Information System of Catering Selection by Using Clustering Analysis

Nataliya Boyko[0000-0002-6962-9363], Khrystyna Shakhovska[0000-0002-9914-229], Lesia Mochurad[0000-0002-4957-1512], Jaime Campos[0000-0001-7048-8089]

Lviv Polytechnic National University, Lviv79013, Ukraine
Linnaeus University, Växjö, Sweden
nataliya.i.boyko@lpnu.ua,
kristin.shakhovska@gmail.com,
lesiamochurad@gmail.com,
jaime.campos@lnu.se

Abstract. The topic of tourism is up-to-date because everyone wants to broaden their mind. An integral part of all trips is food, the difference is only in price and quality. Moreover, today is popular gastronomic tourism. The features of tourism are explored by using geodata analysis, numerical and categorical values. Catering are grouped by certain criteria. Data analysis using RStudio and Tableau is performed. An information system for catering selection is created. A correlation and regression analysis was performed on the analyzed data. The k-means algorithm for the analyzed nutrition data has been implemented. A system of four clusters for the selection of a catering facility was constructed.

Keywords: regression analysis, clustering, k-means algorithm, system, tourism, nutrition.

1 Introduction

With the development of market economy and integration of the world economy the role and place of the restaurant business were reviewed. Changes in economic development in the country require the application of new approaches to management and organization of activities, which should be aimed at maximizing consumer demand and ensuring a high level of efficiency of their productive and economic activity.

Not all modern selection systems take into account given parameters. Another important factor is the price range of the restaurant and reliable reviews of it. After all, the proposed parameters affect the accuracy of the characteristics of the desired catering [10].

There are many geographic information systems, which in most cases are useful by using them in the process of performing application tasks. Each tourist IS along with the input and output modules contains a tools for performing spatial analysis functions and performing specific user tasks. These tools directly depend on the data models that are supported by a particular tourist IS and used for the user's tasks.

2019 DCSMart Workshop.

Quite often, there is a need to predict new trends in the structure of catering, taking into account a variety of factors - environmental, geopolitical, and situational. For most formalized information solutions to existing IS and are reduced to the execution of ready-made solutions provided by mapping tools. In authors opinion, the tools should cover actions from a simple revision of existing catering facilities and the implementation of auxiliary mapping supplements to the issuance of expert gastronomic recommendations [19].

Consequently, the research area described above is interesting for research because it contains many parameters that can be analyzed. It allows to build appropriate relationships between data to form an effective solution. It is also necessary to group catering facilities according to certain criteria.

2 Review of the Literature

A large number of scientists [15-20] describe well-known clustering algorithms in their work and propose their own methods of its application. But most of them cannot provide an effective estimate, as we have a dynamic location database that can be clearly defined as a knowledge base. The point is that with any known operations with attributes, records, etc., there is no need for checking the relations that were created at the beginning of the database creation. The constant accumulation of dynamic data can lead to the restructuring of the knowledge base structure.

Some researchers [18] propose universal multi-parameter clustering methods. However, they contain limited tools for determining spatial models, since spatial real dimensions contain real indicators.

Some scientists have focused their research on spatial data. For example, various statistical approaches have been included in the technique [2], elements of analytical geometry have been used in the Delaunay triangulation analysis [3, 4], and the method of density change in the distribution of random variables has been applied [7, 8] and so on.

This paper provides well-known clustering methods such as: CLIQUE [11], ENCLUS [12], ORCLUS [13] and DOC [14]. However, the complexity of the method calculations is that they find clusters and fine-tune the properties of certain objects under predefined clustering algorithms. Cluster objects are identified by several subjective parameters of the clustering algorithm. For example, in CLIQUE to perform the analysis, it is necessary to determine the value of the interval ξ and the threshold density τ. For ORCLUS it is necessary to determine in advance the number of clusters k and the dimension of the subspaces p. For a DOC, a certain length w, the density threshold α and an equilibrium coefficient β is required over a specified period. The described parameters are determinative and necessary for the condition of passing the clustering algorithm [1, 16-19].

