ESKİŞEHİR TECHNICAL UNIVERSITY JOURNAL OF SCIENCE AND TECHNOLOGY
A- APPLIED SCIENCES AND ENGINEERING

16th Digital Design In Architecture Symposium
16th DDAS (MSTAS) - Special Issue 2022

2022, Vol.23, pp. 86-93, DOI:10.18038/estubtda.1170105

WHAT MAKES A STREET WALKABLE? A DATA ANALYTIC APPROACH TO INVESTIGATING WALKABILITY FACTORS

Nur SİPAHİOĞLU *

Department of Architecture, Yaşar University, Izmir, Turkey

ABSTRACT

Walkability is a hot topic for variety of disciplines, as well as everyday walker. It affects the health, the environment and the liveliness of our neighbourhoods. Walkable streets are necessary for a better lifestyle and sustainable planet. The problem with walkability is that we still don't have a general understanding of the concept. Every study differs in the way they define walkability, thus making walkability a subjective topic. However, the subjectivity causes contradiction in science. In this study, the aim to answer the question of what makes a street walkable by using a data analytic approach. The features used in other studies, as well as new attributes specific to this study, were investigated. Street images were used to extract data. The data was divided into nine categories: Street, Sidewalk, Obstacles, Urban Blocks, Amenities, Transportation, Attractiveness, People, and Vehicles. Data collection was carried out by measuring physical attributes through Remote Sensing images in QGIS, visually analyzing qualitative attributes with Google Street Maps/View and double checking data in Open Street Map Overpass Turbo API. Attributes were translated into scores and normalized where possible. Mutual Information Matrix and Correlation processes were conducted in Rapidminer. The attributes were processed in relation to overall assessment of walkability which was defined with personal rating. As a result, Mutual Information and Correlation matrices are useful in figuring out the relationship and dependencies between different attributes. Applying data analytics to a more comprehensive dataset will help identify the global factors of walkability.

Keywords: Walkability, Data mining, Correlation, Mutual information

1. INTRODUCTION

Walkability has been a topic of interest for the past 20 years. Even though it may be considered an urban design problem at its core, walkability concerns many others from different disciplines from health experts to sociologists; most importantly, the everyday walker. Walkable streets are more than an indicator of healthier lifestyles and livelier neighbourhoods. It provides solutions to the greater environmental problems. Jeff Speck emphasizes that living in walkable cities is greener than living in a sustainable gadget filled house [1]. That's one of the many reasons walkability is crucial to understand and apply to our streets.

The problem of walkability is the lack of general consensus on what it is really. Each study identifies and explains walkability depending on the properties they deem more relevant; physical, environmental, social, and such. However, in the absence of a general definition, studies start to contradict each other. Each study presents valuable contributions to what makes a street walkable, yet they may be in disagreement [2]. Krambeck proposed a global walkability index in a graduate thesis [3]. There are studies that acknowledge the issue from the pedestrian viewpoint [4]. Another study points out that usually the macro-level characteristics (urban blocks, zoning, density) are studied, neglecting the micro-level characteristics (sidewalks, trees, furniture) [5]. Walk21, a walking movement organization, defines the reasons why people avoid walking in certain streets and to some locations as heavy traffic, crime rates, street cleanliness, lack of amenities, quality of pavements, and street lighting (https://walk21.com/). However, it is not possible to deduce walkability to only economic factors (i.e. shops, or malls). Walkability of a street also depends on well-connected streets,

*Corresponding Author: nur.sipahioglu@yasar.edu.tr
Received: 02.09.2022 Published: 23.12.2022

buildings that are human-scale, building density and population, wide sidewalks, trees and greenery [6].

Different studies have used variety of metrics to identify the walkability of streets on a scale. Walkscore (https://www.walkscore.com/) is one of these metrics which considers block length, intersection density, transit score, bike and rail stops, errands, culture, grocery, park, dining, school and shopping scores. Walkscore awards points according to the closest amenities. But it doesn't differentiate amenities. It also doesn't introduce sidewalks, cars, lanes or crime. Walkonomics looks at the issue from a different perspective: Photos shared on social media. It compares the photos and the rated street segments. This comparison proves further the idea that fewer cars equal more walkability. Crime areas are photographed less at night. There is a positive correlation between photos tagged with "'sidewalk', 'clean street', 'tree' and 'architecture'" and pedestrian friendly streets [7]. More studies try to identify and measure walkability with different methods: Measuring qualitative aspects of walkability with the help of an expert panel [8,9]; evaluating walkability aspects selected by experts by the ratings of pedestrians [10]; combining different methods for easier and more accessible data analysis [11]; integrating GIS-based methods [12,13].

