Environmental Research & Technology, Vol. 3 (3), pp. 102-112, 2020

Turkish JournalPark ACADEMIC

Environmental Research & Technology

http://dergipark.gov.tr/ert

RESEARCH ARTICLE

Evaluation of waste management using clustering algorithm in megacity Istanbul

Didem Guleryuz[1,*]

[1]Bayburt University, Industrial Engineering Department, Dede Korkut Campus, 69000 Bayburt, TURKIYE

ABSTRACT

Industrialization and urbanization are increasing with the effect of globalization worldwide. The waste management problems are rising with the rising population rate, industrialization, and economic developments in the cities, which turned into environmental problems that directly affect human health. This study aims to examine waste management performance in the districts located in the city of Istanbul. To ensure that the districts are clustered in terms of the similarities and differences base on waste management. On this occasion, the authorized unit managers of the districts in the same cluster will be able to establish similar management policies and make joint decisions regarding waste management. In addition, the division of districts into clusters according to the determining indicators can provide information about the locations of waste storage centers. Also, these clusters will form the basis for the optimization constraints required to design appropriate logistics networks.

Waste management performance of 39 districts in Istanbul in 2019 was compared by taking into consideration domestic waste, medical waste, population, municipal budget, and mechanical sweeping area. The data were obtained from The Istanbul Metropolitan Municipality (IMM) and Turkey Statistical Institute (TURKSTAT). One of the non-hierarchical clustering methods, the K-means clustering method, was applied using IBM SPSS Modeler data mining software to determine the relations between 39 districts. As a result, the waste management performance of the districts was evaluated according to the statistical data, similarities and differences were revealed by using the determined indicators.

Keywords: Waste management, clustering, K-means, data mining

1. INTRODUCTION

With the rapidly increasing population and the increasing amount of waste, waste management has become an essential field of study. Waste storage areas that do not have the standards required in modern landfill facilities cause serious environmental problems. Therefore, determining these areas and efficiently and collecting waste has great importance. In addition, the urbanization and population that increase in parallel with the increase in industrial activities all over the world cause pressure on the environment. Wastes that accumulate more rapidly with increasing consumption trends have reached threatening environmental and human health due to their quantity and harmful content.

In recent days, the decision-making activities of waste management systems have become more prominent, as recycling of waste has become essential due to the reduced capacity of waste incinerators. The amount of waste per capita in Turkey is always below the average of European countries between 2009 and 2018. However, when the amount of recycled waste per person is examined between the years 2016-2018 (Turkey has registered 3-year data), the average of Turkey is falling far below the average of Europe, as shown in Fig 1 (a-b). The graphs show the need to pay attention to Turkey's waste management and recycling policies

When the studies on waste management are investigated, waste management can be defined as "minimization of domestic, medical, hazardous and non-hazardous waste, separate collection at the source, intermediate storage, determination of transfer centers for waste where necessary, transport of waste, recovery, disposal and operation of disposal

Corresponding Author: dguleryuz@bayburt.edu.tr (Didem Guleryuz)
Received 5 June 2020; Received in revised form 2 August 2020; Accepted 3 August 2020
Available Online 17 August 2020
Doi: https://doi.org/10.35208/ert.764363

facilities, maintenance, monitoring and control processes" [1]. The goal of waste management is to ensure less waste generation, to collect waste, to recycle waste and to eliminate it without harming the environment. Completing these steps will be possible with interdisciplinary approaches efficiently.

Waste services in Turkey are carried out by local governments. For this reason, each municipality proceeds in line with its strategies so that policies implemented for waste management is essential in the megacity Istanbul is Turkey's most populous city. When all regions and Istanbul are examined, it is seen that the amount of waste collected in Istanbul is strikingly higher than in other regions, as seen in Fig 2. For this reason, in this study, waste management in Istanbul is analyzed on a district basis; similarities and differences in the waste management of the districts were tried to be found by the clustering analysis method.

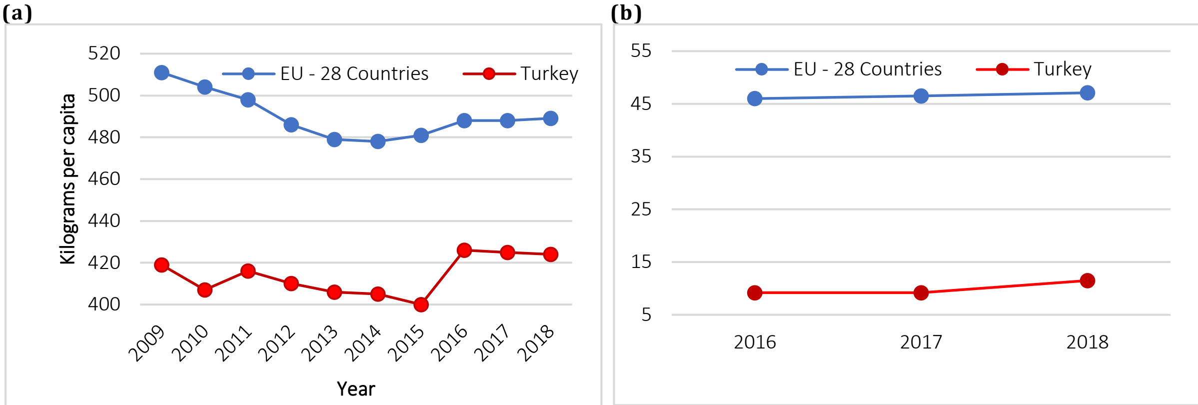

Fig 1. Comparison of waste generation (kilogram per capita; (a) Waste Generation (kilogram per capita), (b) Waste Recycling (kilogram per capita)

