Comparison of Clustering Algorithms for Optimal Restaurant Location Selection Using Location-Based Social Networks Data

Đokić, Kristian; Potnik Galić, Katarina; Štavlić, Katarina

J. J. STROSSMAYER UNIVERSITY OF OSIJEK

EFOS
FACULTY OF ECONOMICS IN OSIJEK

CROATIAN ACADEMY OF ARTS AND SCIENCES
The Institute for scientific and art research work in Osijek

UNIVERSITY IN MARIBOR
Faculty of Economics and Business

UNIVERSITY IN TUZLA
Faculty of Economics in Tuzla

RED X 2021

10th International Scientific Symposium

REGION ENTREPRENEURSHIP DEVELOPMENT

ISSN 1848-9559
9 771848 955005

SVEUČILIŠTE J. J. STROSSMAYERA U OSIJEKU	EKONOMSKI FAKULTET U OSIJEKU	HRVATSKA AKADEMIJA ZNANOSTI I UMJETNOSTI Zavod za znanstveni i umjetnički rad u Osijeku	UNIVERZA V MARIBORU Ekonomsko-poslovna fakulteta	UNIVERZITET U TUZLI Ekonomski fakultet u Tuzli
JOSIP JURAJ STROSSMAYER UNIVERSITY OF OSIJEK	FACULTY OF ECONOMICS IN OSIJEK	CROATIAN ACADEMY OF ARTS AND SCIENCES The Institute for scientific and art research work in Osijek	UNIVERSITY IN MARIBOR Faculty of Economics and Business	UNIVERSITY IN TUZLA Faculty of Economics in Tuzla

10th
INTERNATIONAL SCIENTIFIC SYMPOSIUM REGION, ENTREPRENEURSHIP, DEVELOPMENT

Under the auspices of:

REPUBLIC OF CROATIA
MINISTRY OF SCIENCE AND EDUCATION

Osijek, June 2021

Publishers
Josip Juraj Strossmayer University of Osijek,
Faculty of Economics in Osijek, Croatia

For the publishers
Mirna Leko Šimić, Ph.D., Faculty of Economics in Osijek, Croatia
Boris Crnković, Ph.D., Faculty of Economics in Osijek, Croatia

Editors
Mirna Leko Šimić, Ph.D., Faculty of Economics in Osijek, Croatia
Boris Crnković, Ph.D., Faculty of Economics in Osijek, Croatia

Print
Studio HS internet d.o.o., Osijek

ISSN 1848 – 9559
Previous editions of the Proceedings published under the title
Economy of eastern Croatia – Vision and Growth

Proccedings indexed in:

10th INTERNATIONAL SCIENTIFIC SYMPOSIUM REGION, ENTREPRENEURSHIP, DEVELOPMENT

Content

A scientific paper

Kristian Đokić, Ph.D.
Polytechnic in Pozega
E-mail address: kdjokic@vup.hr

Katarina Potnik Galić, Ph.D.
Polytechnic in Pozega
E-mail address: kpotnikgalic@vup.hr

Katarina Štavlić, Ph.D.
Polytechnic in Pozega
E-mail address: kstavlic@vup.hr

COMPARISON OF CLUSTERING ALGORITHMS FOR OPTIMAL RESTAURANT LOCATION SELECTION USING LOCATION-BASED SOCIAL NETWORKS DATA

ABSTRACT

Machine learning algorithms are increasingly used in various fields. Unlike supervised algorithms that require the engagement and knowledge of experts in a particular area, unsupervised algorithms do not need it and are therefore more comfortable to use. Clustering algorithms belong to unsupervised algorithms and are used to group data according to a given similarity criterion with achieving significant similarity between data within the same group and minor similarities between data belonging to different groups.
In this paper, five clustering algorithms in restaurant location optimization in Zagreb are analyzed. The clustering algorithms' output result lists municipalities in Zagreb city divided into groups with similar properties. Based on these data, the investor can quickly conclude what individual municipalities are similar and based on that, a more objective assessment of the location of a restaurant or catering facility can be made before the investment.
The data based on which the algorithms divided parts of Zagreb into groups were obtained from a social network that can store user locations. One of the essential functions of the used social network is sharing information about restaurants, cafes, and other catering facilities. The common name of these social networks is a location-based social network. The paper compares the Gaussian Mixture Model algorithm, k-means algorithm, Hierarchies algorithm, Agglomerative Clustering algorithm, and Spectral Clustering algorithm. The selected five algorithms have the property that one of their input variables is the number of clusters.

Keywords: *Clustering, big data, restaurant, Foursquare, location-based social network.*

1. Introduction

The location of restaurants and catering facilities in populated areas is an essential factor that must be taken into account before entering the market and starting a business. In the last century, managers have often made location decisions based on experience and personal judgment. More recently, methods based on quantitative data have been available that can help managers make that decision. One of these methods is clustering, which divides objects into groups whose members are similar. Clustering and similar methods belong to machine learning are generally used by managers who manage many restaurants and catering facilities and enter new markets unknown to them.

For clustering to be used when choosing a new restaurant location, it is necessary to collect data on parts of a specific populated area. Based on them, using clustering, divide these parts into similar clusters. By analyzing the obtained clusters, we can objectively assess the general purpose of the inhabited area. Fortunately, clustering data is available on Location-based social networks.

In this paper, five different clustering algorithms are compared, using location-based social network data to help managers choose the optimal location for a restaurant or catering facility. The second section provides a brief overview of the papers that analyze the impact of the location on the business. The third section analyses the clustering algorithms used, while the fourth section describes the data, methodology and results used. The discussion and conclusion are in the last two sections.

2. The impact of location on the business

Choosing a location is a crucial decision that a company must make when starting a business, affecting its success. Location, whether good or bad, always involves the optimal operation of a company. The choice of location affects both revenue and operating costs. More attractive locations also mean higher rental or purchase costs. In comparison, less attractive locations, despite lower rental or purchase costs, can negatively affect attracting customers or attracting promising employees.

Depending on the industry in which it exists and the specifics of the activity in which it is engaged, the company should consider different criteria when choosing a location. For service businesses such as restaurants and cafes, accessibility and proximity to customers are vital. The location of customers and suppliers is also essential for the restaurant's business.

