

Association for Information Systems

**AIS Electronic Library (AISeL)**

---

Wirtschaftsinformatik 2021 Proceedings

Track 6: Digital Retail

---

## Discovering Geographical Patterns of Retailers' Locations for Successful Retail in City Centers

Philipp zur Heiden  
*Universität Paderborn*

Daniel Winter  
*Universität Paderborn*

Follow this and additional works at: <https://aisel.aisnet.org/wi2021>

---

zur Heiden, Philipp and Winter, Daniel, "Discovering Geographical Patterns of Retailers' Locations for Successful Retail in City Centers" (2021). *Wirtschaftsinformatik 2021 Proceedings*. 1.  
<https://aisel.aisnet.org/wi2021/UDigitalRetail/Track06/1>

This material is brought to you by the Wirtschaftsinformatik at AIS Electronic Library (AISeL). It has been accepted for inclusion in Wirtschaftsinformatik 2021 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact [elibrary@aisnet.org](mailto:elibrary@aisnet.org).

# Discovering Geographical Patterns of Retailers' Locations for Successful Retail in City Centers

Philipp zur Heiden, Daniel Winter

<sup>1</sup> Paderborn University, Paderborn, Germany  
philipp.zur.heiden@upb.de, danielkorsmeier@gmail.com

**Abstract.** City centers and resident retail businesses have to react to the continuous growth of online retail. However, some city centers are far more successful concerning the total turnover in relation to its inhabitants. Using machine learning and data analysis methods, we investigate the types and locations of retail businesses inside the city center, comparing successful and unsuccessful city centers. Our results show that success does not come with particular types of shops, but rather with centrality and bundled shopping areas. We provide insights for planning and developing successful retail in city centers to compete and interact with online retail.

**Keywords:** Machine Learning, City Center, Retail, Clustering, Centrality.

## 1 Introduction

The rise of digital retail continues to pressurize retail in German city centers [1], [2]. The downwards spiral of fewer customers and fewer retail businesses inside the city center further increases the number of vacancies in many formerly successful retail areas in city centers [3], [4]. This primarily affects medium-sized cities and smaller retail businesses because they more often cannot focus on increasing their customer experience as a central element of high street retail [5]. Retail businesses are not the only businesses affected by this transition, but also other types of businesses common in high streets, e.g., the service industry, restaurants, and businesses for entertainment and amusement. We continue to use the term businesses to refer to all different types.

With machine learning and data-driven studies transforming whole industries [6], we see new ways to discover novel insights for the planning and development of city centers to strengthen the advantages of city center businesses. Manual and conventional analysis of geographical data cannot deliver insights on comprehensive patterns of business locations. With machine learning, however, one can analyze geographical data and information about businesses making up retail in different city centers by using novel methods. We suspect more factors than the sheer size of the city and its inhabitants to influence the success some city centers have, whereas others fall behind. Therefore, we focus on the following research question: *Which geographical patterns of retailers' locations enable successful retail in city centers?*

## 2 Theoretical Concepts: City Centers and Machine Learning

City centers are network models of different entities, comprising housing, employment, and transportation [7], [8]. Different indices are able to estimate the success of different city centers. One common index for measuring retail success in Germany is the retail centrality index, defined by the percentage value of the quotient of spending capacity of the resident population and turnover of the local retail businesses [9], [10]. A percentage bigger than 100 % shows that a city attracts more customers spending money in the city center from its vicinity than it “looses” customers to other cities. The retail centrality index is only calculated once per city and does not indicate the success of different types of retailers. Naturally, bigger cities and regional centers—e.g., Berlin, Munich—will have a significantly higher retail centrality index [9]. Thus, while comparing different cities based on their retail centrality index, one should always assure that the cities’ populations are even.

Machine learning, in general, describes the automatic extraction of algorithms for data analysis purposes [11]. Research in machine learning revolves around the development and application of new methods to analyze vast amounts of data to gain otherwise hidden patterns and insights [12]. Machine learning comprises three types of learning strategies: supervised learning, unsupervised learning, and reinforcement learning [12]. In our study, we rely on supervised learning to investigate differences between successful and unsuccessful cities. We apply decision trees [13], [14], random forests [15], and support vector machines [16] to answer our research question, using types and locations of businesses as input variables and success classification based on the retail centrality index as our output variable.

## 3 Research Method

Analyzing vast amounts of data has become a popular research method in the IS discipline, as it not only serves to gain insights but also to develop theory to extend the knowledge scope of IS [17]. In this study, we apply the research method of utilizing big data sources and advanced analytics by Müller et al. [18]. They picture the research with unstructured data as a process of four phases: framing a research question, collecting the necessary data, the computational analysis, and, finally, interpreting the results [18].

