Sigma J Eng & Nat Sci 38 (4), 2020, 1681-1692

Sigma Journal of Engineering and Natural Sciences
Sigma Mühendislik ve Fen Bilimleri Dergisi

Research Article

RECYCLING PROJECT WITH RFM ANALYSIS IN INDUSTRIAL MATERIAL SECTOR

Semra ERPOLAT TAŞABAT*[1], Esra AKCA[2]

[1]*Department of Statistics, Mimar Sinan Fine Arts University, ISTANBUL;* ORCID: 0000-0001-6845-8278
[2]*Department of Statistics, Mimar Sinan Fine Arts University, ISTANBUL;* ORCID: 0000-0003-3790-8584

Received: 13.04.2020 Revised: 26.07.2020 Accepted: 01.09.2020

ABSTRACT

With the advancement of technology and the widespread use of the Internet, the concept of big data, which we are often beginning to hear its name, has emerged. Big data can be briefly defined as an unstructured data stack. It aims to transform the data collected from different sources into a meaningful and processable format. One of these methods is RFM analysis. RFM analysis is an effective and practical marketing model that combines the initials of **R**ecency **F**requency and **M**onetary and performs behavioral customer segmentation. In this study, the importance of RFM analysis was emphasized. How to use RFM analysis in estimation models is explained in detail.The applicability of RFM analysis to the recycling project has been demonstrated. The operation of the model and the application of RFM to recycling projects are shown in the original retail data of a company operating in the metal industry. Therefore, they were encouraged to participate in recycling. According to the contribution rate of recycling, it is aimed to establish a profitable relationship between the customer and the company by applying discounts to the customers.
Keywords: RFM, recycle, clustering, CRM, waste.

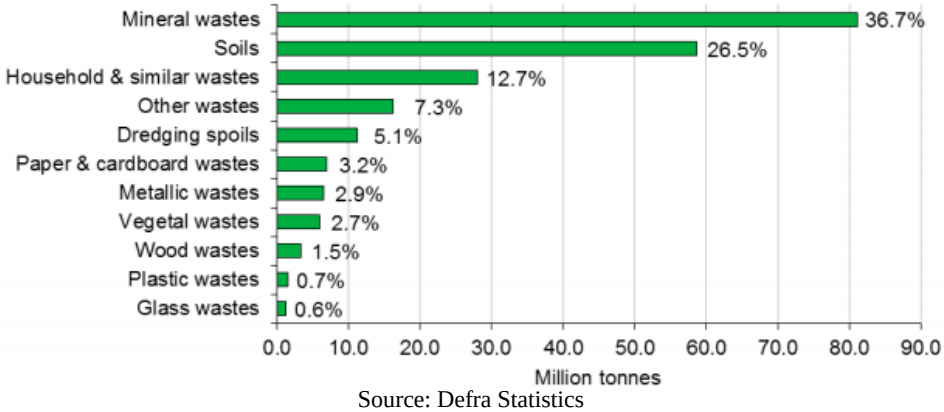
1. INTRODUCTION

The amount of information produced and stored at the global level is unimaginably large and is growing every day. Selecting and analyzing the fit for purpose among these big data stacks plays a major role in making the right decisions by developing effective strategies. The main purpose of data analysis is to ensure that the accumulated data is achieved with logical, useful and effective results by performing purposeful analyses. Big data analysis, which includes a wide range of solution suggestions, can cover different methods depending on the topic covered. In doing so, it is possible to utilize many different analyses, models and algorithms. One of them is RFM analysis.

RFM analysis is a valid marketing model for behavioral customer segmentation. RFM, which can be applied to different fields, can also be applied on recycling projects with this study. Thus, the analysis of big data on the social responsibility project was carried out for the first time with RFM.