There is a lot of scientific and professional research on the importance and impact of location on an organization's business and development. Regardless of the type of organization and its business, location plays a significant role in their business. Numerous authors have investigated the impact of location on hotel profitability and survival (Gémar, Moniche & Morales, 2016) (Vivel-Bua, Lado-Sestayo, & Otero-Gonzalez, 2016), and the impact of various factors on location decision making (Yang, Luo & Law, 2014) (Adam & Mensah, 2014) (Canina, Enz & Harrison, 2005) (Lado-Sestayo, Otero-Gonzalez, Vivel-Bua & Martorell-Cunill, 2016). The location also impacts other tourist offers, accommodation capacities, and tourist behaviour (Shoval, McKercher, Ng & Birenboim, 2011). Bryan Tan believes that demand, brand loyalty, quality are essential for a successful restaurant business. He considers the location of the restaurant to be the most important determinant. Using location-based social networks data, he analyzed three parameters of power consumption, percentage of target customers, and the number of competitors based on the analysis of the optimal location of an Asian restaurant in Toronto. He concluded that location selection is a crucial factor in a restaurant's success and that such a location analysis can affect a restaurant's success (Tan, 2021). Kalnins and Chung discovered a general trend towards Clustering and Shoval and Cohen-Hattab, confirmed that there is a tendency of accommodation to concentrate in the city centre (Kalnins & Chung, 2004) (Shoval & Cohen-Hattab, 2001).

3. Clustering algorithms

Clustering is a powerful data analysis method whose importance is growing with the increasing processing power of computers. Anderberg classifies clustering algorithms into two categories, hierarchical and non-hierarchical. An example of a hierarchical algorithm is agglomerative clustering, while the most used non-hierarchical algorithm is k-means (Anderberg, 1973). Fasulo used other names when dividing clustering methods, so he divided them into partitioning and hierarchical clustering. In partitioning clustering, each object belongs to one group, while in hierarchical clustering, each group with more than one member is divided into smaller groups (Fasulo, 1999). Jain et al. propose a different taxonomy of clustering algorithms, dividing them into heuristic-based, density-based, and model-based (Jain, Topchy, Law & Buhmann, 2004). Rokach has further modified and expanded this division and proposes a division into partitioning, density-based, soft-computing, grid-based, model-based and hierarchical methods (Rokach, 2009). Xu and Tian present a division into two categories, classical and modern clustering algorithms. Classical algorithms are further divided into nine categories: algorithms based on partition, hierarchy, fuzzy theory, distribution, density, graph theory, grid, fractal theory, and model. Modern algorithms are divided into ten categories, algorithms based on a kernel, ensemble, swarm intelligence, quantum theory, spectral graph theory, affinity propagation, density and distance, and algorithms for spatial data, for the data stream and for large-scale data (Xu & Tian, 2015). Five algorithms are used in this paper: agglomerative, BIRCH, a mixture of Gaussians, K-means and spectral clustering. Below is a brief description of each.

3.1. Agglomerative Clustering

The agglomerative clustering method has been known since the 1970s (Cormack, 1971). The specificity of agglomerative clustering ranges from a finite number of objects n to be clustered. In each subsequent step, the two nearest clusters merge, and the number of sets of objects decreases to eventually reach one set (Day & Edelsbrunner, 1984). overtime, the authors have proposed various Agglomerative clustering optimizations, so Kurita offers an algorithm that uses a heap to store the distances of all pairs of clusters, while Ackerman et al. conclude that the method has not been studied theoretically enough and give a detailed overview and explanation (Kurita, 1991) (Ackermann, Blömer, Kuntze & Sohler, 2014).

3.2. BIRCH

The full name of this clustering method is Balanced Iterative Reducing and Clustering using Hierarchies. It is intended for large multidimensional data sets and offers the possibility of parallel processing. Zhang et al. in their paper prove that BIRCH is significantly better than the K-means algorithm in several characteristics (quality, stability, and speed) (Zhang et al., 1997)(Zhang et al., 1996). Lorbeer et al. propose a variation of an A-BIRCH algorithm and use it to perform an automatic threshold estimation for the existing algorithm (Lorbeer et al., 2018)

3.3. Mixture of Gaussians

A Gaussian mixture model is based on assumptions that all data points can be generated from unknown number of Gaussian distributions with unknown parameters. It is defined by Pearson one hundred and twenty years ago (Pearson, 1901). Dasgupta encouraged more intensive use of the model with his paper two decades ago (Dasgupta, 1999).

3.4. K-Means

K-means is one of the better-known clustering algorithms. It is proposed by Steinhous (Steinhaus, 1956). Before starting clustering, it is necessary to define the number of clusters. The algorithm tries to separate the samples into clusters to minimize the sum of the squares of the distance of the samples within the cluster. One of the fundamental disadvantages is the dependence of performance on initial conditions (Scikit-learn developers, 2021).

A large number of authors has analyzed the k-means algorithm, and many optimizations have been proposed. Alsabt et al. proposed an efficient k-means clustering algorithm that significantly reduced computation time (Alsabti, Ranka & Singh, 1997). Na et al. proposed an acceleration method based on storing previous iterations' properties, thus avoiding calculating each object's distance (Na, Xumin & Yong, 2010).

3.5. Spectral Clustering

Spectral clustering creates an affinity matrix for paired samples, normalizes them, and calculates the normalized matrix's eigenvectors. The method is effective for a small number of clusters (De Sa, 2005). Very often, it has better results than the k-means method, and it is easy to implement (Von Luxburg, 2007).

4. Zagreb case study

4.1. Data and methodology

Zagreb is the capital of the Republic of Croatia, with about 800,000 inhabitants. From 2011 to 2018, the number of tourists doubled, i.e. from 600,000 per year to 1400,000 tourists per year (City Office for Strategic Planning and City Development, 2021). The peak in the number of tourists was in 2019. Due to the pandemic caused by the COVID-19 virus in 2020, a significant decline was observed (City Office for Strategic Planning and City Development, 2021). This situation will force the closure of some restaurants and catering facilities, which may be an opportunity to open new restaurants after the pandemic stops.