The studies summarized suggest the importance and the relevance of walkability. However, as can be seen from the limited literature mentioned above, there is still no unity in its definition. Walkability is not in dictionaries like Merriam Webster and Oxford. It is not found in dictionaries of other languages, such as Turkish. The subjectivity of walkability remains. This study asks the still unanswered question: What makes a street walkable? By utilizing data mining methods such as mutual information and correlation matrices, the study aims to answer the following questions:

-What features of walkability affect the personal assessment?
-What is the relationship between walkability attributes with regards to personal assessment?
-Which walkability attributes can be the defining factors of walkability?
-What kind of data can be extracted from street images?

2. MATERIALS AND METHOD

In order to answer the data analytic questions, the study follows a process of data collection, preparation, processing, modeling, and interpretation (Figure 1). Firstly, the attribute data was designed according to the literature review. General attributes used in most of the studies and more customized attributes such as "store owners" were used. Data was collected by measuring physical attributes through Remote Sensing (RS) images in QGIS, which is an open source geographical information system software, visually analyzing qualitative attributes with Google Street Maps, and double checking data in Open Street Map Overpass Turbo API. After a normalization process, Mutual Information Matrix (MIM) and Correlation Matrix (CM) processes were conducted in Rapidminer. MIM and CM reveal positive and negative relationships between each attribute, and their dependencies.

Figure 1. Data analytic process

Deriving from the range of studies mentioned in the previous section the walkability data were gathered under nine categories: Street, Sidewalk, Obstacles, Urban Blocks, Amenities, Transportation, Attractiveness, People, and Vehicles. Each category was also divided into different attributes. Personal rating of walkability was described as the overall assessment which made up the label attribute that other attributes would be associated and compared to. Since there were a lot of attributes, numeric data was normalized to [0, 1] and a scoring system was introduced where possible. This proposed method was applied to the streets of Izmir as case study. In the scope of this study, there were 70 instances and 12 attributes. Instances refer to individual examples which are the streets selected for the case study. Attributes are the characteristics of an instance and they are either nominal or numeric. The attributes of each walkability category are shown in Table 1, which reflects the list of attributes before any reduction. As this study covers a very limited dataset of 70 instances, attributes were condensed into 12 by rescaling, normalization and/or scoring. Some attributes are left out as they had no impact in this study. Attributes indicated in italics indicate temporal data: Data obtained at a specific date and time. Overall assessment is the label attribute. The twelve attributes are Sidewalk Width Normalized, Crossing per 100m, Obstacle Score Normalized, Height category, Facade transparency, Amenity per 100m, Attractiveness Score, People Score Normalized, Store Owners, Lane Width Normalized, Cars, and Bikes.

Table 1. Walkability data categories and attributes

Street	Name Length *Walk duration*	id How long is the street? *meters* How long does it take to walk the whole street? *minutes*
Sidewalk	Width Crossings (per 100m)	What is the width of the sidewalk? How often is the sidewalk cut by other streets?
Obstacles (Score)	*Trash* *Barriers* *Store Stuff* *Street Vendors* Signage *Parking on Sidewalk*	Are there any trash or trash cans? Are there any constructions, road cones, parking signs? Are there any tables, chairs, umbrellas? Are there any (mobile) street vendors? Are there any store signage, advertisements, and traffic signs? Are there any cars occupying the sidewalk?
Urban Blocks	Height Façade Transparency	What is the average floor count of building blocks? transparent, semi-transparent, non-transparent
Amenities (per 100m)	Education Retail Medical Financial Social Leisure	Schools, university, learning centers Convenience stores, markets, shops Pharmacy, hospital, clinics Banks, post office Barbers, gym, library, cinema, restaurants, cafes Parks, Squares
Transportation	Bus stops Metro/Tram Station	How many bus stops? How many metro/tram stations?
Attractiveness (Score)	Trees *Shade* Lighting *Cleanliness*	Are there any trees? Are there any shading elements (tents, big trees)? Are there any street lights? No trash on the ground. Clean or not?
People (Score)	*Age* *Action* *Store Owners*	Young, Adults, Seniors walking, stalling, shopping, eating, sitting, playing Are there anyone we can identify as store owners?
Vehicles	Lane Width *Cars* *Bikes*	What is the width of the lane used by vehicles? Are there any cars? Are there any bikes?
Overall Assessment	Personal Rating	walkable (5) to non-walkable (1)