* Corresponding Author: e-mail: semra.erpolat@msgsu.edu.tr, tel: (212) 236 69 36 /5501

Recycling is a very important and necessary requirement, especially for the world, which is getting more and more polluted every day. Today, recycling is important for sustainability. The fact that progress in different fields is possible by acting in conjunction with natural sciences has further increased the importance of re cycling. Recycling is one of the main factors, especially in a more livable world, such as reducing human health and environmental risks, improving productivity and performance. Recycling, which is also of great importance to the production sector, provides benefits at many points, especially money savings and reduction in production costs. For this reason, enterprises operating in the production sector should take their part in recycling. The estimates of household, commercial and other waste diverted are shown in Figure 1 [1].



Source: Defra Statistics

Figure 1. Tons of waste material, UK, 2016

Recycling can be carried out in different areas. One of these areas is metal. Recycling metals provides protection of natural resources, while using raw materials to produce new products requires less energy than energy. Recycling emits less carbon dioxide and other harmful gases. It saves money and allows manufacturing businesses to reduce production costs. Besides, i.e. since environmental protection awareness contributes to the development of the organization, employees have a sense of "sensitive consumer" and increase their reputation by ensuring that the business has the title of "environmentalist" in national and international marketsare also provided.

While managing big data is an important point; there have been very rare studies of recycling in the world as analysis. One of them is the study of hydrogen storage cost analysis by Cassidy Houchins [2]. The purpose of this analysis has been to reduce costs. Another study is the conversion of used oil into valuable diesel by Anthony Kasozi [3]. As a result of this study, it is aimed to reduce the environmental pollution caused by waste oil.

There has been no study in Turkey using any analysis method in which a segment strategy suitable for recycling is developed. Incentive methods for everyone in general; Plastic bottle piggy bank filling according to the number of bottles, devices that give pens in exchange for recyclable waste, machines that pay “1 TL” in exchange for liquid oil waste are used. For this reason, by segmenting the customers by reaching the right people; the right to special discounts for customers who can contribute to the recycling of metal waste is an incentive method applied for the first time in this study. In this study, which dealt with this important issue, analyses were carried out by RFM method. In the study, a new methodwas proposed for calculating the RFM score account differently. This method and approach has been studied on the 2017 data collected by a company operating in the production of metal kitchen ware in Istanbul.

In the following sections of the study, RFM analysis and different RFM models will be informed. Then, the importance of recycling for the sustainability of human life will be emphasized and a model will be proposed so that RFM can adapt to recycling projects. How to implement this proposed model will be carried out on the selected data set. Finally, the results will be interpreted.

2. REVIEW OF LITERATURE

2.1. For Recycle

Recycling is very important for environment, living life, wellness and the sustainable world. Environmental, air and water pollution lead the world into an uninhabitable environment. For our future natural resources need to be preserved. Recycling and reusing iron, steel, copper, lead, paper, plastic and glass will prevent the consumption of natural resources. A significant portion of recycled materials is metal, plastic and glass waste used in food and beverage packaging, as well as paper and cardboard. Recycling protects our natural resources. Energy savings are provided. The amount of waste is decreasing. Recycling also means investment in the future and the economy. The following are the studies on recycling.

In 2010, the Caught Green Handed campaign was launched in Ohio, USA as one of the "Grab the Prize for Contributing to Recycling" project. Consumers who brought waste to recycling bins were greeted with gift cards from various stores. The few "green handed" at first increased as news spread that prizes were being distributed. This campaign was later implemented in other regions, owned by university students and later spread more [4].

In 2010, an application launched in Barcelona aimed at recycling kitchen waste oils. Consumers who acquired containers called "Olipots" distributed free of charge by the municipality, collected their waste oils in these containers and began to recycle "green dots" placed in certain areas of the city. When the demand from the townspeople was high, the municipality produced more than the Olipots, and each house had one. Waste oils collected and sent to recycling plants are converted into biodiesel [5].

These machines, which pay for buying beverage cans, were first patented by the American company Reverse Vending® in 1920. The first machine to work was manufactured by Sweden's Wicanders in the late 1950s, and the developer was Norway's Aage Tvetian in 1962. For more than 50 years in Europe, these machines, usually in grocery stores, are in use on the principle of collecting empty beverage cans and returning the deposit price to the consumer. There is even a computer game room in Sweden with plastic bottle-operated computers. Public transport tickets, such as subways, can be purchased from these vending machines in Japan and in many countries [6].