There is another issue that occurs when you use the specified clustering algorithm. Namely, it is difficult to implement hierarchical clustering in each subspace, because it requires the adaptation of certain applications and models. It is complex for the user of this algorithm to perform many iterations to find the complete data set.

Therefore, the paper proposes the ICEAGE method [5, 17], which is effective for hierarchical spatial clustering (for two-dimensional spatial points). It is interactive and allows you to achieve complexity *O (n log n).*

Problem statement

Nowadays, the topic of tourism is quite relevant, since everyone wants to broaden their minds and satisfy their gastronomic demands. Indeed, an integral part of all travel is a meal, a difference only in price and quality.

The purpose of the article is to develop improve the efficiency of the catering process and developing the information system for selecting catering by using improved method of clustering based on the analysis of geodata, numeric and categorical values. In the analysis, we will use two tools for analysis: RStudio and Tableau. To achieve the goal, the authors put certain tasks in order:

1. To get acquainted with the means of analysis of R and Tableau.
2. Consider the probable relationship between data.
3. Identify key criteria for finding gastronomic solutions.
4. Split data into clusters.
5. To supplement the general system for finding food establishments.

3 Materials and Methods

Tableau uses the k-means method with a dispersion-based distribution method, thus ensuring consistency between cycles. This goes through automatic pre-processing steps to reduce the cost of data preparation that is required for this type of analysis. These include the standardization of input parameters that automatically scalable data and multidimensional match analysis [1-15].

The K-means method in this study requires the initial specification of cluster centers. So, the analysis begins with one defined classter, which selects variables and calculates an arithmetic mean that is the threshold for splitting the data in half.

In the process of separation, centroids are used to initialize K-means. This allows to optimize the distribution of clusters. The next step is to select one of the two clusters for the splitting operation. Again, the previous algorithm is repeated, i.e.: there is a threshold for splitting the cluster in half. The clusters resulting from the splitting are initialized by the centroids of the two parts of the split cluster and the centroid of the central cluster. The algorithm is repeated until it reaches the set number of clusters [18].

In the process of analysis there is a large amount of multivariable data that need to be processed [19-20]. For this purpose, we apply the Lloyd's algorithm, by which we determine the square of the Euclidean distance for the calculation of clusters formed during the disengagement process. Their collaboration allows us to determine the initial centers for each $k > 1$. It is known that the result of our analysis is directly dependent on the number of clusters, so the resulting clustering is deterministic.

To analyze the data in Tableau, we use the Calinski-Harabasz criterion, which determines the quality of the clusters. The Calinski-Harabasz criterion can be defined as:

$$\frac{SS_B}{SS_W} \times \frac{(N-k)}{(k-1)}, \quad (1)$$

where SSB is the total variance between clusters; SSW - total dispersion within the cluster; k is the number of clusters; N is the number of views.

The value of the Calinski-Harabasz criterion indicates the cluster density. It means that the higher the value of this criterion, the denser the clusters are located (the low dispersion within the cluster), and the lower the cluster distance, the greater the difference between the clusters.

Since the Kalinsky-Harabash index is uncertain for k = 1, it cannot be used to detect instances of creating a single cluster.

Typically, clustering aims to highlight multiple groups of objects with similar characteristics within a group, and between groups - they are different. The peculiarity of co-clustering is the grouping of not only objects, but also the characteristics of these objects themselves. That is, if the data is presented in the form of a matrix, then clustering is a regrouping of the rows or columns of the matrix, and the co-clusterization is a re-grouping of the rows and columns of the data matrix.

In the process of analysis, the authors propose the use of cluster accumulation method. It is based on tools that allow you to divide the map into squares of a given size. Grouping by certain features of certain objects occurs in each square of the map. They also create clusters according to the algorithm described above. The process of splitting clusters takes place until all markers are included in the closest cluster grids [17].