As our study already proposed a research question in Section 1, we continue with our data collection and analysis steps. First, we have to decide on the highly and poorly ranked cities in terms of success. We focus on middle-sized cities (20,000 to 100,000 inhabitants) in Germany and use the retail centrality index to define successful and unsuccessful cities (cf. Section 2)<sup>1</sup>. We use a total of 40 cities for our analysis—20 of them having highly valued retail centrality indices. The other 20 cities have the lowest centrality indices, indicating the poorest performing cities for our research.

---

<sup>1</sup> Data taken from GMA GmbH (2017) and <http://www.mb-research.de/>

As we need geospatial data to find patterns based on the location of the different businesses, we opt for OpenStreetMap (OSM) in favor of Google Maps, because the German community of OSM is much more active and serves as the foundation of high-quality data [19], [20], [21]. We extract the different businesses (entities with OSM categories *shop* and *amenity*) of highly and poorly ranked cities with an API for OSM<sup>2</sup>. For data-cleaning, we unify identical entities found in both *shop* and *amenity* category and remove entities with stopwords (e.g., “wifi,” “bench”) from our data collection. In sum we extracted 37,955 entities for our 40 selected cities, making it an average of about 1,000 entities per city. Removing entities with stopwords leaves us at 26,384 entities for the following analysis. The remaining entities were classified with different categories for the type of business by the community. However, we still have 386 types of businesses of significantly varying sizes. Therefore, we use a k-means clustering [22] based on word2vec annotations [23] to further group these types of businesses. Word2vec can calculate a multi-dimensional vector from a word to quantify its meaning based on likeliness to other words in a dictionary [23]. We used a pre-trained model<sup>3</sup> as our dictionary to calculate the vectors. We input the resulting vectors for each business category into a k-means clustering algorithm, as we then need to cluster the categories into groups. Using an elbow score, we identified  $k = 20$  as a suitable number of clusters. Thus, combining these two algorithms yields 20 clusters of business categories with similar meanings.

For our computational analysis, we apply different techniques of machine learning to gather insights from our dataset, i.e., random forests, decision trees, and support vector machines. For each algorithm, we compare the results of analyzing highly-ranked and poorly-ranked cities to discover factors impacting the success of these cities.

## 4 Results and Discussion

During our analysis, we investigated different geographical features of businesses and their locations inside different city centers. First, the mean number of stores for each category is mostly the same for successful and unsuccessful city centers. This trend only differs for category 13 with a significantly high number of businesses for successful cities ( $\bar{O}=115$ ) compared to unsuccessful cities ( $\bar{O}=76$ ). This category covers gas stations and seaside businesses, comprising, e.g., fishing supply, scuba diving, weighbridges, and loading docks. Gas stations indicate high road traffic from possibly attracted customers to the city center. Thus, we interpret that some of these businesses are present in the city because many people are visiting particular cities, and not that many people are visiting the city for the high number of gas stations. As ports and fisheries are only possible at rivers and seas, the number of seaside businesses is mostly predefined and cannot be improved in many cases.

We analyzed all data we had available and tried to identify possible features significantly influencing the success of city centers. By using different machine learning methods and techniques (cf. Section 3), we are able compare different results

---

<sup>2</sup> <https://developer.mapquest.com/documentation/open/nominatim-search/>

<sup>3</sup> <https://code.google.com/archive/p/word2vec/>

and confirm the validity of the results. However, our data analyses show that the centrality of the locations of businesses in a city center is the main feature significant for the success of the city center. This was consistent across all applications of different machine learning techniques and methods. Figure 1 shows the average distance of locations to the city center divided by the category of the business. For the highly ranked cities, the mean distance of the locations varies from 1.0 km to 2.7 km depending on businesses' categories. In contrast, the mean distance of the locations from unsuccessful cities to the city center is significantly higher, ranging from 3.0 km to 3.5 km. Placing many shops close to the center of the city and fewer in a further range seems to be a factor drastically influencing the cities success.