Transportation attributes were not used as there was no specific difference between the instances. Street length and walk duration were left out due to not being defining factors in this case. Obstacles, amenities, attractiveness and people attributes were scored according to their subcategories. Crossings and amenities were rescaled to "per 100m", and Obstacle Score and Attractiveness Score were calculated as one point for each true (i.e. if there is trash, then true. Add 1 point.). People Score was calculated as the number of age groups multiplied by the number of actions. For example, if there were two age groups on that street (young, adults) and there were six different activities, two was multiplied by six. Building heights were labeled in three categories: Lowrise (1m to 3m), Multistorey (3m to 7m), and Midrise (7m to 9m). After all the categorization and normalization were finalized, the data was processed and modeled in Rapidminer (Figure 2). All the attributes with relation to overall assessment (of walkability) were processed with CM and MIM; the attributes of Amenities, Obstacles and Attractiveness individually with relation to overall assessment (of walkability) were processed further with MIM.

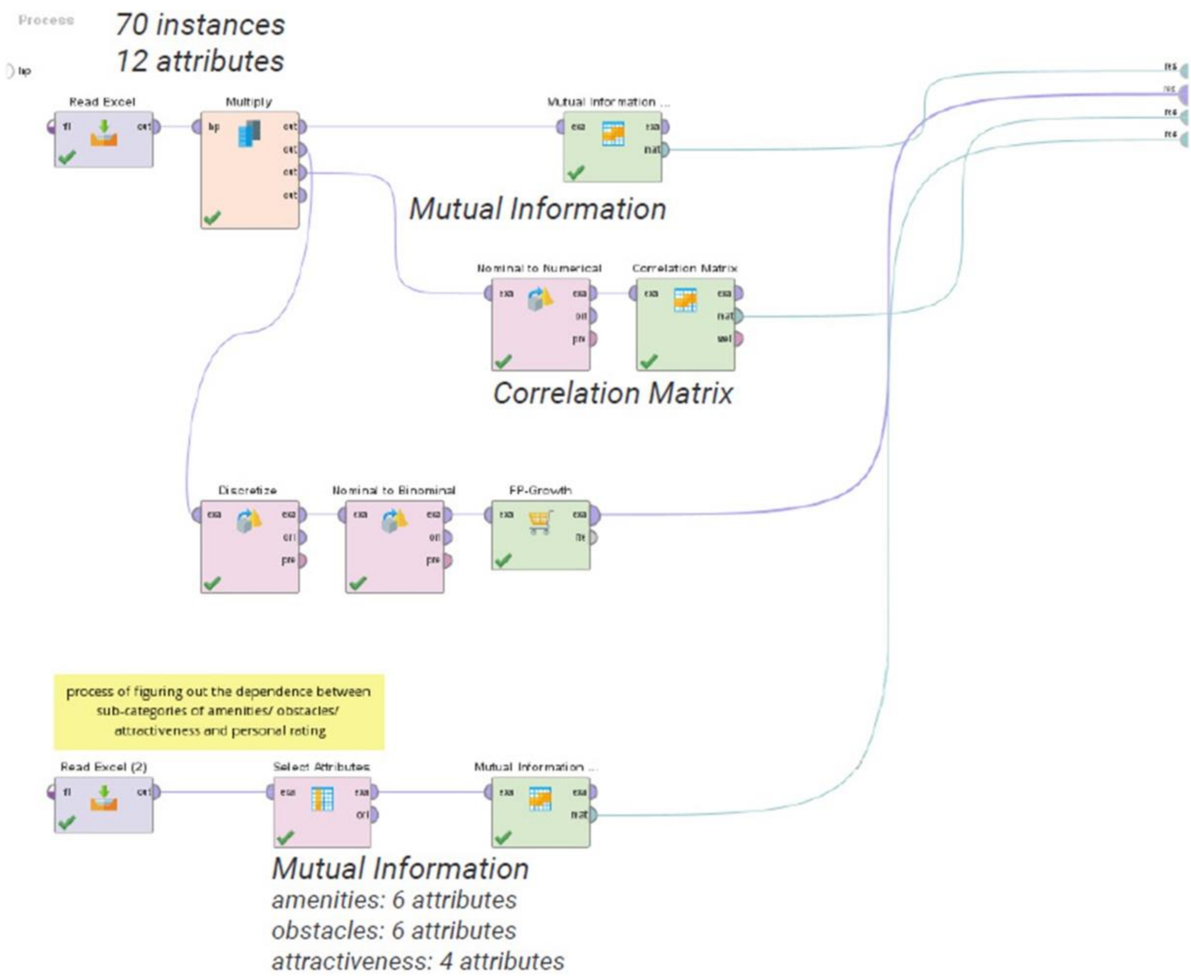

Figure 2. Rapidminer process

3. RESULTS

Results of MIM suggests that lane width, obstacle score, sidewalk width, amenity per 100m and people score tell us the most about personal rating of walkability (Table 2). Other relations uncovered with MIM show that sidewalk width is linked to obstacles and amenities, obstacles more linked to amenities which can be deduced as more amenities mean more store stuff on the sidewalk. Amenities are also linked to lane width, and this tells us that the larger streets may attract more amenities. Façade transparency is linked to amenities the most and it can be explained by amenities having a glass façade where private buildings such as housing don't allow us to peek inside. "Store owners" is an attribute that is interesting to analyze, especially in Turkey where shopkeepers like to spend time socializing outside their shops. Consistently, this attribute is related to amenities and façade transparency.

Table 2. Mutual information matrix based on personal rating of walkability.

Attributes	Personal Rating
Lane Width Normalized	0.653
Obstacle Score Normalized	0.626
Sidewalk Width Normalized	0.611
Amenity per 100m	0.564
People Score Normalized	0.506
Crossing per 100m	0.435
Cars	0.296
Attractiveness Score	0.258
Bikes	0.209
Façade Transparency	0.206
Store Owners	0.131
Height category	0.117

Results of CM confirm the results of MIM. When MIM and CM are compared side by side in terms of personal rating of walkability, the conclusion is that sidewalk width, amenity and people have positive correlation and more dependence; car-free streets, bikes and transparent façade have positive correlation with less dependence; lane width and obstacle score have negative correlation with more dependence (Figure 3).