2.2. For RFM Analysis

RFM analysis has an important role in data mining. There have been very successful results in customer relations. An important study of RFM analysis is target company's "Pregnancy Prediction Score" application, one of the major USA retailers. Analysis stages: A customer is assigned an identification number (ID). So it's being hacked into the system. All purchases are recorded in connection with this customer ID. By examining the history formed, 25 different products purchased according to the stage of pregnancy are determined. As the customer receives these products, the association and meaning continues. When the target score is reached, product promotion and discount vouchers are sent to the customer concerned. Target managed to increase its revenues by 23$ billion between 2002 and 2010 [7].

3. RFM MODEL

RFM concept was first introduced in 1996 by Jan Roelf Bult and Tom Wansbeek as a marketing estimation application of economic models combined with statistical techniques [8]. RFM Mahboubeh Khajvand, part of the decision support system, was used in 2011 to segment customers and find the targeted audience in the most accurate way, a useful, simple and powerful consumer, Customer Relationship Management (CRM) is the application model, hedefined. In addition, from a different point of view, the Pareto Principle the 80/20 rule, or 80% of the results, comes from 20% of the causes. Similarly, 20% of your customers contribute 80% of your total revenue. The Pareto Principle is at the center of the RFM model. Focusing your efforts on critical customer segments will give you a higher return on investment, it contains logic.

RFM is mostly used in e-commerce-marketing discount and campaign application, banking, credit and stock sales. Bult and Wansbeek defined these variables, which combine with the combination of initials, as follows: Recency-Frequency-Monetary means proximity-frequency-expenditure. Recency refers to the customer's most recent purchase, frequency refers to the frequency of the customer's shopping frequency, and monetary refers to the customer's total shopping amount. In the RFM analysis model, differences were also made according to the function. Examples include:

- RFD (Innovation, Frequency, Duration). The time spent here is time spent. Especially when analyzing consumer behavior of audience/reader/surf-oriented products is particularly useful.
- RFE (Innovation, Frequency, Participation). Participation can be a composite value depending on the time spent on the page, pages per visit, bounce rate, social media participation, etc. It is especially useful for online businesses.
- FRAT (Frequency, Renewal, Quantity and Type of Goods) is an extended version. In general, the purpose of this method of analysis is:
  - Who is your client?
  - Which customers are most profitable?
  - Which of your customers has recently visited you?
  - Which of your customers is most loyal?
  - For which campaigns did your customers react?
  - Which customers are missing?
  - Which customers are faced with losing?
  - Is the making of appropriate decisions by finding answers to the questions in the form of?

The underlying logic of all these questions is that we can't do anything about it if the customer has recently made a purchase, the potential to shop again is high, if he frequently makes shopping, there is a high probability of making a purchase, if the total amount of purchases left by the customer is high, he can still shop.

RFM reduces marketing costs due to optimal targeting. Controlled targeting reduces negative responses from customers.

4. DATA MINING AND RFM

4.1. Clustering Using RFM

In recent years, some researchers have considered RFM variables in improving clustering. For example: Hosseini and his friends combined the weighted RFM model with K-Means to improve CRM [9]. An employer in Taiwan has implemented the RFM model and K-Means method in customer value analysis to reinforce customer loyalty for long-term profitable customers [10]. Chuang and Shen first assessed the importance of R, F and M weights with analytical hierarchical

process methodology and then customer life values. Finally, with cluster analysis, it has ranked valuable groups of customers using a self-organizing map method. In short, clustering is used to find similar customer segments.

4.2. Classification Using RFM

The integration of classification techniques and RFM was used by Olson et al. in 2009 to analyze customers' response suppositions to a particular product promotion. In the same year, Cheng and Chen compared and discussed three data mining techniques: The relative change in customer segmentation between logistics regression, decision trees and neural network algorithms. He used the classification decision tree technique to estimate the RFM values of the next customers from current estimates and to see the changes and transition allocations of RFM values over time [11].