If in the process of analysis it is found that some marker is located within several existing clusters, then by the algorithm of the method described above, the distance from each cluster to it is determined. Accordingly, the marker is then added to the closest, closest, cluster, using the fuzzy K-mean approach:

1) Initialization is carried out by accidental filling of matrix of the *F* with preservation in the conditions of normalization $\sum_{i=0}^{c} \mu_{ki} = 1$, returns to step 1 or accidentally fills cluster centroids *Vi* .

2) For each iteration we calculate:

$$V_i = \frac{\sum_{k=1}^{M} \mu_{ki}^m * X_k}{\sum_{k=1}^{M} \mu_{ki}^m}, i = \overline{1,c}, \quad (2)$$

$$D_{ki} = \sqrt{\left\| X_k - V_i \right\|^2}, k = \overline{1,M}, i = \overline{1,c}, \quad (3)$$

$$\mu_{ki}^{m} = \frac{1}{\sum_{j=1}^{c} \left(\frac{D_{ki}}{D_{kj}} \right)^{2/_{m-1}}}, k = \overline{1, M}, i = \overline{1, c} \tag{4}$$

At the end of each iteration, the state of achieving accuracy is checking

$$\max_{k=\overline{1,M}, i-\overline{1,c}} \left(\left| \mu_{ki} - \mu_{ki}^{*} \right| \right) < \varepsilon \tag{5}$$

` ,

where μ_{ki}^{*} is the value that was calculated on the previous iteration.

The result of such clustering is a list of tourist facilities with geodata and roads.

4 Experiment

In order to conduct research and analysis, you need to receive complete, updated information on tourist sites. To do this, we will use the Google Maps API, which has access to Google Maps websites and has a complete set of tools for determining geographic parameters [9, 12].

For analysis is convenient to use MongoDB, with the type of database NoSQL. JSON should be used for describing and storing documents because it is open source and required for use on different platforms [13].

The coordinates of the submitted tourist sites are provided in vector format and described in GeoJSON format.

The data is from kaggle.com, where they are open source. Authors provide certain nutrition characteristics: Id, Name, Cuisine Style, Ranking, Rating, Price Range, NumberofReviews, Reviews, URL_TA, ID_TA. They are presented in **Error! Reference source not found.**.

1		Name	City	Cuisine Sty	Ranking	Rating	Price Rang	Number of	Reviews	URL_TA	ID_TA
2	0	Martine of	Amsterdar	['French', '	1	5	$$ - $$$	136	[['Just like home', 'A Warm V	/Restaurar	d11752080
3	1	De Silverer	Amsterdar	['Dutch', 'E	2	4.5	$$$$	812	[['Great food and staff', 'just	/Restaurar	d693419
4	2	La Rive	Amsterdar	['Mediterr	3	4.5	$$$$	567	[['Satisfaction', 'Delicious ol	/Restaurar	d696959
5	3	Vinkeles	Amsterdar	['French', '	4	5	$$$$	564	[['True five star dinner', 'A su	/Restaurar	d1239229
6	4	Librije's Zu	Amsterdar	['Dutch', 'E	5	4.5	$$$$	316	[['Best meal.... EVER', 'super	/Restaurar	d6864170
7	5	Ciel Bleu R	Amsterdar	['Contemp	6	4.5	$$$$	745	[['A treat!', 'Wow just Wow'	/Restaurar	d696902
8	6	Zaza's	Amsterdar	['French', '	7	4.5	$$ - $$$	1455	[['40th Birthday with my Fan	/Restaurar	d1014732
9	7	Blue Peppe	Amsterdar	['Asian', 'Ir	8	4.5	$$$$	675	[['Great Experience', 'A true	/Restaurar	d697058
10	8	Teppanyak	Amsterdar	['Japanese	9	4.5	$$$$	923	[['Great Food & Service!', 'Su	/Restaurar	d697009

Fig. 1. Initial data

For further analysis, the we used only: Title, City, Rating, Price, Number of reviews. Preprocessing the data by the authors is done in Rstudio. The results are similar to those that are analyzed in everyday life by users: the better the quality, the higher the price. Further analysis by authors was carried out in the Tableau environment. To do this, you must first transform data. The Price.Range text field should be converted to numeric Price by replacing $ -> 1, $$ - $$$ -> 2, $$$$ -> 3 (data categorization) (Fig. 2).

