

Vladislav Totskiy
Master student,
Economic Cybernetics Department,
Taras Shevchenko National University of Kyiv,
Kyiv, Ukraine
totskiy@knu.ua

K-MEANS CLUSTERING OF ONLINE DELIVERY SERVICE USERS IN A METROPOLIS

Highlights

- Used the capabilities of Python in data preprocessing and preparation for modeling
- Elbow method is more effective than dendrogram
- Formation of the meaning of each cluster formed by k-means centroids
- Euclidean distance from centroids to actual data

Abstract

This work is a review of clustering methods in which clusters were identified using graphical dendrogram methods and the stone placer method (elbow method). The data on which the k-means model was tested was collected from a survey of people who use the online delivery services in metropolis.

Key words: customer clustering, hierarchical method, k-means, e-commerce, data science

Introduction

Several studies in the field of electronic commerce in developed countries have highlighted the impact of various barriers to shaping diffusion. The factors determining the spread of electronic commerce can be of a technological, economic and legal nature. However, over time, while some organizations that operate under the influence of e-commerce cannot actually abandon this form of marketing, other companies continue to use traditional distribution channels.

Since e-commerce provides a more convenient and efficient way to collect information from customers, it is more and more popular, especially when a company can no longer do without contextual advertising.

Also, customer data provides an opportunity to analyze this data for the benefit of other larger marketing campaigns. Also, e-commerce has recently begun to develop more and more due to the conditions of the Covid-19 pandemic. The demand for online ordering of goods as well as delivery services has increased significantly.

In this work, a model was created that defines data into clusters using the k-means method. The data on the basis of which clusters were allocated, which are based on data on the marital status of clients, location, and special preferences.

The task is to determine the segmentation of the market, in different aspects. This will help in the future to define the type of client and lead more effectively, and determine the wishes using the model.

Literature review

In this work, articles were analyzed that are related to the intelligent analysis of k-means, and the use of this method to group customers in different areas of the market, respectively, to understand how these groups are different from each other, the using of methods to determine a cluster, a description of the meaning of each cluster.

Article "Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster" by M A Syakur, B K Khotimah, E M S Rochman and B D Satoto authors. 8 clusters were used experimentally, which were determined by the method of stony placer (elbow method). A model was created that was trained on the training data and tested on the test data. Also, the errors modeled by different samples were separated to test how effective the method is and to determine the exact number of clusters, since the data may not quite correctly show obvious clusters in Euclidean space. And checking for outliers in the data.

In this article "Application of K-Means Algorithm for Efficient Customer Segmentation: A Strategy for Targeted Customer Services". The problem was posed about great competition, and the possibility of maximum competitiveness. Moreover, it is important for each enterprise to understand customer preferences in order to create a more effective strategy whose task will be to globally raise sales at the level of a particular market, since each section of the market has its own nature, comparison of which can help introduce statistical mathematical methods to determine many natural things.

Băcilă Mihai-Florin Babeș-Bolyai , Rădulescu Adrian , Ioan Mărar , Liviu emphasizes the relevance of Data Science method (exact Clustering methods). Since the business not only wants to expand its client base, but also to retain current customers, and optimize offers in the same contextual advertising and other strategic marketing and not only campaigns. The author also claims that this clustering method copes well with a large data set, since this work was based on the development of a BI system for a telecom enterprise. it was found out that only subscribers in the 3rd and 4th clusters use their credit for all types of services.

Purpose of own research

Having data on the preferences of food delivery orders, segment customers in order to make further management decisions based on the conclusion.

Hypothesis: people who are self-employed eat irregularly.

The challenge is to cluster data about customer preferences in ordering food delivery.