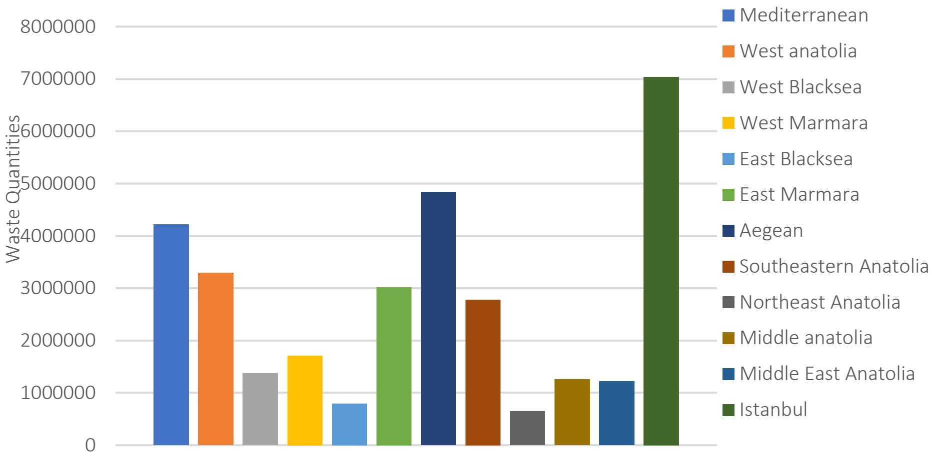

Fig 2. Comparison of waste quantities in Turkey

Domestic waste generated daily in Istanbul with 39 districts is 18,000 tons [2]. Progress in line with 18,000 tons of waste collection and waste management objectives, it is possible with the unification of the districts in a common framework. Thirty-nine districts of Istanbul were clustered using the data mining algorithms, namely the k-means method via five indicators taking into consideration the similarities and differences.

Other applications of the clustering for waste management focus the relationships between the indicators were determined and optimizing waste collection and recycling in previous studies. This study aims to determine the clusters of districts and according to these clusters, waste management policies can be made for differences and similarities for the districts.

There are many studies on waste management in the literature. Management models [3], multi-criteria decision-making methods [4], mathematical models [5,6], and data mining applications [7,8] are some of these studies. Table 1 summarizes the studies which use data mining methods for management cases in the past.

Agovino et al. [1] analyzed the waste management process based on the amount of waste in waste storage areas and they made suggestions to improve waste management activities. Cluster analysis was applied to 103 Italian provinces. As a result of the study, it has been found that the waste disposal rate has a dual structure, and activities that do not directly affect the quality of the institution and the environment are the main factors in the waste management process [1].

Table 1. Previous studies on clustering

Reference	Year	Area	Methods
Agovino et al. [1]	2016	Waste management	Spatial correlation processes and spatial clusters
Sharma et al. [9]	2020	Waste management	K-means clustering
Shi et al. [10]	2014	Environmental risk	K-means clustering
Ecer and Aktas [11]	2019	Healthcare analyzing	K-means clustering
Dorn et al. [12]	2012	Waste disposal facilities	QFD and clustering
Otoo et al. [13]	2014	Waste management	Capacitated clustering
Lin et al. [14]	2011	Food Waste management	K-means clustering
Parfitt et al. [15]	2001	Waste management	Hierarchical clustering
You et al. [16]	2017	Solid Waste management	ANN, ANFIS, SVM, and RF
Niska and Serkkola [17]	2018	Waste management	SOM and K-means clustering
Márquez et al. [18]	2008	Waste management	Clustering, Classification, decision tree
Song et al. [19]	2014	Waste management	Entropy and Spatial Clustering
Caruso and Gattone [20]	2019	Waste management	Unsupervised Classification

Sharma et al. [9] worked on waste management with the K-means method. As a result of their studies, it was expected to facilitate the decision-making process via k-means. For this reason, they made clustering with the solid waste data set, considering the indicators such as land use, financial costs, labor force needs [9].

Otoo et al. [13] have developed an optimization model for logistics and disposal of waste, which is vital for waste management. In this study for the Kumasi region of Ghana, two methods, clustering, and heuristic optimization, were used. In the optimization model, cost and waste transport distances are used as variables. When the results of the study are compared with the existing schedules, the weekly distance decreased by 40% [13].