4.3. Integrated Approach

The use of RFM analysis in data mining; offers a new three-step approach using a few parameters together. In the approach, classification rules are discovered using RFM values, (i) demographic variables using the clustering task to find (ii) customer segments in a similar way (age, gender, education level, etc.), and to predict the future RFM values of customer segments, customer behavior, (iii) at the end; product union rule mining is carried out. This model is meant to help managers develop exactly better marketing strategies using data mining and RFM Analytics [12].

5. RFM ANALYSIS

RFM analysis can be calculated by multiple methods and formulas. The equation (1) contains the general formulation of RFM.

$$(RecencyScore \times RecencyWeight) + (FrequencyScore \times FrequencyWeight) + (MonetaryScore \times MonetaryWeight) \quad (1)$$

"Recency, Frequency, Monetary" in equation (1); date, frequency and monetary are divided into 5 intervals of 20% in itself, the highest 20% of the 5, the highest 2. part 4, 3. part 3, 4. part 2 is calculated by giving 1 point to the lowest or the last lowest or the 20% of the 20%. The weights in the formula underlying this logic differed according to the people and theories. These are listed below, respectively.

5.1. Model 1

$$(RecencyScore \times 100) + (FrequencyScore \times 10) + (MonetaryScore) \quad (2)$$

As can be seen in equation (2), the most important factors in RFM analysis are Recency, Frequency and Monetary. This is how the RFM score is calculated for each customer.

5.2. Model 2

$$\left(R \times 3\right) + (F \times 2) + (M \times 1) \quad (3)$$

Model 2 was published by John Miglautsch in The Journal of Database Marketing on May 28. It was introduced in 2000. The equation (3), logic gives the nearest shoppers more power for return, and also gives some support to the frequency. The logic behind increasing the frequency is

that if two customers have the same rate of recession, the buyer who purchases more frequently will respond if one of them orders only once.

5.3. Model 3

$$\left(R \times 9.9\right) + (F \times 6.6) + (M \times 3.3) \quad (4)$$

The equation (4) was developed by Tsai and Chiu in 2004, and companies that want to act on a 100-point scale of logic in this model use this method to have a score image in the range of 19.8 to 99 points. Instead of multiplying by 3, 2, 1, as in Model 2, the compound number is obtained by multiplying 9.9 by 6.6, 3.3.

5.4. Model 4

$$(R\ Skor) + (F\ Skor) + (M\ Skor) \quad (5)$$

The equation (5), the logic is that because the scores are evenly weighted, by collecting these 3 scores without any weight multipliers; for example, write as a total score ranging from 3 to 15, such as 5 + 2 + 3 = 10 or 1 + 3 + 2 = 6. The basic requirement in this method depends on the company's equal evaluation of the three main features. Since the values will be evaluated equally, instead of being evaluated side by side, it is easier to evaluate the result of the result-oriented and large-to-small ranking, rather than individual evaluations [13].

RFM analysis and score calculations can then be combined with RFM analysis and data mining by performing Basic Components Analysis and Cluster Analysis. While the variables that are weighted in the analysis of the basic components are seen, the clustering analysis creates classes according to average.

6. RECYCLE PROJECT WITH RFM ANALYSIS

In the analysis, the data set; Turkish-based metal kitchen ware sourcing company includes transactions between 01/01/2017 and 31/12/2017. Because it is a manufacturing company, each customer can shop once every month. There is no repetitive data. This is how the data is arranged. The aim of the analysis is to improve the quality of life, to identify the customers who can contribute the most to recycling, which is a social responsibility project. Thanks to the recyclings received, a discount will be applied, and the company's production costs will be reduced and profit-making will be provided. Most frequently, customers who shop and pay the most in the most recent date range are potentially the ones with the most tendency to recycle products they've bought in the past. Analysis will be improved for this purpose. The method used in the study is Model (1), Model (2), Model (3), Model (4).

The data has a total of 284 customers and 3 different variables. Here are the descriptions of the variables in the data set:

Customer Name: Each different buyer who purchases from the company.
Monetary : The total amount of expenditure for each customer within a year.
Frequency : The total number of order bears in a year.
Recency : The last ordered month in a year.

The Table 1 below contains the information of 10 customers as an example.

