trips at off-peak hours.

In our model, both peak and off-peak hour passengers can be rewarded by adjusting the W_T and β parameters. As the passengers travel at off-peak hours, they are rewarded with higher scores (i.e. more free PT tickets) since the T value increases. The previous studies in the literature [48, 52] support our assumption that promoting off-peak hours can be effective in shifting trips outside the peak period. On the other hand, the policymaker can determine a high value for the β parameter to be able to support peak-hour passengers. Hence, different free-pass promotion strategies can be applied by the PT agencies where scoring is done.

The parameters we have formulated for use with the RFLT model hold the promise of a long-term sustainable policy and aim to take the traffic situation in a region under control over time. These are the main policies that are currently employed by municipalities or PT authorities.

5. Conclusion and future works

One of the important policies to increase the use of PT is through free-fare promotions. In order to determine how many free-pass rights the passengers will win the next month, the present study proposes a novel scoring method, namely RFLT. RFLT is a modified version of a well-known marketing method, RFM, and is based on four key metrics: recency, frequency, loyalty, and time. *Recency* is the number of days between the most recent travel time date and today. *Frequency* measures the number of journeys made in a given time period. *Loyalty* is the number of days from the first transport ride until today. *Time* is the ratio of off-peak hour travel counts to total usage counts.

Experimental studies conducted using real-world data collected through an NFC-enabled mobile payment application used in a city in Turkey show that the proposed RFLT method is suitable for PT and can be applied effectively. However, it is observed that the weighted version provides a more balanced score distribution. The sensitivity analysis performed showed that the frequency parameter (F) among the 4 parameters (R, F, L, and T) influences the score results the most. The study results can be used to establish an efficient policy to increase PT ridership.

In a future study, different parameters such as passenger types (adult, student, discounted, etc.), distance covered, type of vehicle (metro, sea transportation, bus, etc.), or demographic properties of passengers (age, economic status, education level, etc.) can be added to the RFLT model according to need and to make it more comprehensive. The parameters in question can be tailored based on the purposes of the municipalities. Another future study can cover different prizes such as giving free theater/movie tickets instead of free-pass rights in order to encourage the use of PT.

Acknowledgment

We thank Kent Kart Ege Elektronik San. Tic. A.S. for supporting this project. Its support gave us the chance to develop our new scoring method and analyze the results.

References

[1] Brög W, Erl E, Mense N. Individualised marketing changing travel behaviour for a better environment. In: OECD Workshop on Environmentally Sustainable Transport; Berlin, Germany; 2002. pp. 6-12.

[2] Jomnonkwao S, Sangphong O, Khampirat B, Siridhara S, Ratanavaraha V. Public transport promotion policy on campus: Evidence from Suranaree University in Thailand. Public Transport 2016; 8 (2): 185-203. doi: 10.1007/s12469-016-0122-2

[3] Cools M, Fabbro Y, Bellemans T. Free public transport: a socio-cognitive analysis. Transportation Research Part A: Policy and Practice 2016; 86: 96-107. doi: 10.1016/j.tra.2016.02.010

[4] Hughes A. Strategic Database Marketing. Chicago, IL, USA: Probus Publishing Company, 1994.

[5] Thogersen J, Moller B. Breaking car use habits: the effectiveness of a free one-month travelcard. Transportation 2008; 35 (3): 329-345. doi: 10.1007/s11116-008-9160-1

[6] Galey D. License to ride: free public transportation for residents of Tallinn. Critical Planning Journal 2014; 21 (1): 19-33.

[7] Belter T, Harten M, Sorof S. Advantages and Disadvantages of Free Public Transport Services. Vaxjo, Sweden: SustraMM, 2012.

[8] Fearnley N. Free fares policies: impact on public transport mode share and other transport policy goals. International Journal of Transportation 2013; 1 (1): 75-90. doi: 10.14257/ijt.2013.1.1.05

[9] Gase LN, Kuo T, Teutsch S, Fielding JE. Estimating the costs and benefits of providing free public transit passes to students in Los Angeles county: lessons learned in applying a health lens to decision-making. International Journal of Environmental Research and Public Health 2014; 11 (11): 11384-11397. doi: 10.3390/ijerph111111384

[10] Blattberg RC, Kim BD, Neslin SA. Database Marketing: Analyzing and Managing Customers. New York, NY, USA: Springer-Verlag, 2008.