1	,"Name","City","Rating","Price.Range","Number.of.Reviews"
2	1,"Martine of Martine's Table","Amsterdam",5,2,136
3	2,"De Silveren Spiegel","Amsterdam",4.5,3,812
4	3,"La Rive","Amsterdam",4.5,3,567
5	4,"Vinkeles","Amsterdam",5,3,564
6	5,"Librije's Zusje Amsterdam","Amsterdam",4.5,3,316
7	6,"Ciel Bleu Restaurant","Amsterdam",4.5,3,745
8	7,"Zaza's","Amsterdam",4.5,2,1455
9	8,"Blue Pepper Restaurant And Candlelight Cruises","Amsterdam",4.5,3,675
10	9,"Teppanyaki Restaurant Sazanka","Amsterdam",4.5,3,923
11	10,"Rob Wigboldus Vishandel","Amsterdam",4.5,1,450
12	11,"The Happy Bull","Amsterdam",4.5,2,295

Fig. 2. Transformed data

Thanks to the generated table on Fig. 3 you can pick up a meal for your needs. For example, in order to choose the catering with the cheapest price range (price = 1) and with the best rated catering (rating = 5), you can check the table by using a table to check if this type is available in a user-defined city. The example is Barcelona with a price range (price = 1), and with a rating of the catering (rating = 4) (Fig. 3).

Price&Rating city

		City									
Price	Rating	Amster..	Athens	Barcelo..	Berlin	Bratisla..	Brussels	Budape..	Copenh..	Dublin	E
1	1	Abc		Abc		Abc	Abc	Abc	Abc		
	2	Abc	Abc	Abc	Abc		Abc	Abc	Abc	Abc	
	3	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	
	4	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	
	5	Abc	Abc	Abc	[illegible]	[illegible]	Abc	Abc	Abc	Abc	
2	1	Abc		Abc	[illegible]	[illegible]	Abc	Abc	Abc	Abc	
	2	Abc	Abc	Abc	[illegible]	[illegible]	Abc	Abc	Abc	Abc	
	3	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	
	4	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	
	5	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	
3	1			Abc			Abc	Abc			
	2	Abc	Abc	Abc			Abc	Abc	Abc		
	3	Abc	Abc	Abc	Abc		Abc	Abc	Abc	Abc	
	4	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	
	5	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	

City: Barcelona
Rating: 4
Price: 1

Fig. 3. Pivote table

In order to verify that the user's assumptions about the place are accurate, you can view requests with reviews (Fig. 4 and Fig. 5).

Number of records price

City	
Amsterdam	543 1 906 105
Athens	674 457 736 67
Barcelona	3 013 1 461 3 659 263
Berlin	3 420 1 147 2 360 134
Bratislava	589 123 326 25
Brussels	1 094 384 1 599 124
Budapest	864 572 1 087 74
Copenhagen	761 262 987 93
Dublin	580 357 1 096 42
Edinburgh	434 410 965 46
Geneva	588 77 779 118
Hamburg	1 864 261 907 91
Helsinki	438 137 606 45
Krakow	456 375 495 25
Lisbon	1 463 758 1 631 109
Ljubljana	159 99 221 17
London	6 037 2 978 8 499 641

City: Barcelona
Price: 3
Number of Records: 263

Fig. 4. Statistical analisys by the reviws

Number of records rating

City	
Amsterdam	186 5 8 63 590 2 220 346
Athens	149 1 6 31 274 1 184 289
Barcelona	626 3 101 326 1 913 4 648 779
Berlin	703 2 40 113 1 183 4 105 915
Bratislava	177 13 46 202 489 136
Brussels	192 2 33 118 844 [illegible]
Budapest	170 21 82 463 1 52[illegible]
Copenhagen	138 3 14 71 488 1[illegible]
Dublin	127 9 50 421 1 309 [illegible]
Edinburgh	104 16 60 365 1 119 191
Geneva	162 1 7 45 342 892 113
Hamburg	394 33 80 586 1 720 310
Helsinki	81 8 43 327 688 79
Krakow	108 3 7 34 212 827 160
Lisbon	271 29 95 783 2 413 370
Ljubljana	32 11 81 317 55