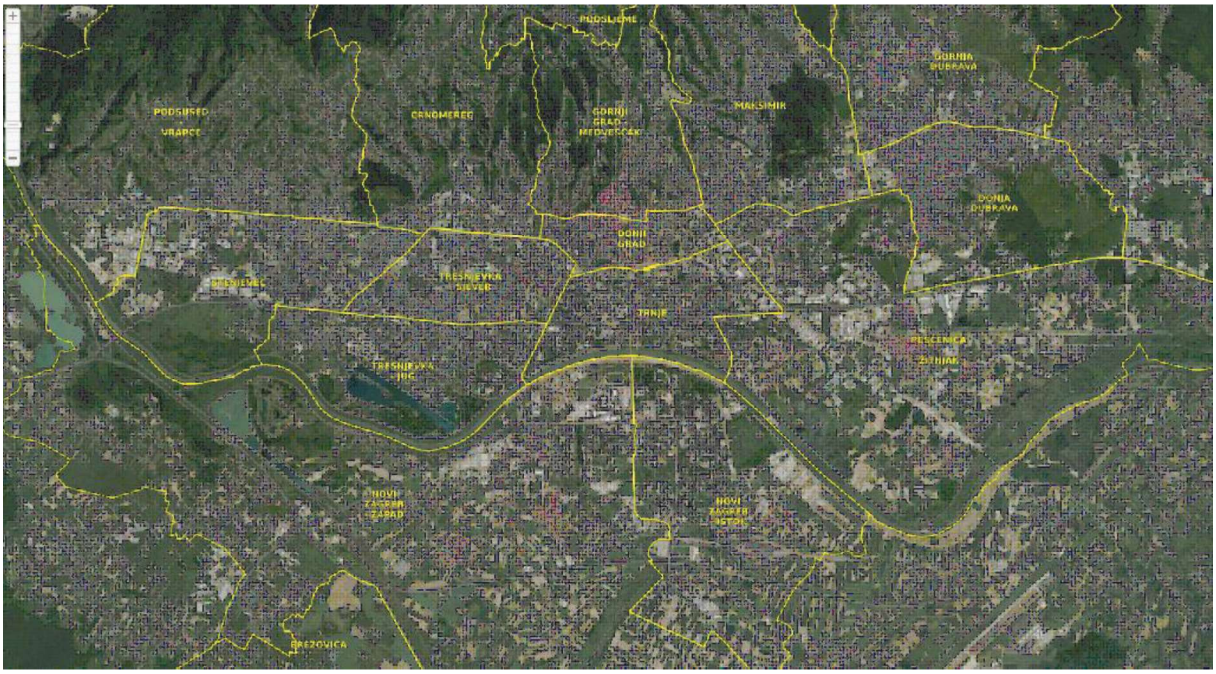
As already mentioned, one of the most critical factors affecting a restaurant's business is its location. For the potential investor to assess the characteristics of individual parts of Zagreb as objectively as possible, the already existing division by municipalities will be used. Table 1 lists all municipalities in the city of Zagreb. The table also shows the area, population in 2001 and 2011, and population density per square kilometre. In figure 1, a satellite image of Zagreb city with visible boundaries of individual municipalities can be seen.

Table 1: Zagreb municipalities list

	Municipalities	Area (km²)	Population (2011)	Population (2001)	Population density (2001)
1.	Donji Grad	3.01	37,123	45,108	14,956.2
2.	Gornji Grad – Medveščak	10.12	31,279	36,384	3,593.5
3.	Trnje	7.37	42,126	45,267	6,146.2
4.	Maksimir	14.35	49,448	49,750	3,467.1
5.	Peščenica – Žitnjak	35.30	56,446	58,283	1,651.3
6.	Novi Zagreb – istok	16.54	59,227	65,301	3,947.1
7.	Novi Zagreb – zapad	62.59	58,025	48,981	782.5
8.	Trešnjevka – sjever	5.83	55,342	55,358	9,498.6
9.	Trešnjevka – jug	9.84	66,595	67,162	6,828.1
10.	Črnomerec	24.33	39,040	38,762	1,593.4
11.	Gornja Dubrava	40.28	62,221	61,388	1,524.1
12.	Donja Dubrava	10.82	36,461	35,944	3,321.1
13.	Stenjevec	12.18	51,849	41,257	3,387.3
14.	Podsused – Vrapče	36.05	45,771	42,360	1,175.1
15.	Podsljeme	60.11	19,249	17,744	295.2
16.	Sesvete	165.26	70,633	59,212	358.3
17.	Brezovica	127.45	12,040	10,884	85.4
	TOTAL	641.43	792,875	779,145	1,214.9

Source: Croatian Bureau of Statistics

Figure 1: Zagreb with municipalities borders



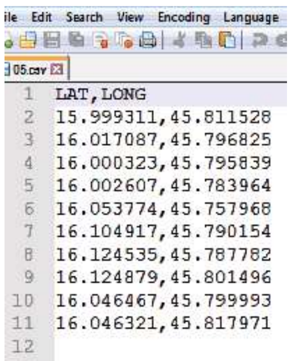
Source: by the authors

To use clustering algorithms and divide individual municipalities of the city of Zagreb into clusters where cluster members have similar characteristics, it is necessary to have data on individual municipalities. Some towns have many publicly available data on their official websites, but Zagreb is not among them. An excellent example of a city with a range of information available on its official website is Hong Kong (The Government of the Hong Kong Special Administrative Region, 2021). For this reason, the paper uses data available on a Location-based social network called Foursquare.

Foursquare Labs Inc. is an American company founded in 2009 in New York. The app they offered in the smartphone market was called Foursquare, and it allowed real-time location sharing and location tagging. The application was free and was among the first to offer this concept, and very quickly gathered many users who used it to mark, rate and comment on restaurants. More than ten years have passed since the company was founded. They currently offer several different products based on the concept of a location social network, but for different categories of users. Applications designed for mobile platforms are the Foursquare City Guide, Marsbot for Airpods, Panel App and Foursquare Swarm (Foursquare Labs Inc, 2021).

One of the services offered by Foursquare Labs Inc is the Foursquare API. It is an interface to their database where it is enough to enter latitude and longitude, and radius to get all the objects they have stored in their database. There are restaurants, coffee bars, shops, monuments, zoos, pet stores, grocery stores, trails and more. Access to the database is free for a small number of queries, while it is charged if the number of queries to their database is above the specified limit. Unfortunately, their database does not have information on individual facilities' location at the municipal level within cities. To obtain good quality results, each municipality is described by a polygon. Using the algorithm for finding the centre of the polygon, each municipality's central point's latitude and longitude are defined. After that, data on facilities were obtained for each municipality in the form of a table. Figure 2 shows an example of a file describing a polygon for a particular municipality.