Table 1. Customer information

CUSTOMER NAME	RECENCY	FREQUENCY	MONEATRY
Customer1	1.10.2017	10	416.938.58 TL
Customer2	1.12.2017	12	187.934.02 TL
Customer3	1.12.2017	12	459.193.52 TL
Customer4	1.12.2017	12	147.629.17 TL
Customer5	1.12.2017	12	356.535.83 TL
Customer6	1.12.2017	9	1.336.913.71 TL
Customer7	1.06.2017	6	35.404.19 TL
Customer8	1.12.2017	12	188.998.24 TL
Customer9	1.12.2017	12	172.308.37 TL
Customer10	1.10.2017	7	124.903.68 TL

RFM has a "1 = weakest customer", "5 = most valuable customer" scale from 1 to 5. It is called a "customer value criterion". This range determination and scoring method depends on individual businesses. Because businesses decide which ranges are ideal for innovation, frequency and monetary values, they can change. However, the programs used today automatically dissect 100% into 20% without having to determine any range and match it to the range of 1-5. The IBM WATSON STUDIO Modeler, which we used in the study, made the distinction in itself. IBM Watson Studio SPSS Modeler provides the tools to work collaboratively and easily with data to build and train scale-based models. In this way, data science is enabled to be implemented in a shorter time. It helps to modernize predictive analytics and machine learning processes and accelerate the time it takes to achieve value. Uses advanced analytical capabilities including predictive analytics, interim statistical analysis, predictive modeling, data mining, text analytics, optimization, real-time scoring, and machine learning. These tools help organizations discover models in data and predict what will happen next, beyond knowing what happened in the past [14].

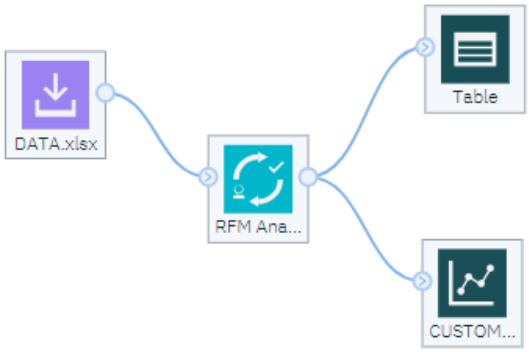


Figure 2. IBM Watson Studiodecision stream

First data has been transferred to IBM Watson Studio. Modeler Flow has been selected and the analysis has begun. RFM Analysis was integrated into the transferred data and a decision

diagram was created. It was requested that the outputs be given in the form of tables and charts (Figure 2).

According to the models wants to apply RFM Analysis, the relevant weight areas are written on the relevant weights and score calculations are made for each model. The images below (Figure 3) show weight calculations of different models. It is observed that all 5 scoring scales are common and there is differentiation in weights. Scores and graphs are taken on the basis of each formula. IBM Watson Studio RFM outputs are provided as follows.