Lin et al. [14] used questionnaires as a method of gathering data sets. The multivariate factor analysis and clustering were used to analyze the results they obtained from the questionnaires. As a result of clustering and factor analysis applications, a more robust decision-making process was aimed to design. They used SWOT analysis to evaluate the results of the two methods. The optimal waste management system was selected [14].

Parfitt et al. [15] proposed a system for accumulating waste and recycling waste to increase the efficiency of local governments in England and Wales. In this system, which is based on hierarchical cluster analysis, related regions are clustered and compared with the existing system. The cluster analysis results indicated that different waste management practices could be used for regular household waste collection [15].

Niska and Serkkola [17] have developed a system that stores information for waste management using the Self-Organizing Map (SOM) and the k-average algorithm. The results showed the potential of an advanced analytical approach to analyze waste management procedures further. Cluster analysis is recommended for planning and optimizing waste collection and recycling [17].

Márquez et al. [18] proposed a management strategy using data mining methods to manage household waste. In the analysis, household waste data from the settlements in Mexicali were used. K-means cluster analysis was applied with socio-economic indicators, and decision tree application was made with clustered data. As a result of their study, the relationships between the indicators were determined [18].

This study aims to cluster waste management practices in the districts of Istanbul by using the K-means clustering algorithm that is a well-known algorithm among data mining methods. Examining the additive waste management performance in the districts of Istanbul and clustering the districts by considering the similarities and differences for waste management. On this occasion, managers authorized to make decisions on waste management will be able to establish similar management policies and make joint decisions on solid waste management.

The following parts of the study are data collection process, explanations of data mining methods and k-means clustering method, determination of the number of clusters, implementation of k-means clustering via IBM modeler using waste management data for Istanbul, identifying of the cluster for districts of the cluster of districts according to results of k-means clustering methods and results and discussion for the case study.

2. MATERIALS AND METHODS

2.1. Data collection process

The waste management performance of 39 districts of Istanbul in the year 2019 was analyzed according to base on the domestic waste amount, population, municipal budget, medical waste amount and mechanical sweeping area variables. As shown in Table 2, data were collected from different data sources. The data set is presented in Table 3 used for the analysis.

Table 2. Information about the variables

Variables	Unit	Sources
Domestic Waste amount (DW)	Ton	IMM[a][2]
Population (PO)	Capita	TURKSTAT[b][21]
Municipal Budget (MB)	Million TRY	IMM[2]
Medical Waste amount (MW)	Ton	IMM[2]
Mechanical Sweeping area (MS)	Square meter	IMM[2]

[a]The Istanbul Metropolitan Municipality
[b]Turkish Statistical Institute

It is essential to normalize the data to make more meaningful model comparisons in data mining applications [22]. The normalization standardization method was used, and the normalization formula can be seen in Eq. 1.

$$z_i = \frac{x_{i-\bar{x}}}{S_x} \qquad (1)$$

where x presents mean and standard deviation of the related variable is shown via S_x in the dataset. The dataset was represented in Table 3 for the clustering analysis.

The statistical significance of the normalized dataset can be seen in Table 4.

2.2. Data mining and clustering

The rapidly growing information pool with the developing technology has made it necessary to work on big data. It is a complicated process to distinguish useful information from big data. For this reason, data mining is processed by automatic or semi-automatic methods in order to analyze large amounts of data and make meaningful results and reaching meaningful results. The most important disciplines of those interested in data mining are Machine learning and artificial intelligence, so developments in these two areas are also significant for data mining. Also, Big data is encountered every day in many areas such as meteorology, complex physics simulations, environmental research, and health services. Therefore, traditional data processing methods cannot respond to Big Data complexity. Especially in many areas, it is necessary to continuously conduct extensive and real-time queries on many unstructured or structured datasets. This demand led to the development of search and sorting technologies to obtain the necessary information from big data [23].

Wu et al. [24] showed that C4.5, k-Means, SVM, Apriori, EM, PageRank, AdaBoost, kNN, Naive Bayes, and CART methods are the top 10 data mining methods [24]. Briefly, clustering is the classification of the observations into groups without supervision. Therefore, the clustering algorithm plays a vital role in a wide variety of real-life applications with its multi-disciplinary application structure.

Clustering algorithms are generally divided into two as hierarchical and non-hierarchical. Two methods are used in hierarchical clustering methods. The first of these methods accept each variable as a cluster initially and continues with iterations that combine the clusters according to their similarities (based on a specified distance measure. For instance, Euclid, Manhattan, Minkowski Distances), so the number of clusters decreases every step. Various visual methods can demonstrate the cluster structure obtained as a result of iterations, such as dendrogram and tree diagrams. The most commonly used algorithms of hierarchical clustering methods in the literature are Single Linkage, Nearest Neighbor, Ward, Centroid, Lance & Williams methods [25]. The most significant disadvantage of hierarchical methods is that it is challenging to decide the proper number of clusters needed to solve the problem.

