Consumer attraction centers in the Metropolitan Zone of the Valley of Mexico. Approach based on data from social networking sites (Twitter)

Abstract

The purpose of this work is to use information from Twitter to investigate the formation of consumer attraction clusters. This work closes the knowledge gap in the study of consumption in Latin American cities with the use of big data and computational science methods. Thus, the work identifies two main types of attraction centers for consumption in the Metropolitan Zone of the Valley of Mexico(MZVM): first, centers that attract consumers from very close distances and others that attract those who travel long distances. Second, centers are characterized by their economic specialization to differentiate the motives for consumption.

Keywords: Consumer clusters, Twitter and Big Data, Machine Learning, Urban economics, consumer data, Mexico City

Introduction

There is a boom in the analysis of consumption in cities (Dolega et al., 2019; Dolega & Lord, 2020). The increase in the consumption of intangible goods and services in contemporary cities has become relevant in recent decades due to the processes of the decentralization of industry, the explosion of the middle classes (Alonso, 1970; de Simone, 2018; Jayne, 2006), urban renewal processes and the creation of mobility infrastructure (Anas, 2007; Massey, 2005). Despite this boom, there is a gap in knowledge about many of the different dimensions of consumption in urban spaces and their implications.

Consumption is the second most important travel attractor in the Metropolitan Zone of the Valley of Mexico (MZVM) after commuting, and many of the instances of consumption are generated in travel segments between home-work or work-home. Additionally, it is important to point out that the economic units dedicated to commerce and services to people exceed 35% of the total economy. That is, one out of every three units is allocated to commerce or to a business that provides services directly to people.

Regarding mobility approaches, data availability has been an issue that prevents decision-making. Although it is true that the first studies were based on transport routes and influx (per hour, day, week, month or year), more information has gradually become available that allows more precise analyses and decision-making practically in real time. However, this information is not available for all cities and social groups. For instance, surveys, video surveillance cameras, and traffic detectors, among others, allow establishing the reasons for these decisions, and the schedules and flows are much more accessible in cities of developed countries. This is where another type of information generated by people, which allows knowing with certain precision the dynamics of mobility and use of public space, can be useful.

Large amounts of data (Big Data) have become fertile ground for many disciplines and governments to fill the lack of information. The use of data to identify spaces and dynamics in city public spaces (Martí et al., 2017) is an interesting aspect, as is knowing the importance of infrastructure that allows accessibility to all people.

Consequently, by using an approach that unites mobility, data and management, there is the potential to address the issues of urban space planning, transport policies (Lloyd et al., 2018) and accessibility by identifying consumption patterns and the mobility associated with them.

Consumption in the metropolitan area of the Valley of Mexico

In the case of Mexico City and the Metropolitan Zone of the Valley of Mexico (MZVM), the phenomenon of consumption has been studied from the approaches of urban geography, sociology, and urban economy. These studies can be outlined in three aspects: gentrification and public space, shopping centers and self-service stores, and cultural elements of the consumption of goods and services.

The first approach links the gentrification process, particularly economic gentrification, with the patterns of consumption of services focused on the middle classes who live in middle and upper class neighborhoods, such as Polanco or Condesa. Similarly, urban elements of mixed land use have been reported, where residential properties are mixed with spaces aimed at selling elite goods and services (Olivera-Martínez, 2017a, 2017b).

Supermarkets have been identified as one of the most important elements of consumption. Consumption in supermarkets has also brought to the fore when examining issues such as marketing chains, habits, and the impacts on population nutrition (Durand, 2015; Orozco-Hernández & García-Luna-Villagrán, 2014).

The cultural aspects of consumption form the last approach. The focus of this topic has been how the socioeconomic strata and social configuration of urban space play a role in attracting differentiated people based on cultural habits. For example, the literature on this approach has studied cultural aspects of elite consumption in shopping centers, consumption in alternative markets, in particular, and hipster goods or intangible goods such as jazz music (Duhau & Giglia, 2016; Vargas Magaña, 2022).

In this context, almost all the studies previously mentioned have used qualitative and case study methodologies. This can probably be explained by the complexity of the consumption phenomenon on a fine grain spatial scale. At the same time, general studies have faced the problem of the limited space-time granularity of official data sources. In this way, one of the objectives of the work is to close the gap by offering a fine-grained analysis of consumption and its associated mobility in the MZVM based on alternative sources of information.