City: Bratislava
Rating: 4
Number of Records: 489

Fig. 5. Statistical analisys by the rating

5 Results and Discussion

First of all, we find direct relations between parameters using the language R.

```
lm1 <- lm(Rating ~ Number.of.Reviews, data = data_restaurants)
qqnorm(lm1$residuals, col="orange", pch=20)
qqline(lm1$residuals, col = "blue") (Fig. 6)
```

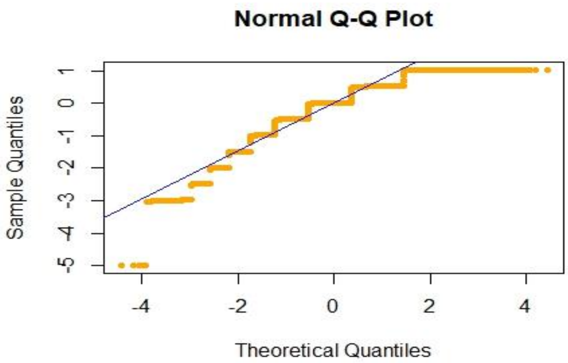

Fig. 6. Linear regression

```
model1=rpart(Rating ~ Price.Range, data=data_restaurant3)
p = predict(model1, data_restaurant3)
plot(model1) (Fig. 7).
```

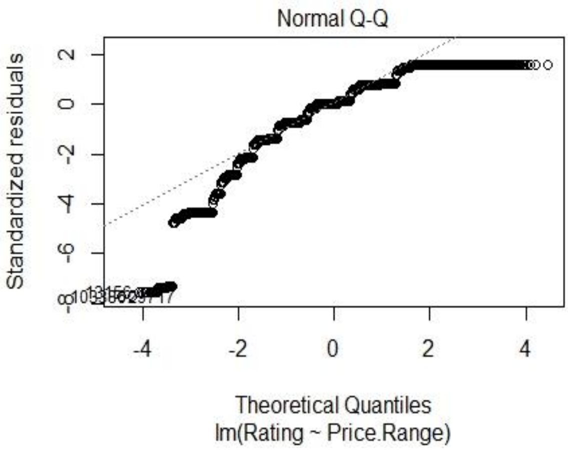

Fig. 7. Standardized residuals

The next step is calculation of the determination coefficient for linear regression models.

```
>lm = lm(Rating ~ Number.of.Reviews, data = data_restaurants)
>summary(lm)$r.squared
[1] 0.001152904
>lm1 = lm(Rating ~ Price.Range, data = data_restaurants)
>summary(lm1)$r.squared
[1] 0.01045306
```

```
>lm2 = lm(Number.of.Reviews ~ Price.Range, data = da-
ta_restaurants)
>summary(lm2)$r.squared
[1] 0.08443799
```

According to the results of the experiment, the determination coefficient is too small and the linear regression for the data set selected by the authors is not appropriate.