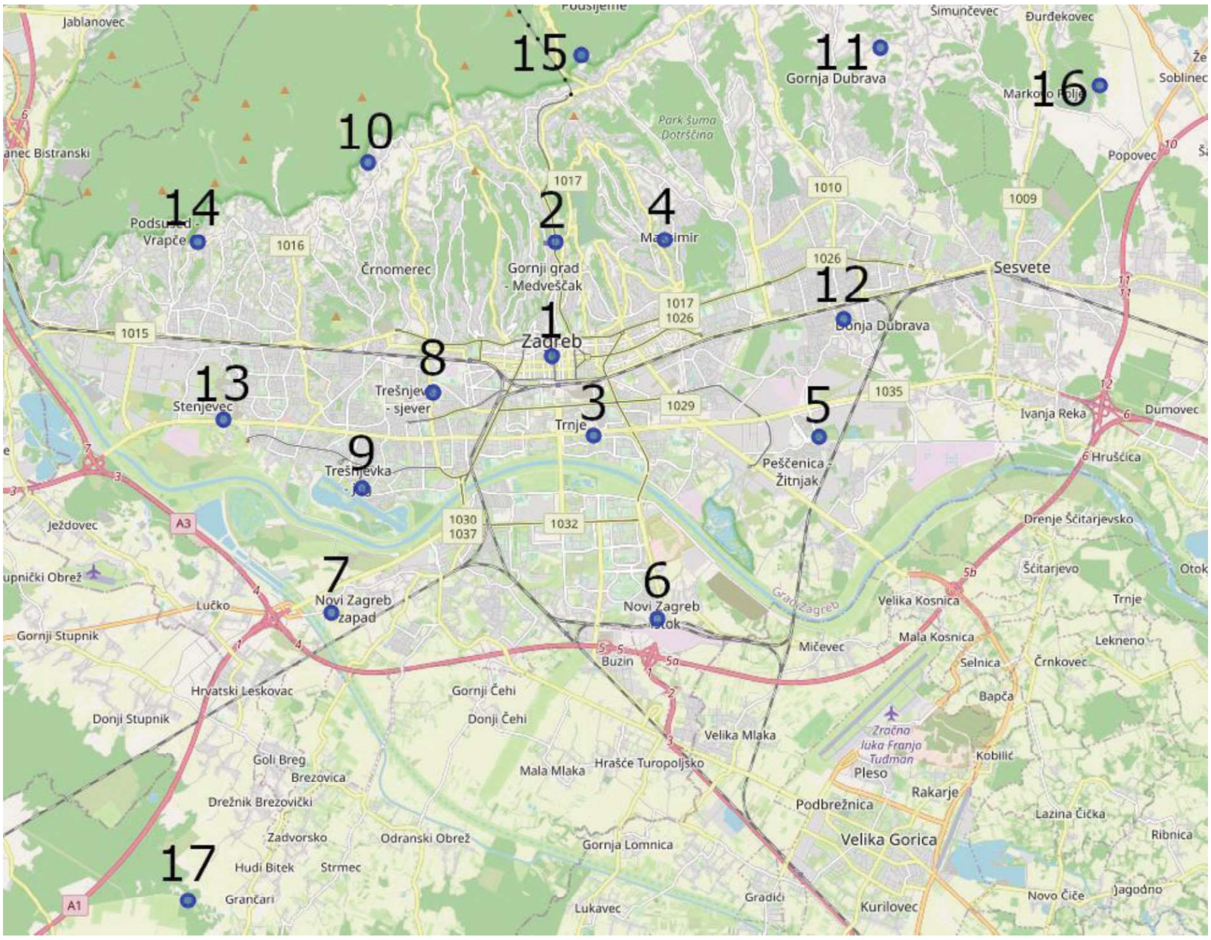
Figure 2: Municipality polygon example



Source: authors

Figures 3 show the central points of individual municipalities with the numbers corresponding to the first column in Table 1.

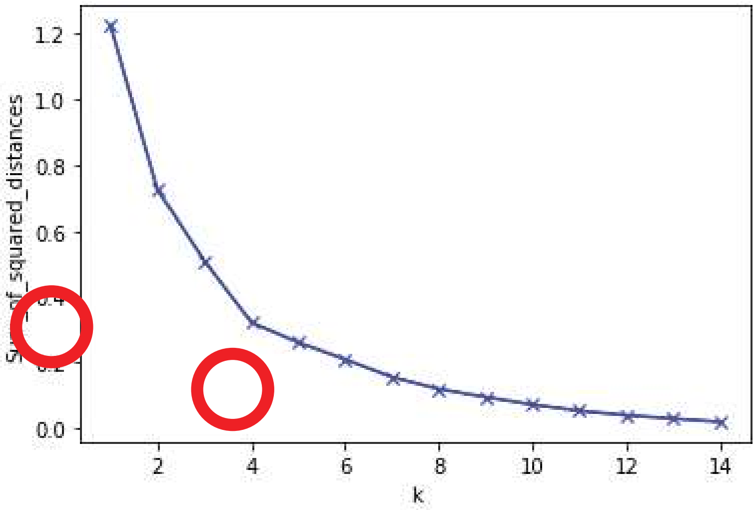
Figure 3: Central points of all municipalities



Source: authors

Xu et al. list four basic steps in applying a clustering algorithm: feature selection or extraction, clustering algorithm design or selection, cluster validation, and interpretation results (Xu & Wunsch, 2005). This is the completion of the first step, feature selection or extraction. Before using the above five algorithms, it is necessary to define the number of clusters. One of the most famous methods is the so-called elbow method. The technique consists of displaying the explained variation or sum of squared distances depending on the number of clusters. The elbow of the curve is selected as the cluster number.