Clustering is one of the most extensive data analysis techniques applied to gain knowledge about the structure of the data. Although the data in different clusters have different properties and data in the same subgroup have very similar statistical properties, it can also be defined as the task of identifying subgroups in data. In this study, the k-means algorithm, which is accepted as one of the most used clustering algorithms, will be used for clustering 39 districts because of its ease of application and excellent results.

2.3. K-Means Clustering Algorithm

Suppose x = (x_1, x_2, ...,x_n) is the dataset of observed values. The clustering method aims to split the dataset into K sub-groups, considering the clustering criterion. There are several clustering methods, and the sum of the squared Euclidean distances between each variable is one of the most commonly used clustering criteria. This criterion is known as cluster error and bases on cluster centers. The cluster error formula represents in Eq. 2.

$$E(m_1, m_2, \ldots, m_M) = \sum_{i=1}^{N} \sum_{k=1}^{M} I(x_i \in C_k) \, \|x_i - m_k\|^2 \qquad (2)$$

where if x is true I(X) = 1 and otherwise I(X) = 0.

Where x_i represents each data point, c_k is cluster-k, and the center of the cluster is denoted by m_k. The K-means algorithm determines the most suitable results locally regarding cluster error. In many clustering applications, it is a fast-iterative algorithm that is employed. In addition, it is also a point-based clustering method that initially begins with cluster centers placed at random locations and continues with each step centered by the cluster to minimize cluster error. The primary drawback of the method is that it is sensitive to the starting point since it is based on the initial positions of the cluster centers. Therefore, to obtain the most suitable solutions using the k-means algorithm, several iterations should be done [26].

Table 3. DW, PO, MB, MW and MS values for the 39 districts of Istanbul (2019)

District	Domestic Waste	Population	Municipal Budget	Medical Waste	Mechanical Sweeping
Adalar	16718	15238	41	3	9082500
Arnavutkoy	93010	282488	311	127	87659454
Atasehir	174355	425094	518	947	48441360
Avcilar	155042	448882	325	379	54300948
Bahcelievler	212956	611059	426	1250	29185767
Bagcılar	278547	745125	525,4	1863	656353,9512
Bakirkoy	121614	229239	481	966	102057107
Basaksehir	208181	460259	540	126	68632635
Bayrampasa	124328	274735	301	495	24971001
Besiktas	123926	182649	436	616	84754551
Beykoz	123766	248260	465	282	59293920
Beylikduzu	113246	352412	481	533	59448773
Beyoglu	133928	233323	325	248	142649721
Büyükcekmece	108522	254103	409	146	33305508
Catalca	29868	73718	88	38	18678348
Cekmeköy	97751	264508	270	81	27396912
Esenler	140148	450344	375	320	64918291,5
Esenyurt	356789	954579	900	635	73692424
Eyup	148273	400513	380	109	200202139,5
Fatih	234880	443090	391,8	2562	204703315
Gaziosmanpasa	154480	491962	379	1244	25908042
Gungoren	111236	289441	245	188	14957499
Kadıkoy	209382	482713	670	1502	75382581
Kagithane	156949	448025	370	295	60961389
Kartal	160725	470676	615	1276	96889383
Kucukcekmece	322731	792821	650	1420	88936494
Maltepe	171185	513316	488,4	883	123889441
Pendik	233929	711894	610	1546	141648519
Sancaktepe	142699	436733	466	238	34515090
Sarıyer	164783	347214	421,9	988	67267368
Silivri	86341	193680	254,5	224	8582196
Sultanbeyli	120453	336021	313	286	27600975
Sultangazi	178280	534565	435	361	89081175
Sile	27487	37692	85	13	14128357,5
Sisli	157137	279817	670	2599	116745879
Tuzla	122584	267400	325,6	412	87903284
Umraniye	258042	710280	550	1017	81827844
Uskudar	222645	531825	650	2023	88358349
Zeytinburnu	139747	293574	505	822	103765050

Table 4. The statistical information of normalized data set

	DW	PO	MB	MW	MS
Mean	-1,6E-16	6,83E-17	3,2E-16	2,03E-16	-7,7E-17
Std. Dev.	1	1	1	1	1
Min	-1,97	-1,89	-2,28	-1,08	-1,45
Max	2,80	2,76	2,77	2,71	2,80
Skewness	0,62	0,56	0,06	1,17	0,95
Kurtosis	1,03	0,51	0,85	0,76	1,08

2.4. Determination of k

In order to specify the number of clusters, there are various methods are used in the literature. The most common methods are The Elbow Method and The Silhouette Method. To identify the optimal number of clusters, the Elbow method is the best-known method among them. Within-Cluster-Sum of Squared Errors (WCSS) value is based on the different values that k can take. The number of clusters is specified by selecting the k value where the WCSS begins to decrease. On the WCSS-versus-k chart, this situation is an elbow shape.