Theory

Consumption in cities: a territorial approach

Consumption in cities has been studied from the perspectives of different disciplinary areas, mainly social and economic ones. In these approaches, the role of the individual or the household level has been emphasized in terms of the consumption decision (Tang et al., 2020) or in the supply and availability of individuals to move and thereby satisfy a demand. The main theory from regional science and urban geography in this field is the Central Place Theory and its revisions (Christaller, 1966; Mulligan et al., 2012; Tabuchi & Thisse, 2011; van Weerdenburg et al., 2019).

In the field of geography, the consumption phenomenon has been analyzed from the perspectives of two subdisciplines: economic geography and urban geography. The first has tried to determine how individuals make decisions on specific aspects. In the case of consumption, mobility behaviors and constraints have been modeled, for example, by discovering the maximum distance that an individual can/wants to travel to satisfy a need. Some conclusions have been reached that are, at first glance, obvious, but that are nonetheless essential to understand how a certain level of supply and how restrictions (sex, age, income, race, among others) affect the type and level of consumption in cities.

The second approach is based on urban spatial economy models that use the location elements of companies and how they determine consumption, and this is how certain valleys or supply deserts are created. These types of analyses are common in urban economics and economic geography studies that seek to find agglomeration economies. For this reason, taking advantage of location is central in the supply of either similar goods (for example, specialization) or diverse goods (urbanization economies).

Mobility, consumption, and big data

Big data is a relatively recent source for urban studies, particularly in the aspects such as transportation, mobility, taste, sustainability, and so on. It has been widely believed that the availability of information can lead to better decisions, and they can be made practically in real time. The data generated by the governments themselves (through specific apps) or by private companies (that do not have a specific purpose in urban matters) are usually seen as sources that allow the knowledge of specific phenomena and that can be modified (or reinforced) in the short term.

The business nature of big data usually entails the challenge of inferring certain characteristics that are elementary dimensions in the academic field, such as sociodemographic or economic characteristics. Therefore, the first challenge when using big data is the repurposing of big data obtained from private sources to answer social research questions (Salganik, 2018). In this sense, the implementation of novel methods and techniques from computational sciences allows obtaining robust approximations that allow social researchers to answer social questions. In these cases, the mobility and consumption patterns in cities are identified.

The study of everyday mobility in the urban environment has been concentrated on the commuting motive (Hipp et al., 2022). The core information of these studies has been statistical built surveys focused on mobility phenomena. The paradigmatic case in Mexico is the Origin-Destination Survey given to households in the Metropolitan Zone of the Valley of Mexico, which offers data on the subject for 2017 (INEGI, 2018).

In contrast, the massive use of smartphones, Wi-Fi connections and other digital devices in cities offer new data, and reuse strategies have emerged to approximate daily mobilities for a variety of motives (Milusheva et al., 2020). For example, mobility for consumption purposes has been estimated by the frequency of the location of mobile phones or records of access to open Wi-Fi networks made by mobile phones (Anenberg & Kung, 2015; Miyauchi et al., 2021; Murcio et al., 2018). Another approach has been carried out through the analysis of loyalty cards or purchases with credit cards (Lloyd et al., 2018).

Another novel data methodology that links both dimensions, mobility and consumption, has been the use of data and metadata of social network sites or websites focused on the evaluation of places of consumption, such as Foursquare, Yelp, or even Google Maps. In this regard, through these platforms, consumption segregation has been studied (Davis et al., 2019), and the processes of gentrification have also been studied. This is due to the nature of the platforms for user reviews of specific places of consumption, such as restaurants. Consumption of cultural goods and services, such as museums or parks (van Weerdenburg et al., 2019) and identification of consumption patterns associated with sustainable practices on Twitter have also been noted (Brzustewicz & Singh, 2021).

Material and method

As mentioned, the rising interest in consumption phenomena in cities has been, to a large extent, due to the availability of fine-grained spatiotemporal data. In this sense, Twitter has been little explored as a source of information on consumer issues in general in international literature, particularly in Latin American cities.

Twitter data

For the present work, the free Twitter API has been used to harvest data in its streaming modality, specifically, tweets from within the polygon that makes up the MZVM.[1]. The collected tweets cover the period from March 2017 to October 2021. Based on the collection algorithm, all the metadata of the tweets have been extracted, although only 5 fields have been recovered: timestamp, user, text, latitude and longitude. These variables allow us to carry out the analyses of identification of consumption, identification of the home block and mobility toward consumption. The total base comprises approximately 25 million tweets, of which only 2.7 million are georeferenced.