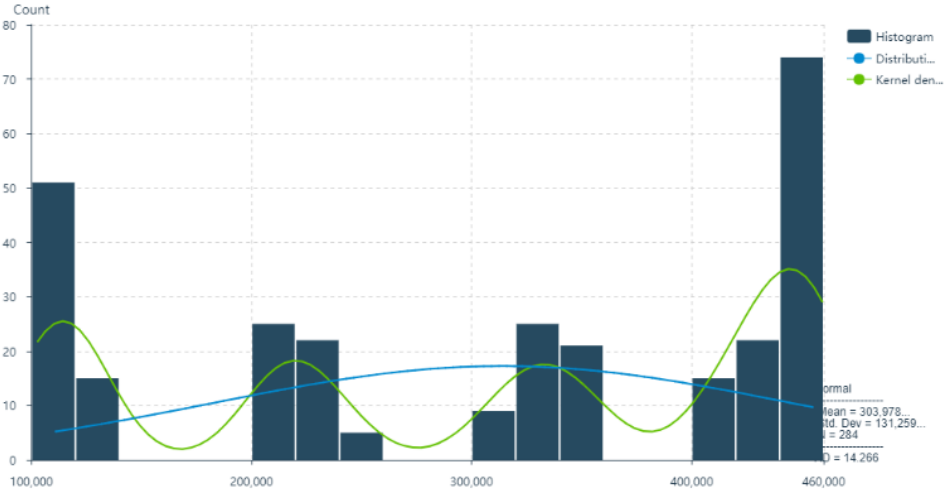


Figure 3. Model (1) Modeler RFM score chart

Table 2. Key Segments of RFM Analysis

SEGMENTS	RFM	EXPLANATION	MARKETING MODEL
Best Customers	555	Those who have recently made purchases, often performing high amounts of purchases.	They can be the main goal of work on new products and loyalty programs.
Loyal Customers	X5X	Most purchasers.	They can be the target audience of new products and loyalty programs, it is useful to divide them into sub-segments.
Large Spending	XX5	Customers with the highest amount of shopping.	You can market your most expensive products, they like to play big.
Almost Lost Customers	355	Those who have not done shopping recently but who have done frequent shopping in the past.	You can convince them back with aggressive discount offers.
Lost Customers	155	Those who have not done shopping for a long time but who have done frequent shopping in the past.	With discount offers you can convince them back.
Escaped Customers	111	Those who haven't been shopping for a long time but are not very active in the past.	It is important to regain the customer, but there is no need to spend a significant budget.

The average was 303,978 as shown in the chart. The normality distribution is shown in blue and the core density estimate is green.

The calculated results of the Model 1 can be interpreted in Table 2. Decisions can be made on the right customer groups by acting in accordance with the segment groups found. This segment distinction will be meaningful in order to focus on campaigns and be effective in communication activities. In this way, customer potentials are determined according to the score results and their return to the company will be easier and strategic.

Benefits to the Company; more increase in response rates and conversion rates. As a result, the revenue generated from marketing efforts through targeting will be increased. The image below shows that the size of the boxes varies according to RFM score frequencies for Model 2 (Figure 4). While the minimum frequency is of 11, it can be said that the maximum frequency belongs to 27.

Figure 4. Model (2) Data Refinety Flow RFM score chart

A graph of the RFM scores of Model (3) is shown in Figure 5. Scores are more divided by weight productivity.

Finally, the graph of the Model (4) is shown in Figure 6. Frequency distributions are observed according to RFM scores with a different graph representation. The dimensions of the balls indicate frequency size. The lowest frequency is for the score of 11, while the highest frequency belongs to the 6 score.

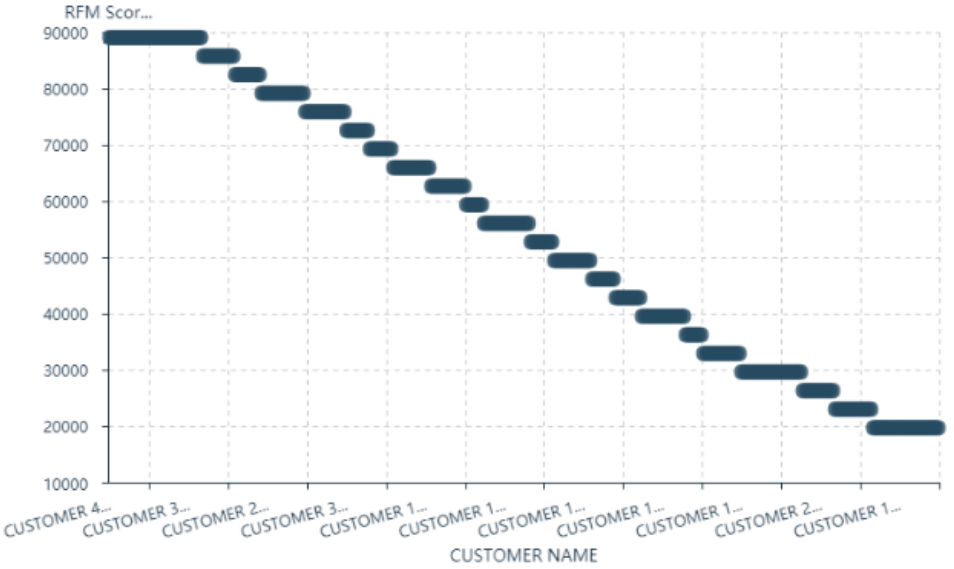


Figure 5. Data Refinety Flow RFM skor chart for Model (3)