The method starts with k = 2, the number of clusters is increased by 1, and each step calculates the amount of error. For a one-unit change for k, the point at which the amount of error is dramatically reduced is determined. The error is recalculated by changing the number of clusters. If the error continues stably or increases in the next change, the value of k with a dramatic decline is determined as the number of clusters [27]. As seen in Fig 3 the Elbow method was applied, and the number of clusters was determined as 5. Calculations for the Elbow method was performed via Python.

Fig 3. Diagram of the Elbow method

As seen in Fig 3, the error of k = 2 is 1.52. The error is 1.22 when the number of clusters is raised to 4. According to the result of the elbow method, the number of clusters was determined as 4. As shown in Table 5, the most considerable value belongs to 4 clusters with 0.4808.

Table 5. The results of Silhouette method

k	2	3	4	5
Silhouette	0.4291	0.4342	**0.4808**	0.4581

3. RESULTS AND DISCUSSION

3.1. Results of K-Means

Thirty-nine districts of Istanbul are clustered by using k-means method for the optimization of waste management activities, considering the variables of DW, PO, MB, MW and MS. All clustering analyses were performed via IBM SPSS Modeler. Since the number of districts owned by each cluster is different, the number of districts included in the clusters as a percentage and map display of clustered districts are given in Fig 4.

The obtained clusters can be seen as a 3-dimensional in Fig 5.

As a result of the analysis, obtained clusters have different statistical properties. Cluster-1 comprises nine districts, cluster-2 has six districts, cluster-3 has four districts and cluster fourth includes 20 districts. The clusters can be seen in Table 6.

The significance levels of clusters by the k-means method are shown in Table 7. Variables with significance levels above 0.90 are significant on the cluster in IBM SPSS Modeler. According to Table 7, it can be concluded that the effects of all variables on four clusters are significant.

3.2. Statistical Evaluation

In determining the path to be followed in waste management, it will be easy to include the relevant variables in the system and separate the problem into sub-problems. Therefore, the variables determined for cluster analysis are essential. The five variables selected for waste management determined in this study are DW, PO, MB, MW, and MS values. The predictor importance values of these variables can be seen in Fig 6.

Table 8 presents descriptive statistics of clusters based on observed data for all variables.

When domestic waste average is analyzed, it is seen that cluster-1 consisting of 9 districts has the least amount. Although cluster - 4 covers 20 districts, the municipal budget does not have the highest average. The cluster with the highest municipal budget is cluster-2, consisting of 6 districts. Considering the amount of medical waste, the cluster with the highest amount of medical waste is cluster-3, covering four districts. The most mechanical sweeping area belongs to cluster-4. Districts in cluster-2 are in cooperation with medical waste management and can apply standard rules. Likewise, counties located in cluster-1 can determine standard policies for domestic waste management. Clustering the districts according to the variables associated with waste management will be useful in the province of Istanbul regarding zero waste, which is also among the goals of sustainable development. Recently, studies on reducing environmental pollution from supply chains have increased in the literature, considering the sustainability goals. Supply chain planning for municipalities will also be more easy than the k-means cluster analysis results made in this study [28].