We selected a polynomial regression model that exactly matches the study described above:

```
>lmpoly = lm(Rating ~ Number.of.Reviews+ I(Number.of.Reviews^2)
+ I(Number.of.Reviews^3), data = data_restaurants)
>summary(lmpoly)$r.squared
[1] 0.001346703
>lmpoly = lm(Rating ~ Number.of.Reviews+ I(Number.of.Reviews^2)
, data = data_restaurants)
>summary(lmpoly)$r.squared
[1] 0.001346518
>lmpoly = lm(Rating ~ Number.of.Reviews+ I(Number.of.Reviews^2)
+ I(Number.of.Reviews^3)+ I(Number.of.Reviews^4), data = da-
ta_restaurants)
>summary(lmpoly)$r.squared
[1] 0.001418925
>lmpoly = lm(Rating ~ Number.of.Reviews+ I(Number.of.Reviews^2)
+ I(Number.of.Reviews^3)+ I(Number.of.Reviews^4)+
I(Number.of.Reviews^5), data = data_restaurants)
>summary(lmpoly)$r.squared
[1] 0.001629522
lmpoly = lm(Price ~ Number_of_reviewers+
I(Number_of_reviewers^2) + I(Number_of_reviewers^3)+
I(Number_of_reviewers^4)+ I(Number_of_reviewers^5), data = da-
ta_restaurants6)
>summary(lmpoly)$r.squared
[1] 0.05857986
>lmpoly = lm(Price ~ Number_of_reviewers+
I(Number_of_reviewers^2) + I(Number_of_reviewers^3), data = da-
ta_restaurants6)
>summary(lmpoly)$r.squared
[1] 0.04679744
>lmpoly = lm(Number_of_reviewers ~ Price+ I(Price^2) +
I(Price^3), data = data_restaurants6)
>summary(lmpoly)$r.squared
[1] 0.02783621
>lmpoly = lm(Price ~ Number_of_reviewers+
I(Number_of_reviewers^2) + I(Number_of_reviewers^3)+
```

```
I(Number_of_reviewers^4)+ I(Number_of_reviewers^5), data = da-
ta_restaurants6)
>summary(lmpoly)$r.squared
[1] 0.05857986
```

In this experiment, as in the previous ratio is too small. In parallel, the authors make clustering of data according to the given parameters.

According to the results of regression analysis, researchers have chosen two methods of splitting into clusters: first, according to the number of reviews; and secondly, in the price range. The results of clustering on the number of feedbacks are shown on Fig. 8, where the orange circles are marked by the cities with the most reviews. This color indicates the most accurate information with a range of values 1 537 453 – 2 136 471.

The red color is marked with somewhat fewer reviews, with a range of values ranging from 810,267 to 1,020,548. The smallest number of reviews is indicated by blue, with values from 41,434 to 455,280.

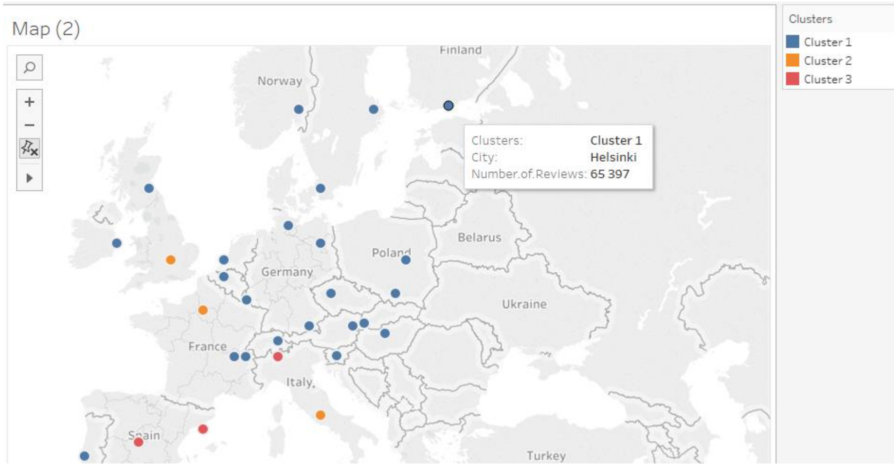

Fig. 8. Clustering results

On fig.9 is shown system of cluster, which is based on price criteria. Different types of clusters differ by colors:

- Red (1,609 – 1,722)
- Blue (1,735 – 1,807)
- Orange (1,829 – 1,916)
- Light blue (1,933 – 2,042)
- Green – noise data.