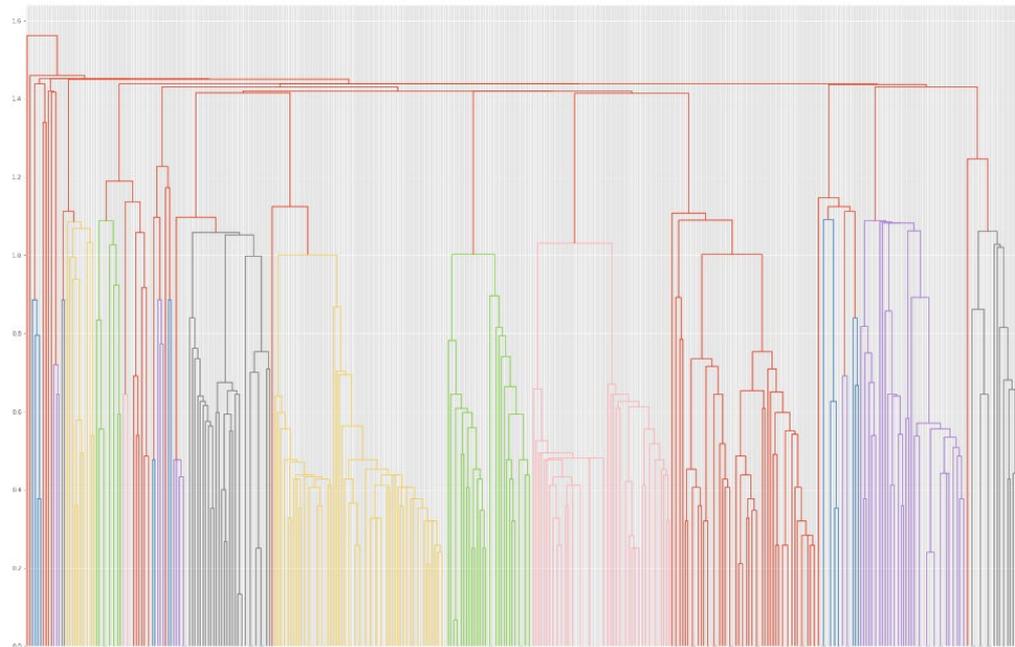
Results

First of all, we will select the data from the dataset that we need in clustering.

In this work, I took the data for clustering: Age, Sex, Marital status, Employment, Family size, Preferred meal and other parameters that relate to service, and what things in service are important for these people. scale (from agree to disagree 1-5). The data was normalized before identifying the clusters.

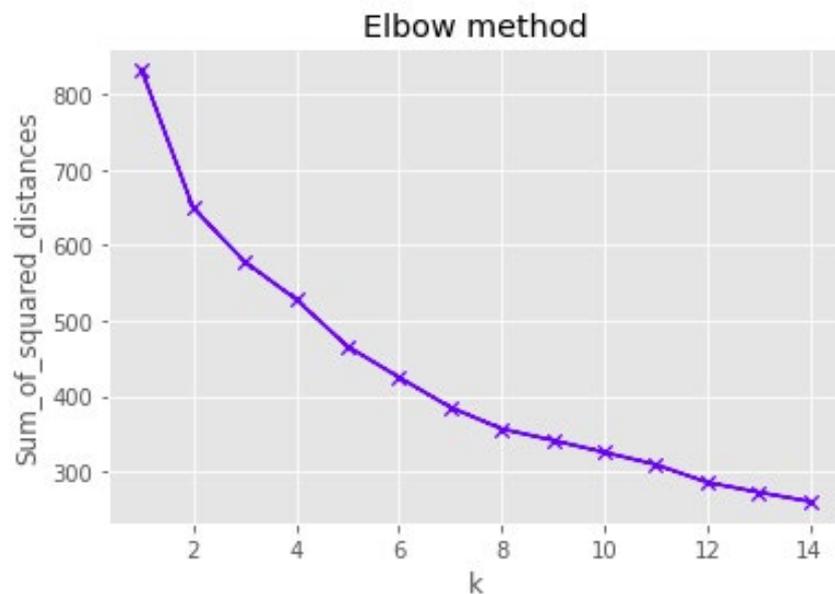
Hierarchical analysis and k -means were performed.

The dendrogram showed quite a lot of clusters, which raises doubts about their correctness, and these clusters were also checked and we can say that these clusters do not have clear information, therefore it was replaced with k-means. And the definition of clusters using the stone placer method or as it is also called the elbow method.