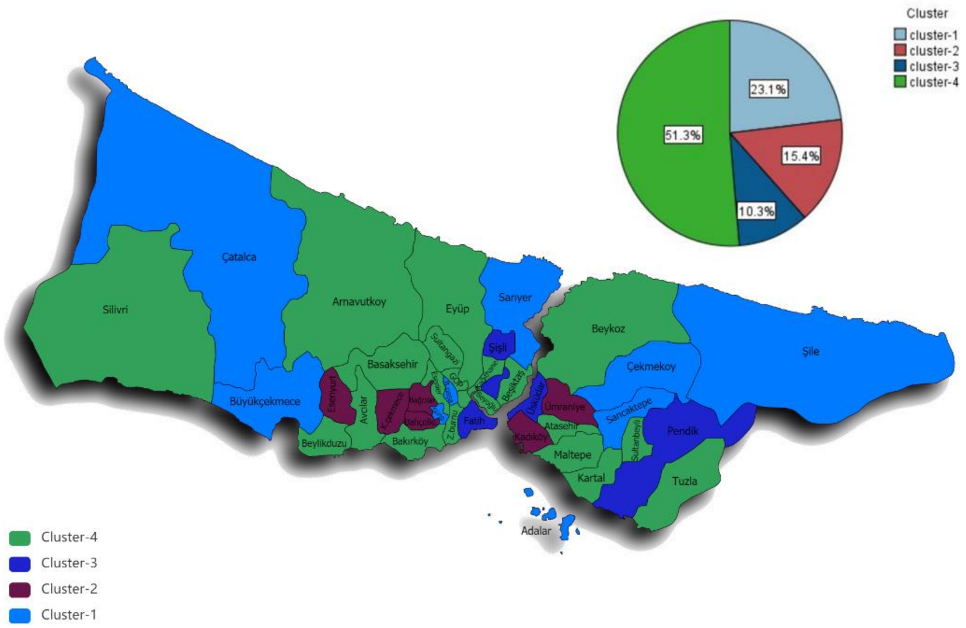

Fig 4. Cluster sizes and map display of clustered districts

Table 6. Clusters and distances from the centers

Cluster-1 (9)		Cluster-2 (6)		Cluster-3 (4)		Cluster-4 (20)	
District	**Distance**	**District**	**Distance**	**District**	**Distance**	**District**	**Distance**
Adalar	0.350	Bahcelievler	0.340	Fatih	0.429	Arnavutkoy	0.287
Bayrampasa	0.222	Bagcılar	0.379	Pendik	0.348	Atasehir	0.266
Büyükcekmece	0.250	Esenyurt	0.547	Sisli	0.353	Avcilar	0.217
Catalca	0.255	Kadıkoy	0.338	Uskudar	0.268	Bakirkoy	0.257
Cekmeköy	0.117	Kucukcekmece	0.234			Basaksehir	0.299
Gungoren	0.142	Umraniye	0.180			Besiktas	0.217
Sancaktepe	0.279					Beykoz	0.222
Sarıyer	0.193					Beylikduzu	0.164
Sile	0.285					Beyoglu	0.373
						Esenler	0.166
						Eyup	0.603
						Gaziosmanpasa	0.410
						Kagithane	0.185
						Kartal	0.372
						Maltepe	0.294
						Sancaktepe	0.279
						Sarıyer	0.193
						Sultangazi	0.210
						Tuzla	0.193
						Zeytinburnu	0.189

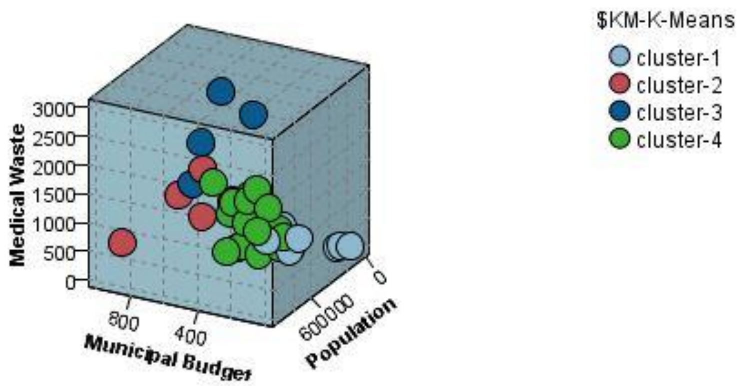

Fig 5. Clusters of districts

Table 7. The normalized mean values of variables for each cluster

Variables	Cluster-1 (9)	Cluster-2 (6)	Cluster-3 (4)	Cluster-4 (20)	Importance
Domestic Waste	-1.081	1.623	0.769	-0.154	★1.0000 Important
Population	-1.013	1.575	0.464	-0.109	★1.0000 Important
Municipal Budget	-1.205	1.130	0.896	0.024	★1.0000 Important
Medical Waste	-0.849	0.782	2.098	-0.272	★1.0000 Important
Mechanical Sweeping	-1.053	-0.251	1.410	0.267	★1.0000 Important

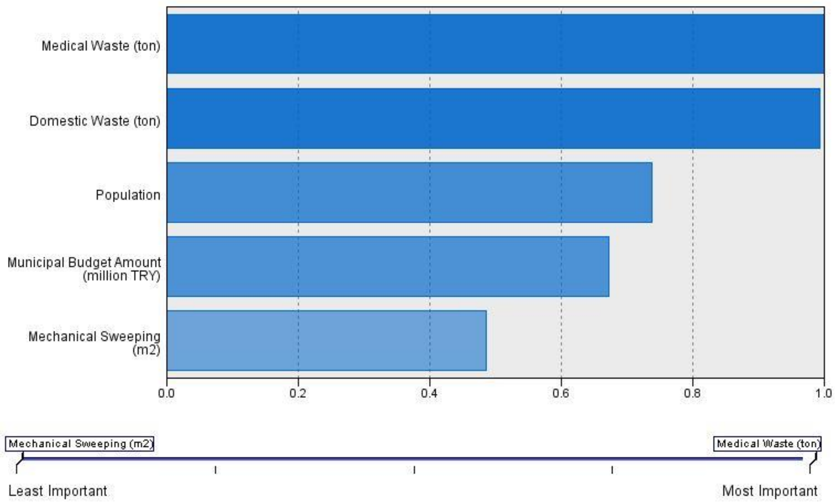

Fig 6. Predictor importance of variables

The k-means clustering is relatively simple to implement to find similarities and differences of the districts that scale to extensive waste data and guarantees convergence. The data mining process allows municipalities to collect useful information they can use in waste management. The data can be analyzed from several different perspectives to provide valuable information that can reduce waste management costs. With this application, the relationships and patterns between variables determined on waste management and data were analyzed. The statistical results found can be used in decision making for administrative activities. The data related to waste management was provided to be "analyzed in detail," and more information was obtained from the waste management data in the archive by using the k-means clustering method. The relationships between external factors such as internal factors and cost factors, personnel skills and demographic characteristics can be examined. For example, using one of the data mining clustering methods for waste management can help identify subgroups with different characteristics in the district of waste management. Variables analyzed by clustering can have a significant impact on internal processes and citizen satisfaction.