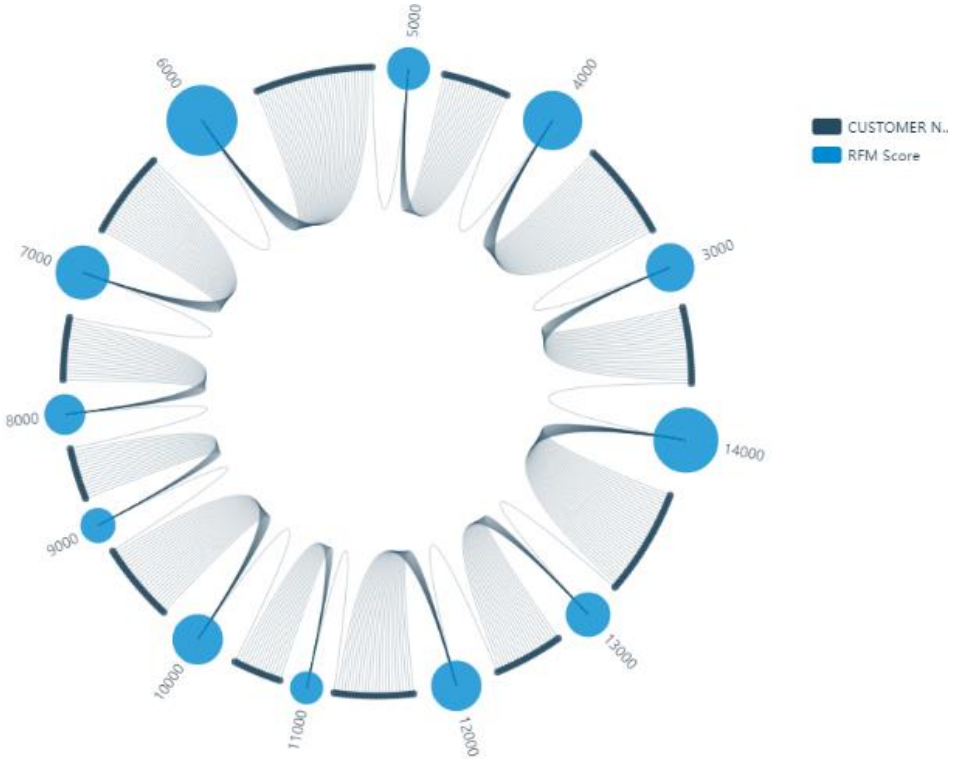


Figure 6. Data Refinety Flow RFM score chart for Model (4)

The score results of the first 8 data according to Model (4) are included in Table 3.

Table 3. Scores for 4 different formulas

CST NAME	RFM SKOR-MODEL1	CST NAME	RFM SKOR-MODEL2	CST NAME	RFM SKOR-MODEL3	CST NAME	RFM SKOR-MODEL4
Customer 43	455	Customer 43	27	Customer 43	89.100	Customer 43	14
Customer 6	455	Customer 6	27	Customer 6	89.100	Customer 6	14
Customer 2	455	Customer 2	27	Customer 2	89.100	Customer 2	14
Customer 5	455	Customer 5	27	Customer 5	89.100	Customer 5	14
Customer 64	454	Customer 64	26	Customer 64	85.80	Customer 64	13
Customer 76	454	Customer 76	26	Customer 76	85.80	Customer 76	13
Customer 44	454	Customer 44	26	Customer 44	85.80	Customer 44	13
Customer 83	454	Customer 83	26	Customer 83	85.80	Customer 83	13

The important thing here is the ranking of the corresponding customers when the points are ranked from large to small. Although the weights of the formulas are different, when we sort them by values in themselves, it is seen that we reach the same result in all 4 formulas and that the segment distributions of customers are the same. Since these 4 formulas are for the same purpose under the logic of RFM analysis, it is normal for the results to be the same, and if the results were different, it would have been said that the formulas were incorrect or not for the same purpose.