[11] Stormi K, Lindholm A, Laine T, Korhonen T. RFM customer analysis for product-oriented services and service business development: an interventionist case study of two machinery manufacturers. Journal of Management and Governance 2020; in press. doi: 10.1007/s10997-018-9447-3

[12] Vlasselaer VV, Bravo C, Caelen O, Rad TE, Akoglu L et al. APATE: a novel approach for automated credit card transaction fraud detection using network-based extensions. Decision Support Systems 2015; 75: 38-48. doi: 10.1016/j.dss.2015.04.013

[13] Ghodousi M, Alesheikh AA, Saeidian B. Analyzing public participant data to evaluate citizen satisfaction and to prioritize their needs via K-means. Cities 2016; 55: 70-81. doi: 10.1016/j.cities.2016.03.015

[14] Dursun A, Caber M. Using data mining techniques for profiling profitable hotel customers: an application of RFM analysis. Tourism Management Perspectives 2016; 18: 153-160. doi: 10.1016/j.tmp.2016.03.001

[15] Moreno JB, Lorente AT, Porcel C, Viedma EH. A new model to quantify the impact of a topic in a location over time with Social Media. Expert Systems with Applications 2015; 42 (7): 3381-3395. doi: 10.1016/j.eswa.2014.11.067

[16] Ravasan AZ, Mansouri T. A fuzzy ANP based weighted RFM model for customer segmentation in auto insurance sector. In: Khosrowpour M (editor). Intelligent Systems: Concepts, Methodologies, Tools, and Applications. Hershey, PA, USA: IGI Global, 2018, pp. 1050-1067.

[17] Chang HC. Developing EL-RFM model for quantification learner's learning behavior in distance learning. In: IEEE Proceedings of 2nd International Conference on Education Technology and Computer; Shanghai, China; 2010. pp. 452-454.

[18] Wang YJ, Di G, Yu J, Lei J, Coenen F. Feature representation for customer attrition risk prediction in retail banking. In: Perner P (editor). Advances in Data Mining: Applications and Theoretical Aspects, Lecture Notes in Computer Science. Berlin, Germany: Springer, 2013, pp. 229-238.

[19] Chen YS, Cheng CH, Lai CJ, Hsu CY, Syu HJ. Identifying patients in target customer segments using a two-stage clustering-classification approach: a hospital-based assessment. Computers in Biology and Medicine 2012; 42 (2): 213-221. doi: 10.1016/j.compbiomed.2011.11.010

[20] Wei JT, Lin SY, Weng CC, Wu HH. A case study of applying LRFM model in market segmentation of a children's dental clinic. Expert Systems with Applications 2012; 39 (5): 5529-5533. doi: 10.1016/j.eswa.2011.11.066

[21] Wei JT, Lee MC, Chen HK, Wu HH. Customer relationship management in the hairdressing industry: an application of data mining techniques. Expert Systems with Applications 2013; 40 (18): 7513-7518. doi: 10.1016/j.eswa.2013.07.053

[22] Babaiyan V, Sarfarazi SA. Analyzing customers of south Khorasan telecommunication company with expansion of RFM to LRFM Model. Journal of AI and Data Mining 2019; 7 (2): 331-340. doi: 10.22044/JADM.2018.6035.1715

[23] Chiu CY, Kuo IT, Chen PC. A market segmentation system for consumer electronics industry using particle swarm optimization and honey bee mating optimization. In: Chou SY, Trappey A, Pokojski J, Smith S (editors). Global Perspective for Competitive Enterprise, Economy and Ecology Advanced Concurrent Engineering. London, UK: Springer, 2009, pp. 681-689.

[24] Huang D, Zhou J, Wang H. RFMS method for credit scoring based on bank card transaction data. Statistica Sinica 2019; 28 (4): 2903-2919. doi: 10.5705/ss.202017.0043

[25] Peker S, Kocyigit A, Eren PE. LRFMP model for customer segmentation in the grocery retail industry: A case study. Marketing Intelligence and Planning 2017; 35 (4): 544-559. doi: 10.1108/MIP-11-2016-0210

[26] Zhang Y, Bradlow ET, Small DS. Predicting customer value using clumpiness: from RFM to RFMC. Marketing Science 2015; 34 (2): 195-208. doi: 10.1287/mksc.2014.0873

[27] Chen Y, Liu Z, Wang Y, Qin Y. Research on user segmentation based on RFL model and k-means clustering algorithm. In: International Conference on Logistics, Engineering, Management and Computer Science (LEMCS 2015); Shenyang, China; 2015. pp. 1499-1503. doi: 10.2991/lemcs-15.2015.298

[28] Cho YS, Kim KA , Moon SC, Park SH, Ryu KH. Effective purchase pattern mining with weight based on FRAT analysis for recommender in e-commerce. In: Park J, Stojmenovic I, Jeong H, Yi G (editors). Computer Science and its Applications, Lecture Notes in Electrical Engineering. Berlin, Germany: Springer, 2015, pp. 443-454.

[29] Liu Z, Chen G. RFAT customer segmentation based on improved K-means algorithm. Journal of Nanjing University of Science and Technology 2014; 38 (4): 531-536.

[30] Al-Shayea QK, Al-Shayea TK. Customer behavior on RFMT model using neural networks. In: World Congress on Engineering (WCE 2014); London, United Kingdom; 2014. pp. 49-52.