Table 8. Statistics of clusters for variables

Cluster	Mean	Standard Deviation	Standard Error
Cluster-1 (9 Districs)			
DW	80300.44	43330.92	14443.64
PO	193237.33	120016.41	40005.47
MB	222.94	123.92	41.31
MW	163.70	158.172	52.724
MS	19855921.83	8840655.49	2946885.163
Cluster-2 (6 Districs)			
DW	273074.5	58992.84	24083.73
PO	716096.2	160709.54	65609.39
MB	620.233	163.27	66.65
MW	1281.12	423.138	172.745
MS	58280243.99	35188980.23	14365841.023
Cluster-3 (4 Districs)			
DW	212147.75	37092.47	18546.23
PO	491656.5	180141.4	90070.7
MB	580.45	128.217	64.108
MW	2182.40	499.42	249.71
MS	137864015.5	49593686.18	24796843.094
Cluster-4 (20 Districts)			
DW	146346.050	26584.57	5944.49
PO	375846.4	107739.83	24091.36
MB	432.14	82.62	18.5
MW	558.8	379.34	84.82
MS	83126956	39769453.6	8892720.19

4. CONCLUSIONS

Due to globalization causes an enormous amount of consumption and that induce to increase the amount of waste by leaps. In terms of environmental resources, waste is one of the main issues that need to be taken before it reaches dangerous levels. In many parts of the world, academic studies are carried out for the solution to environmental problems. Hence, reducing the environmental damage of waste and recycling should be the primary target of all countries. Also, three of the seventeen targets determined for sustainable development are directly related to waste management. Therefore, well-planned waste management decisions will directly contribute to sustainable development. For this reason, waste management data is carefully recorded in European countries. Waste management differentiated according to the structural and geographical features of the country that are carried out by municipalities in Turkey.

In this study, Turkey's most populous city of Istanbul, which has 39 districts, is divided into clusters using Domestic waste, medical waste, population, municipal budget, and Mechanical Sweeping area. The data for the variables were obtained from IMM for 2019. In order to divide the districts into clusters, the k-means clustering method, which is the most familiar explorative data analysis technique in data mining, was used. In the first step, the data is normalized to state the number of clusters. Then, the elbow method and the silhouette method calculations, which are frequently used in the literature, were performed to specify the number of clusters. According to these calculations, the number of clusters was determined as 4. Thirty-nine districts are distributed as cluster-1 involves nine districts, cluster-2 comprises six districts, cluster-3 has four districts, and cluster-4 contains 20 districts. Based on statistics, it was concluded that all variables were significantly affected on all four clusters. As a result, it has been observed that there are significant differences in the clusters of the districts obtained by using domestic waste, medical waste, population, municipal budget and Mechanical Sweeping area variables.

For future research, an extensive database can be used. Other indicators that are important according to the regional conditions can be included in the model as a variable. Different clustering algorithms in the literature such as Mean-Shift Clustering, Gaussian Mixture Model, Agglomerative Hierarchical Clustering would be used to compare the clustering results.

Besides the practical results, this study contributed to the existing literature by creating clusters for districts in Istanbul for waste management in order to develop the necessary policies and to reduce costs and environmental impact in waste management activities. In addition, supportive policies can assist in carrying out waste management activities

REFERENCES

[1]. M. Agovino, M. Ferrara and A., "Garofalo An exploratory analysis on waste management in Italy: A focus on waste disposed in landfill," *Land Use Policy*, Vol. 57, pp. 669–81, 2016.

[2]. IMM. IMM open data portal. 2020, Available: https://data.ibb.gov.tr/, (accessed 13 June 2020).

[3]. D.C. Wilson, "Development drivers for waste management," *Waste Management Research*, Vol. 25, pp. 198-207, 2007.

[4]. S. Cheng, C.W. Chan and G.H. Huang "Using multiple criteria decision analysis for supporting decisions of solid waste management," *Journal of Environmental Science and Health*, Part A, Vol. 37, pp. 975-990, 2002.

[5]. S. Rathi, "Alternative approaches for better municipal solid waste management in Mumbai, India," *Waste Management*, Vol. 26, pp. 1192-1200, 2006.

[6]. S.O. Benítez, G. Lozano-Olvera, R.A. Morelos and C.A. de Vega, "Mathematical modeling to predict residential solid waste generation," *Waste Management*,Vol. 28, pp. 7-13, 2008.

[7]. D.Q. Zhang, S.K. Tan and R.M. Gersberg, "Municipal solid waste management in China: Status, problems and challenges," *Journal of Environmental Management*, Vol. 91, pp. 1623-1633, 2010.