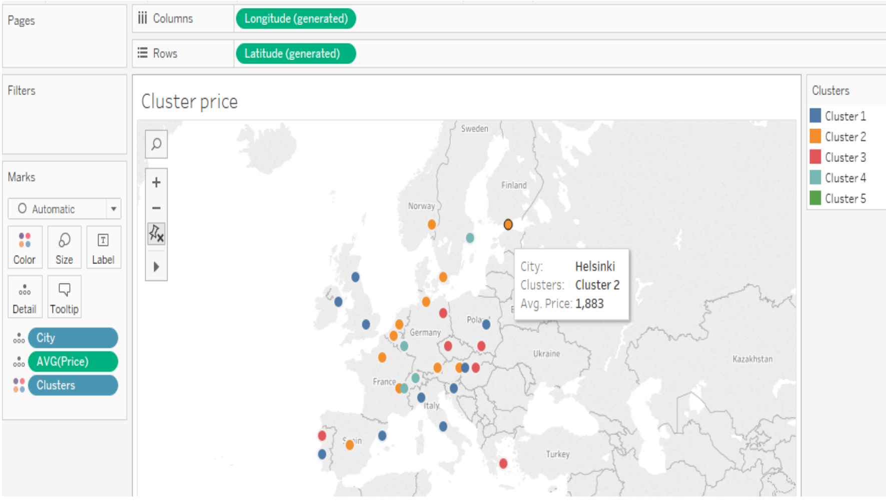

Fig. 9. Cluster representation

By way of overlaying two types of clustering, we received four finite clusters. The following tables 1-4 show detailed information about each of the clusters.

Table 1 demonstrates input data, particularly, which variables we use and clustering details.

Table 1. Inputs for Clustering

Variables:	Avg. Price, Sum of Number.of.Reviews
LevelofDetail:	City
Scaling:	Normalized

Table 2 is useful for analyst. For example, it helps to consider number of clusters.

Table 2. Summary Diagnostics

NumberofClusters:	4
NumberofPoints:	32
Between-groupSumofSquares:	0.8524
Within-groupSumofSquares:	0.0057509
TotalSumofSquares:	0.85815

Table 3 was created for viewers. In our case, it demonstrates which cluster contains a price that we need.

Table 3. Clusters legend

Clusters	NumberofItems	Avg. Price	SumofNumber.of.Reviews
Cluster 1	10	1.764	2.1361e+05
Cluster 2	11	1.8763	1.7895e+06
Cluster 3	6	1.6859	3.9e+05

Cluster 4	4	1.9894	2.49e+05
NotClustered	0		

Table 4 helps us to investigate clusters.

Table 4. Analysis ofVariance

Variable	F-statistic	p-value	Model Sum of Squares	DF	Error Sum of Squares	DF
Avg. Price	6.705	0.0006979	0.8524	4	0.8582	27
Sum of Number.of.Reviews	12.88	0.0001079	1.711	2	1.859	28

6 Conclusions

The paper represents the several methods of analysis usage, namely regression and clustering, for catering selection. The system can be used for Big data processing. The results of analysis have shown that the criteria do not have a direct relationship between each other (the correlation coefficient is very small), but data grouping and clustering gives the opportunity to form a proper description of the catering. Accordingly, for better perception, data was visualized in Tableau. As a result of the study, the authors received an information system of clusters, through which you can define a meal on the gastronomic criteria of the user and find out how accurate are the reviews about it. Also associative rules can be used for analysis.