7. RESULTS

Data analysis contains important points not on the size of the data, but on the correct, efficient, purposeful use of the data and what to do with the information. Analyzing data is very important so that we can make smart decisions and make future decisions. The summary and importance of data mining is the separation of useful information from the accumulated information and transported to the scanning process.

Analyzing data in ways appropriate to the company's purpose will bring a lot of return to the companies. Today, data analysis has gained more importance and value with the development of Industry 4.0 and the orientation towards more "human" focus. Companies tend to analyze logically with appropriate techniques.

In this study, in which RFM analysis was included in data mining techniques, customers were segmented according to RFM scale based on customers' purchasing habits. As a result, potential customers have the potential to contribute the most to metal waste recycling. In this way, a useful environmental-customer-vendor relationship has been created by gaining discounts as much as the

rate at which customers contribute to recycling. At the same time, sensitivity to the environment has been shown and valued.

REFERENCES

[1] Katie Fisher, (March 19th, 2020), UK Statistics on Waste, UK.

[2] Brian D. James & Jennie M. Moton & Whitney G. Colella, (May 14th, 2013). U.S. Department of Energy's (DOE's) Annual Merit Review and Peer Evaluation Meeting (AMR) for the Hydrogen and Fuel Cell Technologies (FCT) Program, Arlington, Virginia.

[3] Anthony Kasozi, (2016). A Recycle Analysis for Converting Used Oil into Diesel to Support the Power System During the Drilling Process, SPE Nigeria Annual International Conference and Exhibition, 2-4 August, Lagos, Nigeria, ISBN: 978-1-61399-487-0, United States of America.

[4] Internet: Recycle Mania 2009 Hits Otterbein Campus, (2009).https://www.thisweeknews.com/content/ stories/westerville/news/2009/01/21/0122wvrecyclemania_ln.html

[5] Internet: Olipot in Barcelona Cooking Oil is Recycled!, (2016).https://www.livingcircular.veolia.com/en/eco-citizen/olipot-barcelona-cooking-oil-recycled

[6] Internet: Recycling Reverse Vending Machine(2014).https://reversevending.wordpress.com/2013/09/16/ the-history-of-reverse-vending-1920-to-2013/

[7] Internet: How Companies Learn Your Secrets(2012).https://www.nytimes.com/2012/02/19/magazine/ shopping-habits.html

[8] Bult. J. R. & Wansbeek. T.. "Optimal selection for direct mail". Marketing Science. Vol. 14. No. 4. (1995) 378-394. ISSN: 0732-2399.

[9] Hosseini, S.M.; Maleki, A. Gholamian, M.R. (2010). Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty, Expert Systems with Applications, Vol. 37, No. 7, (July 2010) 5259-5264, ISSN:0957-4174.

[10] Chuang, H. & Shen, C. (2008). A study on the applications of data mining techniques to enhance customer lifetime value – based on the department store industry, Proceedings of the 7th International Conference on Machine Learning and Cybernetics, pp. 168-173, ISBN:978-1424420964, Kunming, China, July 2008, IEEE.

[11] Ha, S.H. (2007). Applying knowledge engineering techniques to Customer analysis in the service industry, Advenced Engineering Informatics, Vol. 21, No.3, (July 2007) 293-301, ISSN:1474-0346.

[12] Derya Birant (2011). Data Mining Using RFM Analysis, Knowledge-Oriented Applications in Data Mining, Prof. Kimito Funatsu (Ed.), ISBN: 978-953-307-154-1, InTech.

[13] Internet: Yenilik analizi, Sıklık analizi ve Parasal (RFM) analiz ayarlama, (2017). https://docs.microsoft.com/tr-tr/dynamics365/unified-operations/retail/set-up-rfm-analysis

[14] Internet: IBM Watson Studio ile yapay zekanın gücünden istediğiniz ölçekte yararlanın.https://www.ibm.com/tr-tr/marketplace/watson-studio.