[31] Golsefid SMM, Turksen IB, Zarandi MHF. A systematic type-2 fuzzy optimization model for global market analysis and its application. In: Annual Meeting of the North American Fuzzy Information Processing Society (NAFIPS); Berkeley, CA, USA; 2012. pp. 1-6.

[32] Chang HC, Tsai HP. Group RFM analysis as a novel framework to discover better customer consumption behavior. Expert Systems with Applications 2011; 38 (12): 14499-14513. doi: 10.1016/j.eswa.2011.05.034

[33] Bizhani M, Tarokh MJ. Behavioral rules of bank's point-of-sale for segments description and scoring prediction. International Journal of Industrial Engineering Computations 2011; 2 (2): 337-350. doi: 10.5267/j.ijiec.2010.04.002

[34] Yan B, Chen G. AppJoy: personalized mobile application discovery. In: International Conference on Mobile Systems; Bethesda, MD, USA; 2011. pp. 113–126.

[35] Yeh IC, Yang KJ, Ting TM. Knowledge discovery on RFM model using Bernoulli sequence. Expert Systems with Applications 2009; 36 (3): 5866–5871. doi: 10.1016/j.eswa.2008.07.018

[36] Wei Z, Shan X. RFM high-speed railway customer value classification model based on spark, IOP Conference Series, Materials Science and Engineering 2019; 563 (5): 1-6. doi: 10.1088/1757-899X/563/5/052081

[37] Chiang WY. Establishing high value markets for data-driven customer relationship management systems: an empirical case study. Kybernetes 2019; 48 (3): 650-662. doi: 10.1108/K-10-2017-0357

[38] Qian C, Yang M, Li P, Li S. Application of customer segmentation for electronic toll collection: a case study. Journal of Advanced Transportation 2018; article id 3635107. doi: 10.1155/2018/3635107

[39] Dehghanizadeh M, Fathian M, Gholamian M. LDcFR: a new model to determine value of airline passengers. Tourism and Hospitality Research 2018; 18 (3): 357-366. doi: 10.1177/1467358416663821

[40] Chiang WY. Analysing weekend city travellers' customer value to business. Urban Design and Planning 2017; 170 (4): 144-150. doi: 10.1680/jurdp.16.00026

[41] Chiang WY. Discovering customer value for marketing systems: an empirical case study. International Journal of Production Research 2017; 55 (17): 5157-5167. doi: 10.1080/00207543.2016.1231429

[42] Ma J, Luo S, Wu L, Luo G, Li X et al. Research on highway passenger segmentation based on Canopy-kmeans clustering algorithm under parallel computing framework. In: Proceedings of IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation; San Francisco, CA, USA; 2017. pp. 1-6.

[43] Chiang WY. Applying data mining with a new model on customer relationship management systems: a case of airline industry in Taiwan. Transportation Letters 2014; 6 (2): 89-97. doi: 10.1179/1942787514Y.0000000014

[44] Chiang WY. Applying a new model of customer value on international air passengers' market in Taiwan. International Journal of Tourism Research 2012; 14: 116-123. doi: 10.1002/jtr.840

[45] Carmona M. Relationship marketing in urban public transport: a new approach for building competitive advantages. In: Proceedings of 9th International Conference on Competition and Ownership in Land Transport; Lisbon, Portugal; 2005.

[46] Sun S, Duan Z. Modeling passengers' loyalty to public transit in a two-dimensional framework: a case study in Xiamen, China. Transportation Research Part A: Policy and Practice 2019; 124: 295-309. doi: 10.1016/j.tra.2019.04.007

[47] Minser J, Webb V. Quantifying the benefits: application of customer loyalty modeling in public transportation context. Transportation Research Record 2010; 2144 (1): 111-120. doi: 10.3141/2144-13

[48] Currie G. Quick and effective solution to rail overcrowding: free early bird ticket experience in Melbourne, Australia. Transportation Research Record: Journal of the Transportation Research Board 2010; 2146 (1): 35-42. doi: 10.3141/2146-05

[49] Kohavi R, Parekh R. Visualizing RFM segmentation. In: Proceedings of the 4th SIAM International Conference on Data Mining; Florida, USA; 2004. pp. 391-399.

[50] Olson DL, Chae BK. Direct marketing decision support through predictive customer response modeling. Decision Support Systems 2012; 54 (1): 443-451. doi: 10.1016/j.dss.2012.06.005

[51] Hess DB. Decrypting fare-free public transport in Tallinn, Estonia. Case Studies on Transport Policy 2017; 5 (4): 690-698. doi: 10.1016/j.cstp.2017.10.002

[52] Lovric M, Raveau S, Adnan M, Pereira FC, Basak K et al. Evaluating off-peak pricing strategies in public transportation with an activity-based approach. Transportation Research Record: Journal of the Transportation Research Board 2016; 2544 (1): 10–19. doi: 10.3141/2544-02

[53] Gitman LJ, McDaniel C, Shah A, Reece M, Koffel L et al. Introduction to Business. Houston, TX, USA: OpenStax, 2018.