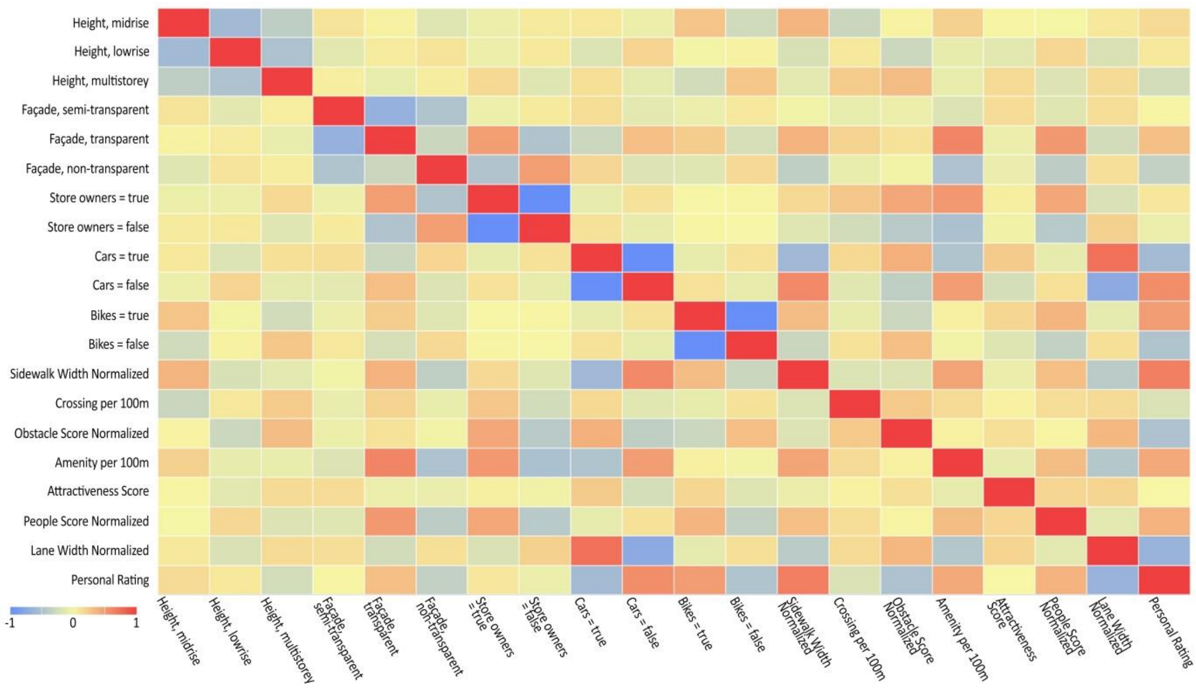

Figure 3. Correlation matrix where blue indicates negative and red indicates positive.

The dependence between the sub-attributes of amenities shows that there is no significance in the kind of amenity there is on the street with regards to walkability. Only retail and financial attributes show stronger dependence which may be speculated as more shopping equals more need for cash or banks tend to cluster around economically active areas (Table 3).

Table 3. Mutual Information Matrix of amenities

Attributes	Personal Rating	Educational	Retail	Medical	Financial	Social	Leisure
Personal Rating	2.228	0.327	0.380	0.270	0.332	0.341	0.227
Retail	0.380	0.243	2.272	0.511	0.774	0.433	0.236
Social	0.341	0.317	0.433	0.327	0.595	1.923	0.181
Financial	0.332	0.366	0.774	0.594	2.014	0.595	0.231
Educational	0.327	1.376	0.243	0.257	0.366	0.317	0.085
Medical	0.270	0.257	0.511	2.201	0.594	0.327	0.068
Leisure	0.227	0.085	0.236	0.068	0.231	0.181	1.242

In terms of obstacles; parking on sidewalks, trash and signage have a more significant effect on walkability rating (Table 4). Attractiveness, on the other hand, relies more on cleanliness (Table 5).

Table 4. Mutual Information Matrix of obstacles

Attributes	Personal Rating	Trash	Barriers	Store Stuff	Street Vendors	Signage	Parking on Sidewalk
Personal Rating	2.228	0.363	0.175	0.033	0.171	0.323	0.458
Parking on Sidewalk	0.458	0.149	0.194	0.009	0.058	0.223	0.913
Trash	0.363	0.844	0.054	0.021	0.025	0.075	0.149
Signage	0.323	0.075	0.099	0.004	0.008	0.722	0.223
Barriers	0.175	0.054	0.692	0.054	0.013	0.099	0.194
Street Vendors	0.171	0.025	0.013	0.041	0.913	0.008	0.058
Store Stuff	0.033	0.021	0.054	0.627	0.041	0.004	0.009

Table 5. Mutual Information Matrix of attractiveness