Figure 4: Elbow method for optimal cluster number

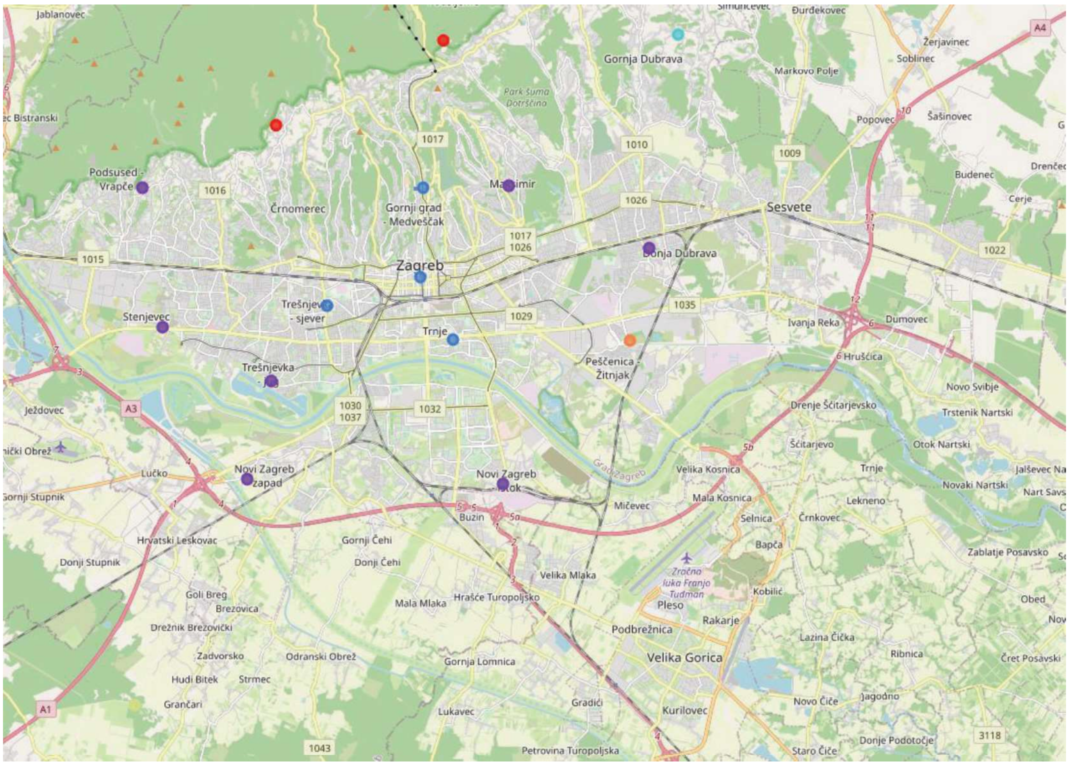


Source: authors

Figure 4 shows a graph obtained from the data, indicating that the optimal number of clusters is 3. After testing with only three clusters, the algorithms would put one municipality in each of the two clusters, while in the third cluster, there were 15 municipalities left. For this reason, the number of clusters has been increased to 7, and in figure four are rounded the places where the elbows are located. The elbow above number 4 is much more noticeable than the one above number 8.

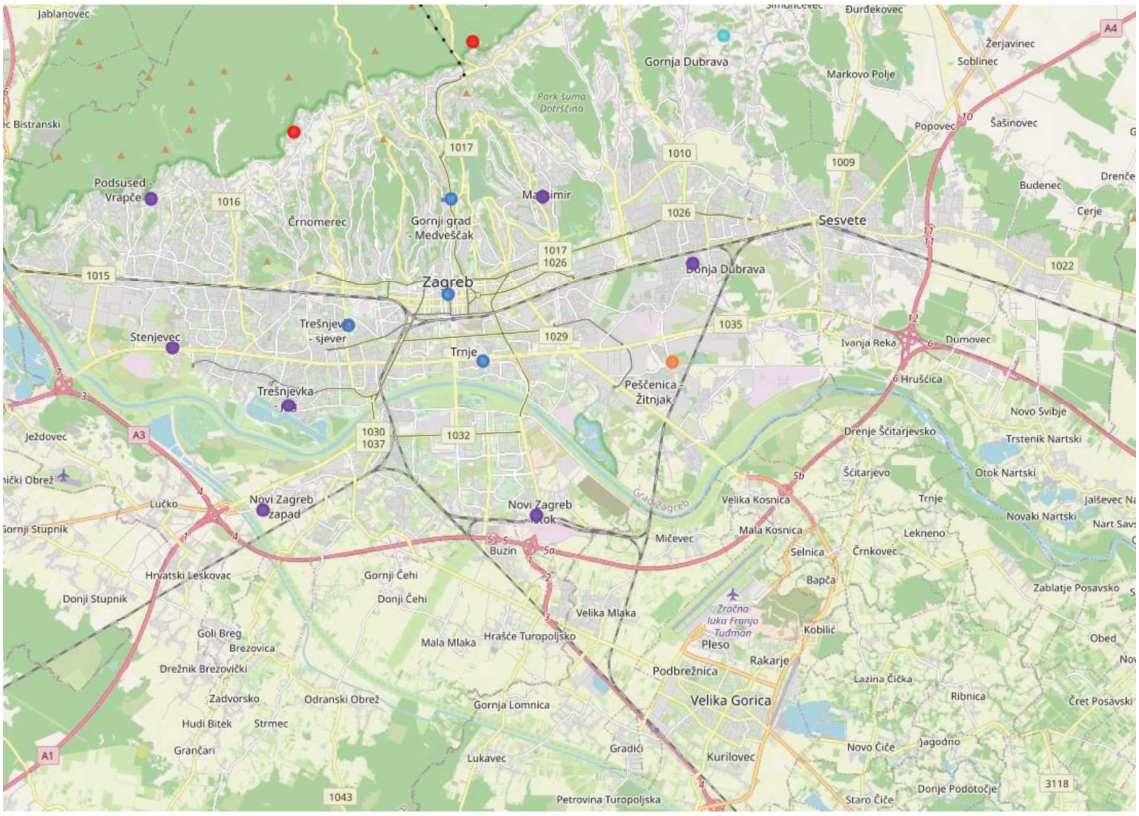
4.2. Results

Figure 5: Municipalities divided into seven clusters (Agglomerative Clustering)