References

1. What is Big Data, http://datascience.berkeley.edu/what-is-big-data/, last accessed 2000/11/16
2. Zhang, C., Murayama, Y.: Testing local spatial autocorrelation using /fc-order neighbors. Intern. J. of Geogr. Inform. Science, vol. 14, pp. 681–692 (2000).
3. Estivill-Castro, V., Lee, I.: Amoeba: Hierarchical clustering based on spatial proximity using Delaunay diagram. In: 9th Intern. Symp. on spatial data handling, pp. 26–41, Beijing, China (2000)
4. Kang, H.-Y., Lim, B.-J., Li K.-J.: P2P Spatial query processing by Delaunay triangulation , Lecture notes in computer science, vol. 3428, pp. 136–150, Springer, Heidelberg (2005)
5. Boehm, C., Kailing, K., Kriegel, H., Kroeger P.: Density connected clustering with local subspace preferences. In: Proc. of the 4th IEEE Intern. conf. on data mining, pp. 27–34, Los Alamitos: IEEE Computer Society (2004)
6. Wang, Y., Wu, X.: Heterogeneous spatial data mining based on grid, Lecture notes in computer science, vol. 4683, pp. 503–510,Springer/Heidelberg (2007)
7. Harel, D., Koren, Y.: Clustering spatial data using random walks. In: Proc. of the 7th ACM SIGKDD Intern. conf. on knowledge discovery and data mining, pp. 281–286, San Francisco, California (2001)
8. Turton, I., Openshaw, S., Brunsdon, C. et al.: Testing spacetime and more complex hyperspace geographical analysis tools. In: Innovations in GIS 7, pp. 87–100, L.: Taylor & Francis (2000)

9. Tung, A.K. H., Hou, J., Han, J.: Spatial clustering in the presence of obstacles. In: The 17th Intern. conf. on data engineering (ICDE'01), pp. 359–367, Heidelberg (2001)
10. Veres, O., Shakhovska, N.: Elements of the formal model big date. In: The 11th Intern. conf. Perspective Technologies and Methods in MEMS Design (MEMSTEH), pp. 81-83, Polyana (2015)
11. Agrawal, R., Gehrke, J., Gunopulos, D., Raghavan, P.: Automatic sub-space clustering of high dimensional data. In: Data mining knowledge discovery, vol. 11(1), pp. 5–33 (2005)
12. Guimei, L., Jinyan, L., Sim, K., Limsoon, W.: Distance based subspace clustering with flexible dimension partitioning. In: Proc. of the IEEE 23rd Intern. conf. on digital object identifier, vol. 15, pp. 1250-1254 (2007)
13. Aggarwal, C., Yu, P.: Finding generalized projected clusters in high dimensional spaces. In: ACM SIGMOD Intern. conf. on management of data, pp. 70-81 (2000)
14. Procopiuc, C.M., Jones, M., Agarwal, P.K., Murali, T.M. : A Monte Carlo algorithm for fast projective clustering. In: ACM SIGMOD Intern. conf. on management of data, pp. 418–427, Madison, Wisconsin, USA (2002)
15. Ankerst, M., Ester, M., Kriegel, H.-P.: Towards an effective cooperation of the user and the computer for classification. In: Proc. of the 6th ACM SIGKDD Intern. conf. on knowledge discovery and data mining, pp. 179-188, Boston, Massachusetts, USA (2000)
16. Peuquet, D.J.: Representations of space and time, N. Y., Guilford Press (2002)
17. Guo, D., Peuquet, D.J., Gahegan, M.: ICEAGE: Interactive clustering and exploration of large and high-dimensional geodata. Geoinformatica, vol. 3, N. 7, pp. 229-253 (2003)
18. Boyko, N.: Advanced technologies of big data research in distributed information systems. Radio Electronics, Computer Science, Control, vol. 4, pp. 66-77, Zaporizhzhya: Zaporizhzhya National Technical University (2017)
19. Mochurad, L., Boyko, N.: Solving Systems of Nonlinear Equations on Multi-core Processors. Advances in Intelligent Systems and Computing IV Selected Papers from the International Conference on Computer Science and Information Technologies, CSIT 2019, September 17–20, pp. 90-106, Lviv, Ukraine (2019)
20. Shakhovska, N., Boyko, N., Zasoba, Y., Benova, E.: Big data processing technologies in distributed information systems. Procedia Computer Science, 10th International conference on emerging ubiquitous systems and pervasive networks (EUSPN-2019), 9th International conference on current and future trends of information and communication technologies in healthcare (ICTH-2019), Vol. 160, 2019, pp. 561–566, Lviv, Ukraine (2019)