Attributes	Personal Rating	Trees	Shade	Lighting	Cleanliness
Personal Rating	2.228	0.097	0.088	0.072	0.158
Cleanliness	0.158	0.053	0.003	0.084	0.881
Trees	0.097	0.863	0.024	0.089	0.053
Shade	0.088	0.024	0.371	0.010	0.003
Lighting	0.072	0.089	0.010	0.422	0.084

4. CONCLUSION

As the results show, many features affect the personal assessment of walkability. According to the data analytic process lane width, obstacles, sidewalk width, amenities and people on the street affect walkability significantly more. Comparing the attributes that have positive correlation with personal assessment, bikes and façade have less dependence whereas sidewalk width, amenities and people have more dependence. The relationship between walkability attributes is more intricate. Both MIM and CM uncovered different relationships between attributes in the scope of this study as mentioned in results; however, larger dataset might help unpack these relationships even more. The results show that lane and sidewalk width notably affect walkability, yet they can't be labeled as defining factors. Many other attributes have an impact on the walkability assessment. Amenities, shade, cleanliness might make a street with wider lanes more walkable compared to a street with ideal lane and sidewalk width with nothing to do and nowhere to rest. Data in this study was mostly extracted by simple observation of street images. It was possible to find out many features of a street via images in a plot study like this. As humans we can recognize and understand images, categorize shapes and things, make judgments regarding the image in question. However, bias and interpretation might hinder the data extraction process by human gaze. Computer vision will be better suited for decreasing bias and speeding up the process in a larger dataset.