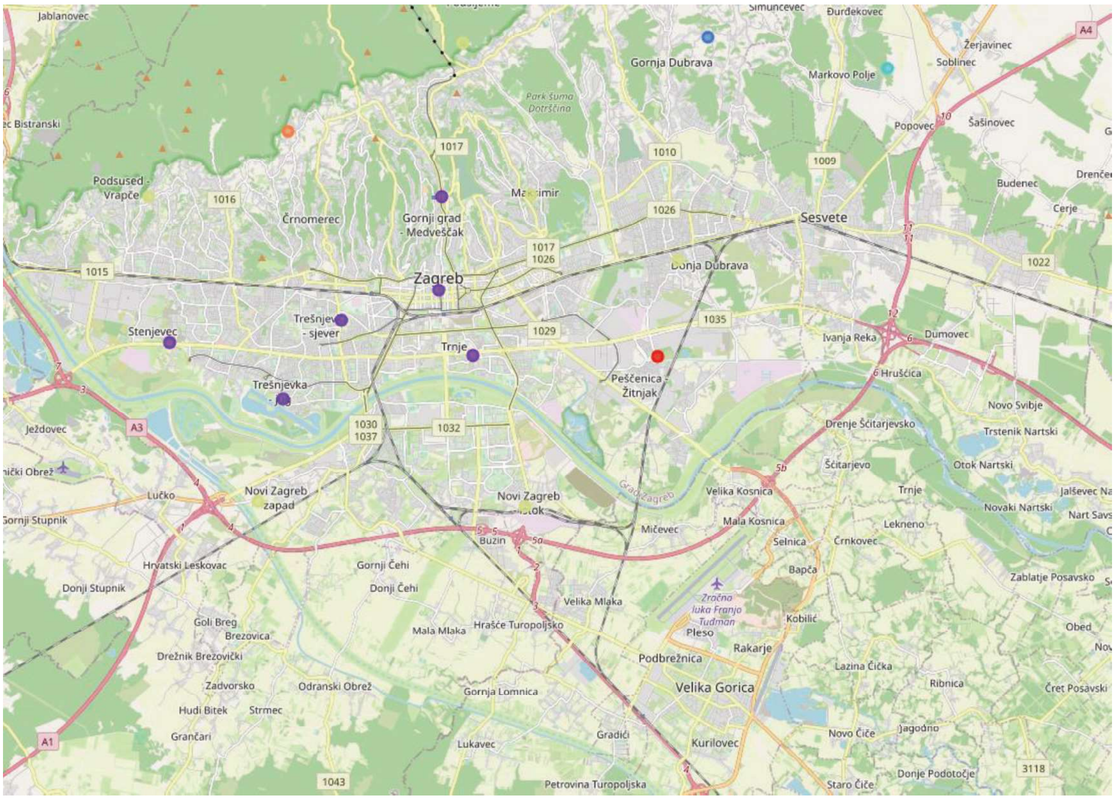
Source: authors

Figure 6: Municipalities divided into seven clusters (BIRCH Clustering)



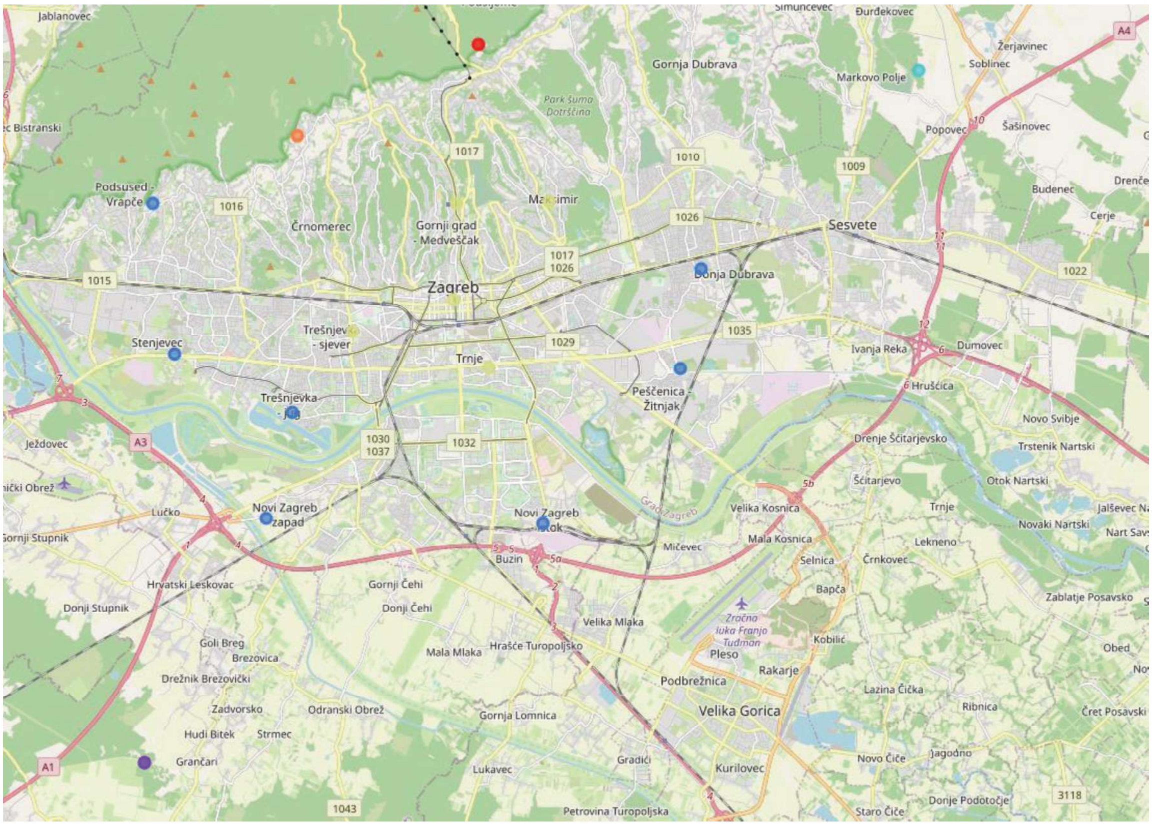
Source: authors

Figure 7: Municipalities divided into seven clusters (Mixture of Gaussians)



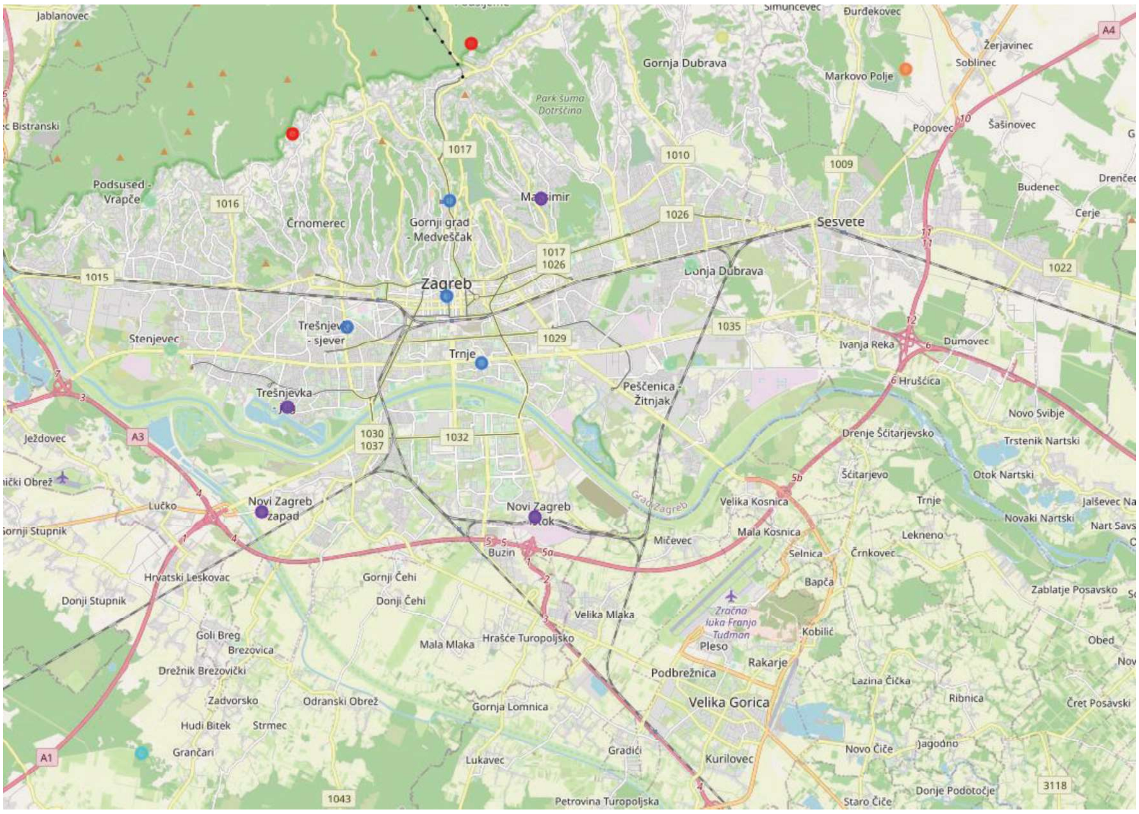
Source: authors

Figure 8: Municipalities divided into seven clusters (K-means)