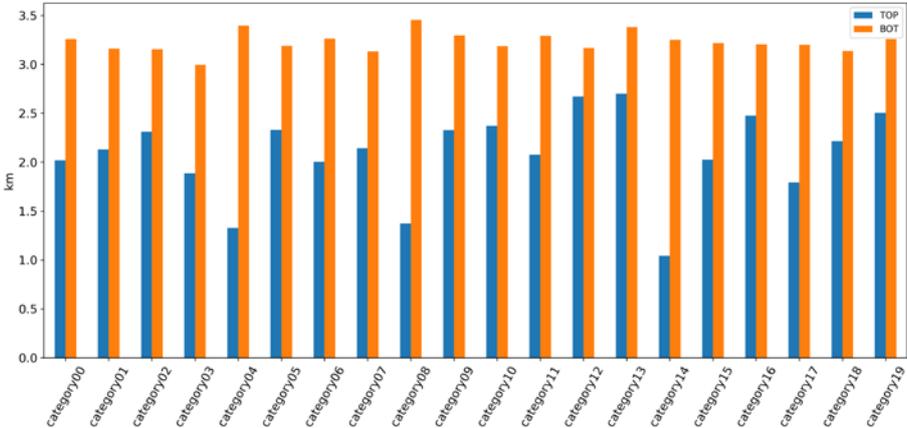


Figure 1: Average distance from the location to the middle of the city by category

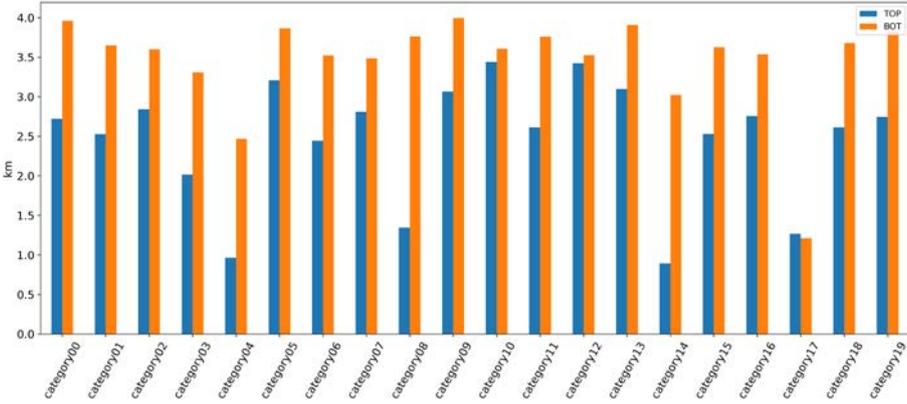


Figure 2: Average distance between locations of the same category in the same city by category

Similar results can be seen for the average distance between locations of the same category. Figure 2 visualizes these averages for every category of business locations. Again, average distances are smaller for successful city centers than average distances

of businesses in less successful cities. Generally, the average distance is lower for better-ranked cities. Our results suggest that centrality is one of the critical features of a successful city center.

With online retail threatening the existence of many retail businesses [3], [4], grouping different retail businesses (independent of their categories) seems to be a foundation for successful retail in city centers. Especially planned shopping areas [24], e.g., shopping malls and outlet centers, excel at pooling retailers on central locations because they do not have to consider natural, private, and governmental hindrances. Further, they are managed centrally in contrast to solitary retailers in organically grown shopping areas. Customers seem to prefer cities with shopping malls or similar structures. Thus, planning shopping areas with a dense presence of retailers could be a focus to improve the attractiveness and success of retail in city centers and to prevent city center retail to perish due to online retail. This does not need to be done in a shopping mall, but can be utilize vacancies in formally organically grown shopping areas. Town center management as a strategic way to improve city center attractiveness [25] can, e.g., benefit from recommending fitting areas to retailers who want to open new business sites in city centers [26].

## 5 Conclusion

In this paper, we used different methods of supervised machine learning to investigate patterns of retailers' locations in city centers. We discovered that the number of retailer businesses of different types does not influence the city center's success. Still, the centrality of successful cities is much higher, i.e., businesses are geographically closer to businesses of the same category. Our insights are limited by the selection of machine learning methods, the restriction to medium-sized cities in Germany, and by neglecting historical events transforming city centers (e.g., war damage, political agendas).

As our findings on distances between city centers and businesses as well as distances between businesses have to be contextualized, we plan to extend our research for considering further structural attributes, e.g., elevation, courses of roads, and public transportation networks. Our paper serves as an entry point for a detailed analysis of structural attributes in city centers and their influence on the success of retailers. As further analyses have to cope with an increasing volume of data, we consider machine learning techniques to be able to handle such analyses.