---

[1]The coordinates of the polygon are at the southwest point: -99.654.18.922 and at the northeast point: -98.609.20.067

Identification of consumer tweets and household blocks

The text in the tweets can provide information about the users' consumption behavior and places where they consume. To identify tweets related to consumption, the naive Bayes classifier (NBC) is carried out. This algorithm has been trained to classify twitter data. First, one database with human supervised labels is created and, second, one with a semisupervised database based on text markers in tweets.

The NBC algorithm is a classifier that, due to its simplicity and predictive power, is widely used in text classification in different contexts. In the case of economics and urban studies, it has been used to carry out sentiment analysis during the COVID-19 lockdown (Yao et al., 2021), for the modeling and forecasting of urban expansion and land use (Xiao et al., 2021) or to study differences in traffic lights(Xu et al., 2018) .

The NBC algorithm is based on the bag-of-words principle and assumes independence between the documents that make up the corpus. Formally, the classifier is expressed as follows: where fi is the i-th word of the total of n words on tweets classified as consumer or nonconsumer and multiplied by the probability of finding consumer tweets in the total corpus. The formal expression of NBC is:

$$Y^t = \underset{Y_k}{\operatorname{argmax}} P(Y = y_k) \prod_i P(f_i \mid Y = y_k) \quad (1)$$

For the study, a training set built by means of semisupervised and supervised sets of tweets with reference to consumption is implemented. In the first instance, the semisupervised training set was built with a text marker found in the 2.7 million georeferenced tweets. This text marker is the *I'm* at text pattern, which is a trace that the Twitter platform assigns to a user's publication when performing a georeferenced (check-in) in a particular place (Fig. 1).

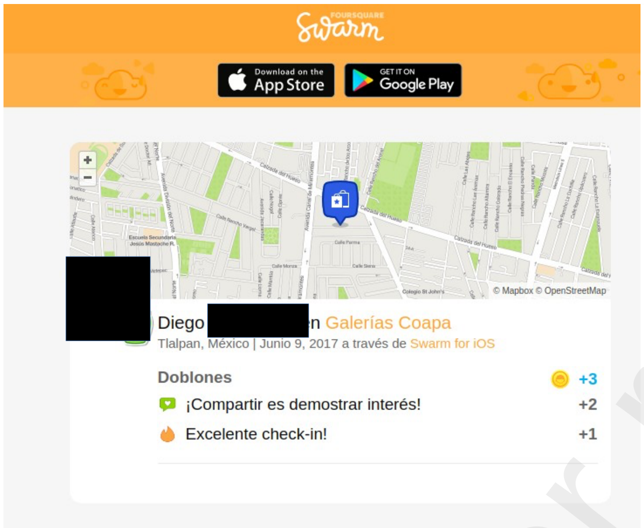

Fig. 1. Example of a georeferenced tweet with a text registration tag.

Source: author's image

These text markers can offer nonsemantic information about consumption patterns. It has been found that 20% of georeferenced tweets have this tag and contain no text. After this identification, a word frequency analysis is carried out to identify consumption/nonconsumption named entities on the corpus. On the one hand, there are named entities related to nonconsumption practices, e.g., public places or boroughs, of Mexico City. Due to the text markers, these tweets have been labeled nonconsumption tweets. On the other hand, name entities related to service establishments, restaurants, shopping centers or concert venues can be found. These tweets have been labeled with consumption tags. This semisupervised method to label tweets has allowed a greater set to train the NBC.

The second training set for NBC consists of supervised tweets. At this point, a random sample of 2,000 tweets is retrieved from tweets without a text marker, and consumption practices such as attendance at restaurants, parties, shopping or visits to shopping malls are identified. In sum, both sets of tweets make up the training set to identify consumption in georeferenced tweets.

The identification of home blocks based on tweet emission has been carried out using a mixed strategy that combines previous results of the NBC and official population on urban blocks of the 2020 Population and Housing Census. First, tweets with nonconsumption markers are taken and an extra label is added based on their timestamp, in particular for emissions at night (0–6 h) on business days (Monday to Thursday). Second, tweets with two previous filters identify the most frequent block with emissions with a condition of 10 or more inhabitants in that block. This last criterion has been implemented to eliminate, as much as possible, the possibility that it is a broadcast in a block without any housing and that, therefore, the tweets are categorized as referring to other practices in the city at night.