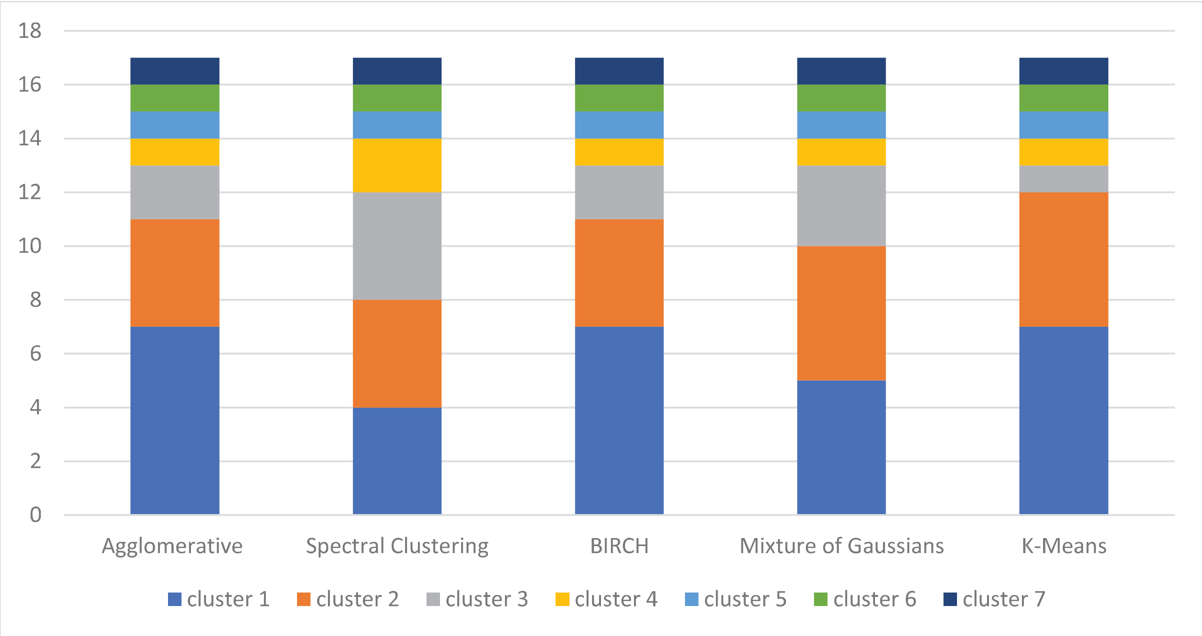
Source: authors

Figure 9: Municipalities divided into seven clusters (Spectral Clustering)



Source: authors

Figure 10: Comparison of the number of municipalities within a cluster



Source: authors

5. Discussion

Several authors have used Foursquare API and clustering algorithms in their research. Novovic et al. used the foursquare API for user's mobility flows aggregation and analyzed dynamics over two years in ten cities (Novović, Grujić, Brdar, Govedarica, & Crnojević, 2020). Sun used the same data source in analyzing the locality of spatial interactions in New York City, and Noulas et al. combine data obtained from the Foursquare API with data from mobile operators to infer user activity in urban environments (Sun, 2016) (Noulas, Mascolo & Frias-Martinez, 2013). It is evident that the foursquare service data are of sufficient quality to draw conclusions based on them. In the case of applying different clustering algorithms based on that data, it can be seen that the results are similar. The similarity is best seen in Figure 10, which shows the number of municipalities per cluster. In figures 5, 6, 7, 8 and 9 can be seen that all algorithms put four municipalities in the centre of Zagreb's city (Donji grad, Gornji grad, Trnje and Trešnjevka sjever) in one cluster. Also, two municipalities south of the river Sav (Novi Zagreb istok and Novi Zagreb zapad) belong to the same cluster in all five algorithms.

Furthermore, Figure 5 and Figure 6 are identical, which may indicate an error. This is not an error, but Agglomerative and BIRCH clustering algorithms give the same complete result. This is also visible in figure 10.

The proposed model can be compared with other models, but it is challenging to evaluate the results obtained by each algorithm. One possibility is to assess the results obtained by real estate professionals, where the number of professionals should be large enough to avoid subjectivity.

6. Conclusion

The paper points out that clustering algorithms with data available on location-based social networks can be used to analyze the properties of individual municipalities within the city of Zagreb. Based on the available data, the potential investor has a more objective picture of the assets of individual municipalities, which is very important if they do not know the city well enough. Also, the clustering results can be attractive to different professions, and finally, the

results themselves can be used by the city administration itself to analyze the city's development. Such analyzes should be done continuously because the addition of a time-domain could follow trends in individual municipalities.
The presented approach can be improved in several ways. First, data from other location-based social networks should be obtained. One of the exciting characteristics that would affect the division is undoubtedly the monthly income per family member within a municipality. Besides, the algorithm could be improved in a way that more precisely locations objects within individual municipalities.
One of the shortcomings of the paper is certainly the lack of evaluation of the obtained results. This could be realized by having the quality of the results assessed by experienced real estate agents and this is one of the possible future paper upgrades.
Finally, the Zagreb City Administration should be aware that the public availability of data on individual municipalities can increase investors' interest for the simple reason that they can more objectively assess which location is more attractive for them to invest in.

REFERENCES

Ackermann, M. R., Blömer, J., Kuntze, D., & Sohler, C. (2014***). Analysis of agglomerative clustering***. *Algorithmica, 69*, 184–215.

Adam, I., & Mensah, E. A. (2014). ***Perceived spatial agglomeration effects and hotel location choice***. *Anatolia, 25*, 49–60.

Alsabti, K., Ranka, S., & Singh, V. (1997). ***An efficient k-means clustering algorithm***.

Anderberg, M. (1973). R., 1973. ***Cluster analysis for applications***.