## References

1. Handelsverband Deutschland e.V. (HDE): Zahlenspiegel 2018, 42 (2018)
2. Handelsverband Deutschland e.V. (HDE): Umsatz im Einzelhandel im engeren Sinne in Deutschland in den Jahren 2000 bis 2019\* (in Milliarden Euro). Statista - Das Statistik-Portal, 2019 (2019)
3. Hagberg, J., Sundstrom, M., Egels-Zandén, N.: The digitalization of retailing: an exploratory framework. *Intl J of Retail & Distrib Mgt* 44, 694–712 (2016)

4. Verhoef, P.C., Kannan, P.K., Inman, J.J.: From Multi-Channel Retailing to Omni-Channel Retailing. *Journal of Retailing* 91, 174–181 (2015)
5. Verhoef, P.C., Lemon, K.N., Parasuraman, A., Roggeveen, A., Tsiros, M., Schlesinger, L.A.: Customer Experience Creation: Determinants, Dynamics and Management Strategies. *Journal of Retailing* 85, 31–41 (2009)
6. Jordan, M.I., Mitchell, T.M.: Machine learning: Trends, perspectives, and prospects. *Science (New York, N.Y.)* 349, 255–260 (2015)
7. Keating, W.D.: Linking downtown development to broader community goals: An analysis of linkage policy in three cities. *Journal of the American Planning Association* 52, 133–141 (1986)
8. Wegener, M.: Operational Urban Models State of the Art. *Journal of the American Planning Association* 60, 17–29 (1994)
9. Stable, E.G.: Die Einzelhandelszentralität westdeutscher Städte : Ein Beitrag zur Methodik der Zentralitätsmessung ( Retail Centrality of West German Urban Centres : A Contribution to the Methodology for Measuring Centrality ). *JSTOR* 1, 46–59 (1996)
10. Everling, O., Jahn, O., Kammermeier, E.: Rating von Einzelhandelsimmobilien. Gabler | GWV Fachverlage GmbH, Wiesbaden (2009)
11. Provost, F., Fawcett, T.: *Data Science for Business*. O'Reilly Media, Inc (2013)
12. Alpaydin, E.: *Introduction to machine learning*. MIT Press (2009)
13. Swain, P.H., Hauska, H.: Decision Tree Classifier: Design and Potential. *IEEE Trans Geosci Electron GE-15*, 142–147 (1977)
14. Rasoul, S., David, L.: A Survey of Decision Tree Classifier Methodology. *IEEE Transactions on systems, man, and cybernetics* 21, 660–674 (1991)
15. Breiman, L.: Random forests. *Machine Learning* 45, 5–32 (2001)
16. Pontil, M., Verri, A.: Support vector machines for 3D object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20, 637–646 (1998)
17. Berente, N., Seidel, S., Safadi, H.: Research Commentary—Data-Driven Computationally Intensive Theory Development. *Information Systems Research* 30, 50–64 (2019)
18. Müller, O., Junglas, I., Brocke, J., Debortoli, S.: Utilizing big data analytics for information systems research: challenges, promises and guidelines. *EJIS* 25:4, 289-302 (2016)
19. Neis, P., Zipf, A.: Analyzing the Contributor Activity of a Volunteered Geographic Information Project — The Case of OpenStreetMap. *ISPRS International Journal of Geo-Information* 1, 146–165 (2012)
20. Neis, P., Zielstra, D.: Recent Developments and Future Trends in Volunteered Geographic Information Research: The Case of OpenStreetMap. *Future Internet* 6, 76–106 (2014)
21. Mooney, P., Minghini, M.: A Review of OpenStreetMap Data. In: Foody, G., See, L., Fritz, S., Mooney, P., Olteanu-Raimond, A.-M., Fonte, C.C., Antoniou, V. (eds.) *Mapping and the Citizen Sensor*, pp. 37–59. Ubiquity Press (2017)
22. Kanungo, T., Mount, D.M., Netanyahu, N.S., Piatko, C., Silverman, R., Wu, A.Y.: An Efficient k-Means Clustering Algorithm: Analysis and Implementation. *Proceedings of the Sixteenth ACM Symposium on Computational Geometry*, 100–109 (2000)
23. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient Estimation of Word Representations in Vector Space, 1–12 (2013)
24. Gilbert, D.: *Retail marketing management*. Financial Times Prentice Hall, Harlow (2002)
25. Coca-Stefaniak, J.A., Parker, C., Quin, S., Rinaldi, R., Byrom, J.: Town centre management models: A European perspective. *Cities* 26, 74–80 (2009)
26. zur Heiden, P., Berendes, C.I., Beverungen, D.: Designing City Center Area Recommendation Systems. In: Gronau, N., Heine, M., Poustecchi, K., Krasnova, H. (eds.) *WI2020 Zentrale Tracks*, pp. 506–521. GITO Verlag (2010)