Graph 1. Dendrogram

Starting with the k-means method, we calculated centroids and Euclidean distances from points in vector space. On the basis of which on graph 2 you can see how many clusters are acceptable to take.



Graph 2. Elbow Method

It was decided to take 8 clusters in the study.

Clients were classified, and one cluster was noticed that has almost all clients who are doing business. It was noticed that they prefer each of the types of meal with different

frequency, there is no exact correspondence with what entrepreneurs prefer specifically breakfast, lunch, snacks, or dinner.

Cluster number 4

	Age	Gender	Marital Status	Student	House wife	Employee	Self Employed	Family size
cluster								
0	22.600000	0.581818	0.090909	1.000000	0.000000	0.000000	0.000000	3.400000
1	22.878788	0.651515	0.015152	1.000000	0.000000	0.000000	0.000000	3.333333
2	26.681159	0.521739	0.608696	0.000000	0.028986	0.971014	0.000000	3.362319
3	26.907692	0.723077	0.615385	0.000000	0.061538	0.615385	0.323077	3.646154
4	26.333333	0.500000	0.633333	0.000000	0.033333	0.000000	0.966667	3.466667
5	23.081081	0.459459	0.027027	0.864865	0.027027	0.108108	0.000000	2.891892
6	23.115385	0.461538	0.000000	1.000000	0.000000	0.000000	0.000000	2.865385

Family size	Lunch	Dinner	Snacks	Breakfast	Easy Payment option	Good Food quality	Missing item	Late Delivery
3.400000	0.000000	0.000000	1.0	0.000000	3.545455	3.381818	2.927273	3.327273
3.333333	1.000000	0.000000	0.0	0.000000	3.924242	3.439394	3.378788	3.378788
3.362319	0.565217	0.304348	0.0	0.130435	3.043478	3.057971	3.086957	3.318841
3.646154	0.000000	0.000000	1.0	0.000000	3.492308	3.815385	2.446154	2.969231
3.466667	0.266667	0.566667	0.0	0.166667	3.800000	3.566667	2.166667	3.300000
2.891892	0.000000	0.000000	0.0	1.000000	3.810811	3.729730	2.783784	2.810811
2.865385	0.000000	1.000000	0.0	0.000000	3.923077	3.673077	3.288462	3.307692

Conclusions

In this paper, it was found that k-means is more appropriate for identifying clusters. In the data on the clusters, it can be seen that the hypothesis that businessmen eat irregularly. The data showed the opposite trend for students who preferred either lunch only, lunch only, or snacks only.

References

1. C. P. Ezenkwu, S. Ozuomba. Application of K-Means Algorithm for Efficient Customer Segmentation: A Strategy for Targeted Customer Services. - (IJARAI) International Journal of Advanced Research in Artificial Intelligence, Vol. 4, No.10, 2015. - Available at: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.736.3182&rep=rep1&type=pdf>

2. L. Chen, F. Wang. Preference-based clustering reviews for augmenting e-commerce recommendation. Knowledge-Based Systems, Volume 50, September 2013, Pages 44-59. - Available at:
<https://www.sciencedirect.com/science/article/abs/pii/S095070511300155X>
3. M. A. Syakur, B. K. Khotimah, E. M. S. Rochman and B. D. Satoto. Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster. IOP Conference Series: Materials Science and Engineering, 2018. - Available at:
<https://iopscience.iop.org/article/10.1088/1757-899X/336/1/012017/meta>
4. M. Aryuni, B. Nusantara, E. D. Madyatmadja, E. Miranda. Customer Segmentation in XYZ Bank Using K-Means and K-Medoids Clustering. Published in: 2018 International Conference on Information Management and Technology (ICIMTech). - Available at:
<https://ieeexplore.ieee.org/abstract/document/8528086>
5. Băcilă Mihai-Florin Babeş-Bolyai, Rădulescu Adrian, Ioan Mărar, Liviu. PREPAID TELECOM CUSTOMERS SEGMENTATION USING THE K-MEAN ALGORITHM. - Available at:
<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.1079.6133&rank=1>