Canina, L., Enz, C. A., & Harrison, J. S. (2005). ***Agglomeration efects and strategic orientations: Evidence from the US lodging industry***. *Academy of management journal, 48*, 565–581.

City Office for Strategic Planning and City Development. (12. February 2021). ***TOURIST ARRIVALS AND NIGHTS IN THE CITY OF ZAGREB, JULY 2020.*** Dohvaćeno iz www.zagreb.hr: https://www.zagreb.hr/userdocsimages/arhiva/statistika/2020/turizam%202020/Turizam%20VII.%202020._web.pdf

City Office for Strategic Planning and City Development. (12. February 2021). ***TURIZAM srpanj 2019***. Dohvaćeno iz www.zagreb.hr: https://www.zagreb.hr/userdocsimages/arhiva/statistika/2019/turizam%202019/Turizam%20%20VII.%202019%20.pdf

Cormack, R. M. (1971). ***A review of classification***. *Journal of the Royal Statistical Society: Series A (General), 134*, 321–353.

Dasgupta, S. (1999). ***Learning mixtures of Gaussians***. *40th Annual Symposium on Foundations of Computer Science (Cat. No. 99CB37039)*, (str. 634–644).

Day, W. H., & Edelsbrunner, H. (1984). ***Efficient algorithms for agglomerative hierarchical clustering methods***. *Journal of classification, 1*, 7–24.

De Sa, V. R. (2005). ***Spectral clustering with two views***. *ICML workshop on learning with multiple views*, (str. 20–27).

Fasulo, D. (1999). ***An analysis of recent work on clustering algorithms***. Tech. rep., Technical report.

Foursquare Labs Inc. (23. January 2021). ***About***. Dohvaćeno iz foursquare.com: https://foursquare.com/about

Gémar, G., Moniche, L., & Morales, A. J. (2016). ***Survival analysis of the Spanish hotel industry***. *Tourism Management, 54*, 428–438.

Jain, A. K., Topchy, A., Law, M. H., & Buhmann, J. M. (2004). ***Landscape of clustering algorithms***. *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, *1*, str. 260–263.

Kalnins, A., & Chung, W. (2004). ***Resource-seeking agglomeration: a study of market entry in the lodging industry***. *Strategic Management Journal, 25*, 689–699.

Kurita, T. (1991). ***An efficient agglomerative clustering algorithm using a heap***. *Pattern Recognition, 24*, 205–209.

Lado-Sestayo, R., Otero-Gonzalez, L., Vivel-Bua, M., & Martorell-Cunill, O. (2016). ***Impact of location on profitability in the Spanish hotel sector.*** *Tourism Management, 52*, 405–415.

Lorbeer, B., Kosareva, A., Deva, B., Softić, D., Ruppel, P., & Küpper, A. (2018). ***Variations on the clustering algorithm BIRCH***. *Big data research, 11*, 44–53.

Malmi, E., Do, T. M., & Gatica-Perez, D. (2013). ***From Foursquare to my square: Learning check-in behavior from multiple sources***. *Proceedings of the International AAAI Conference on Web and Social Media*, *7*.

Na, S., Xumin, L., & Yong, G. (2010). ***Research on k-means clustering algorithm: An improved k-means clustering algorithm***. *2010 Third International Symposium on intelligent information technology and security informatics*, (str. 63–67).

Noulas, A., Mascolo, C., & Frias-Martinez, E. (2013). ***Exploiting Foursquare and cellular data to infer user activity in urban environments***. *2013 IEEE 14th international conference on mobile data management*, *1*, str. 167–176.

Novović, O., Grujić, N., Brdar, S., Govedarica, M., & Crnojević, V. (2020). ***Clustering Foursquare Mobility Networks to Explore Urban Spaces.*** *World Conference on Information Systems and Technologies*, (str. 544–553).

Pearson, K. (1901). LIII. ***On lines and planes of closest fit to systems of points in space***. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 2*, 559–572.

Rokach, L. (2009). ***A survey of clustering algorithms***. U *Data mining and knowledge discovery handbook* (str. 269–298). Springer.

Scikit-learn developers. (22. February 2021). ***Clustering***. Dohvaćeno iz Scikit-learn: https://scikit-learn.org/stable/modules/clustering.html#k-means

Shoval, N., & Cohen-Hattab, K. (2001). ***Urban hotel development patterns in the face of political shifts.*** *Annals of Tourism Research, 28*, str. 908–925.

Shoval, N., McKercher, B., Ng, E., & Birenboim, A. (2011). ***Hotel location and tourist activity in cities.*** *Annals of tourism research, 38*, 1594–1612.

Steinhaus, H. (1956). ***Sur la division des corps matériels en parties***. *Bull. Acad. Polon. Sci, 1*, 801.

Sun, Y. (2016). ***Investigating "locality" of intra-urban spatial interactions in New York city using foursquare data***. *ISPRS International Journal of Geo-Information, 5*, 43.

Tan, B. (1. March 2021). ***Strategic Clustering for Establishing a Restaurant***. Dohvaćeno iz https://towardsdatascience.com/strategic-location-for-establishing-an-asian-restaurant-c3aecf2496b1

The Government of the Hong Kong Special Administrative Region. (15. February 2021). ***Population and Household Statistics Analyzed by District Council District***. Dohvaćeno iz https://www.censtatd.gov.hk/hkstat/sub/sp150.jsp?productCode=B1130301

Vivel-Bua, M., Lado-Sestayo, R., & Otero-Gonzalez, L. (2016). ***Impact of location on the probability of default in the Spanish lodging industry: A study of MSMEs***. *Tourism Economics, 22*, 593–607.

Von Luxburg, U. (2007). ***A tutorial on spectral clustering***. *Statistics and computing, 17*, 395–416.

Xu, D., & Tian, Y. (2015). ***A comprehensive survey of clustering algorithms***. *Annals of Data Science, 2*, 165–193.

Xu, R., & Wunsch, D. (2005). ***Survey of clustering algorithms***. *IEEE Transactions on neural networks, 16*, 645–678.

Yang, Y., Luo, H., & Law, R. (2014). ***Theoretical, empirical, and operational models in hotel location research.*** *International Journal of Hospitality Management, 36*, 209–220.

Zhang, T., Ramakrishnan, R., & Livny, M. (1996). ***BIRCH: an efficient data clustering method for very large databases.*** *ACM sigmod record, 25*, 103–114.

Zhang, T., Ramakrishnan, R., & Livny, M. (1997). BIRCH: ***A new data clustering algorithm and its applications.*** *Data Mining and Knowledge Discovery, 1*, 141–182.