[8]. L.A. Manaf, M.A.A. Samah and N.I.M. Zukki, "Municipal solid waste management in Malaysia: Practices and challenges," *Waste Management*, Vol. 29, pp. 2902-2906, 2009.

[9]. N. Sharma, R. Litoriya and A. Sharma, "Application and Analysis of K-Means Algorithms on a Decision Support Framework for Municipal Solid," *International Conference on Advanced Machine Learning Technologies and Applications*, pp. 267–276, 2020.

[10]. W. Shi and W. Zeng, "Application of k-means clustering to environmental risk zoning of the chemical industrial area," *Frontiers of Environmental Science & Engineering*, Vol. 8, pp. 117-127, 2014.

[11]. B. Ecer, A. Aktaş, "Clustering of European Countries in terms of Healthcare Indicators," *International Journal of Computational and Experimental Science and Engineering,* Vol. 5, pp. 23-26, 2019.

[12]. T. Dorn, M. Nelles, S. Flamme and C. Jinming, "Waste disposal technology transfer matching requirement clusters for waste disposal facilities in China," Waste Management, Vol. 32, pp. 2177-2184, 2012.

[13]. D. Otoo, S.K. Amponsah and C. Sebil, "Capacitated clustering and collection of solid waste in kwadaso estate Clustering of European Countries in terms of Healthcare Indicators, " *Journal of Asian Scientific Research*, Vol. 4(8), pp. 460-472, 2014.

[14]. C. Lin, E.M.Y. Wu, C.N. Lee and S.L. Kuo, "Multivariate statistical factor and cluster analyses for selecting food waste optimal recycling methods," *Environmental Engineering Science*, Vol. 28, pp. 349-356, 2011.

[15]. J.P. Parfitt, A.A. Lovett and G.A. Sünnenberg, "A classification of local authority waste collection and recycling strategies in England and Wales," *Resources, Conservation and Recycling*, Vol. 32, pp. 239-257, 2001.

[16]. H. You, Z. Ma, Y. Tang, Y. Wang, J. Yan, M. Ni, K. Cen and Q. Huang, "Comparison of ANN (MLP), ANFIS, SVM, and RF models for the online classification of heating value of burning municipal solid waste in circulating fluidized bed incinerators," *Waste Management*, Vol. 68, pp. 186-197, 2017.

[17]. H. Niska and A. Serkkola, "Data analytics approach to create waste generation profiles for waste management and collection," *Waste Management*, Vol. 77, pp. 477-485, 2018.

[18]. M.Y. Márquez, S. Ojeda and H. Hidalgo, "Identification of behavior patterns in household solid waste generation in Mexicali's city: Study case," *Resources, Conservation and Recycling,* Vol. 52, pp. 1299-1306, 2008.

[19]. J. Song, Y. Liao, J. He, J. Yang and B. Xiang, "Analyzing complexity of municipal solid waste stations using approximate entropy and spatial clustering," *Journal of Applied Science and Engineering*, Vol. 17(2), pp. 185-192, 2014.

[20]. G. Caruso and S.A. Gattone, "Waste management analysis in developing countries through unsupervised classification of mixed data," Social Sciences, Vol.8, pp. 186, 2019.

[21]. TURKSTAT. Turkish Statistical Institute 2020. Available: http://www.tuik.gov.tr/Start.do, (accessed June 13, 2020).

[22]. M.M. Rahman, Y. Ghasemi, E. Suley, Y. Zhou, S. Wang and J. Rogers, "Machine Learning Based Computer Aided Diagnosis of Breast Cancer Utilizing Anthropometric and Clinical Features," IRBM, https://doi.org/10.1016/j.irbm.2020.05.005, 2020.

[23]. R. Sowmya and K.R. Suneetha, "Data Mining with Big Data," IEEE, pp. 246–250, https://doi.org/10.1109/ISCO.2017.7855990, 2017.

[24]. X. Wu, V. Kumar, Q.J. Ross, J. Ghosh, Q. Yang, H. Motoda, et al. "Top 10 algorithms in data mining," *Knowledge and Information Systems*, Vol. 14. PP. 1-37, 2008.

[25]. P.A. Berkhin, "Survey of Clustering Data Mining Techniques," *Grouping Multidimensional Data*, 2006. In: Editors. J. Kogan, C. Nicholas, M.

Teboulle, Springer, Berlin, Heidelberg, pp. 25-71, 2006.

[26]. A. Likas, N.J. Vlassis, J. Verbeek, "The global k-means clustering algorithm," *Pattern Recognition*, Vol. 36, pp. 451-461, 2003.

[27]. B. Purnima and K. Arvind, "EBK-Means: A Clustering Technique based on Elbow Method and K-Means in WSN," *International Journal of Computer Applications*, Vol. 105, pp. 17-24, 2014.

[28]. M.A. Yerlikaya, B. Efe, O.F. Efe, "Çevresel Atık Kriteri Temelli Tedarikçi Seçim Problemi," *The International New Issues in Social Sciences*, Vol. 5, pp. 311-322, 2017.