Walkability is subjective. Attributes change and/or transform in every other study. By using data mining methods, as seen in this study, we are able to figure out the relation between our understanding of what walkability is and its characteristics. In order to work with as many attributes as in this study, we need a huge amount of data. More data will help uncover more relations under a complex phenomenon such as the city. It will be easier to expose different dimensions of walkability. But simply more data is not the best way to go. More data may also cause noise which is a problem that requires more time spent on preparation and preprocessing. As the amount of data increases, the need for automation rises. In terms of extracting data from images like in this study, automation where possible (such as image processing and recognition) will make the data collection and analysis easier than manually filling the data table.

This is a pilot study which gives us clues on how to combine data mining methods with the subjectivity of walkability problem. The aim is to develop this study further with more attributes that haven't been included due to limitations in data collection and construction. The attributes planned to be include are listed, but not limited to, as follows:

- Quality of pavement: Height, material, no hazardous pits.
- Slope of the street: Is it too steep to walk?
- Children: Is there a correlation between walkability and the number of children spending time / playing on the street?
- Stray dogs: Despite being members of our communities, sometimes a large number of dogs is scary for many people.
- Tram lines: Just because a street is closed to car traffic doesn't mean there is no danger from other means of transport.

ACKNOWLEDGEMENTS

I'd like to offer my sincere gratitude to Ahu Sökmenoğlu Sohtorik (PhD) and Ceyhun Burak Akgül (PhD) for their guidance and encouragement, and to my fellow CaaDM Collective colleagues who have been showing great enthusiasm and effort for the continuation and expanding the limits of walkability studies.

CONFLICT OF INTEREST

The author stated that there are no conflicts of interest regarding the publication of this article.

REFERENCES

[1] Speck J. Walkable City. 1st ed. New York, NY, USA: North Point Press, 2012.

[2] Forsyth A. What is a walkable place? The walkability debate in urban design. Urban Des Int 2015; 20(4): 274–292.

[3] Krambeck H V. The Global Walkability Index. MCP, Massachusetts Institute of Technology, Massachusetts, USA, 2006.

[4] Lo R. H. Walkability: What is it? J. Urban 2009; 2(2): 145–166.

[5] Tiemann T K, Scott, A C, Atkins K N. Sidewalks, Streets and Walkability. Spaces Flows Int J Urban Extra Urban Stud 2012; 2(3).

[6] Campoli J. Made for Walking: Density and Neighborhood Form. 1st ed. Cambridge, Massachusetts, USA: Lincoln Institute of Land Policy, 2012.

[7] Quercia D, Aiello LM, Schifanella R, Davies A. The Digital Life of Walkable Streets. Proceedings of the 24th International Conference on World Wide Web, 2015.

[8] Ewing R, Handy S. Measuring the unmeasurable: Urban design qualities related to walkability. J Urban Des 2009; 14(1): 65–84.

[9] Ewing R, Handy S, Brownson R C, Clemente O, Winston E. Identifying and Measuring Urban Design Qualities Related to Walkability. J Phys Act Heal 2016; 3(1): 223–240.

[10] Moura F, Cambra P, Gonçalves A B. Measuring walkability for distinct pedestrian groups with a participatory assessment method: A case study in Lisbon. Landsc Urban Plan 2017; 157: 282–296.

[11] M, Hagen O H, Tennøy A, Rynning M K. Exploring ways of measuring walkability. Transp Res Procedia 2019; 41: 264–282.

[12] Majic I, Pafka E. AwaP-IC—An Open-Source GIS Tool for Measuring Walkable Access. Urban Sci 2019; 3(2): 48.

[13] D'Orso G, Migliore M. A GIS-based method for evaluating the walkability of a pedestrian environment and prioritised investments. J Transp Geogr 2020; 82: 102555.