Identification of the economic specialization of blocks

The second part of the methodological process is to identify the economic specialization of the block. This task allows us to give a sectorial interpretation of consumption based on Twitter. For this task, the National Statistical Directory of Economic Units (known as DENUE in Spanish) from March 2021 is used. Based on the classification of the North American Industrial Classification System (NAICS), trade and services are grouped into 13 categories described in Table 1.

Service	General description	Number of branches
Wholesale	Wholesale trade of different items, from food, clothing, pharmaceuticals, appliances, etc.	18
Retail	Retail trade of different items, from food, clothing, pharmaceuticals, appliances, etc.	15

Conspicuous consumption	Trade in automobiles, motorcycles, perfumes, and entertainment items	5
Entertainment services	Parks with recreational facilities or casinos	3
Financial services	Central banks, exchange houses and financial advisory institutions	5
Cultural services	Audiovisual industry, museums, theater companies, arts	5
Care services	Nursery services, work and social orientation, asylums, and residences for elderly individuals, for people with intellectual disabilities and rehabilitation problems	7
Higher education services	Higher education services: undergraduate, postgraduate and higher technical	2
General services	Basic education services, personal, multiple banking, cleaning, rent, among others	53
Logistics services	Transport services for people and goods, mail, and packaging	26
Medical services	General and specialty hospitals, medical and dental offices	10
Professional services	Design, consulting, accounting, software editing, corporate, scientific, and technical services, etc.	22
Travel services	Services related to tourist activity, hotels, camps, food, and beverage preparation, among others	11

Table 1. Grouping of trade and services and number of branches grouped in each category.

Source: own elaboration based on DENUE 2021

Once commerce and service activities are grouped into these categories, a location quotient (LQ) is calculated for each activity group at the block spatial level. Classically, a spatial unit with an LQ greater than 1 in a given activity is considered to be specialized in that activity. However, to grant it a single specialization activity, it was decided to reclassify the category with the maximum value of the LQ at the block level. Thus, in formal terms, the economic specialization of the block is expressed as follows:

$$QL_i = \text{argmax}\left(\frac{\frac{eu_{ij}}{EU_j}}{\frac{EU_i}{EU}}\right) \quad (2)$$

where:

Eu_ij is the number of establishments in block i and in service group j; EU_j is the number of establishments of group j in the total MZVM; EU_i is the total number of service establishments in block i; and EU is the total number of establishments in the universe of blocks of the MZVM. Given that each block i can report an LQ in an interval of [0, ∞), it is decided to assign the group that reports the maximum LQ of the 13 groups.

Identification of attraction clusters

Once the possible home blocks and the economic specialization of the blocks where a consumption is made have been identified , the average distance that Twitter users travel from their home to the consumption blocks is calculated according to the type of economic specialization that they identified in each of them. In this way, the average distance is calculated as follows:

$$\overline{D_c} = \frac{1}{\eta_{CH}}\sum_{h=1}^{\eta_{CH}} d_{ch} \quad (3)$$

where:

$$d_{ch} = \sqrt{(lat,lon)_c^2 + (lat,lon)_h^2} \quad (4)$$

Equations 3 and 4 imply the average Euclidean distance of all identified consumer tweets in the block from the user's home block. In other words, an approximate variable of the level of attraction that each point of consumption has can be obtained.

Finally, clusters of attraction toward consumption blocks are identified according to the distance traveled by Twitter users and if these cores correspond to certain patterns of the urban structure of the MZVM. For this task, the local index of spatial autocorrelation (LISA) has been implemented to identify the core of attraction toward consumption based on each block of consumption. The LISA has been chosen to be implemented to capture the spatial structure of the distance to consumption and the synergy between neighboring blocks that can attract certain types of Twitter users and the conditions of smaller or greater distance traveled.

Formally, the LISA is expressed:

$$I_i = \frac{\sum_j w_{ij} z_i z_j}{\sum_i z_i^2} \tag{5}$$

Equation 5 is the classic equation proposed by(Anselin, 1995), in which it recovers the univariate correlation of a phenomenon in a spatial unit i, in this case blocks, correspondence of the same variable in neighboring spatial units j. In this way, Moran's I is built based on a matrix of geographic weights w_ij; in this case, this matrix has been built with the centroids of the blocks with a bandwidth of 9 km to guarantee that all the blocks have at least three neighbors. The z values of the distance variable refer to the standard normalization, where there is a mean of 0 and a standard deviation of 1 in the mean distance traveled to each block of consumption. Based on this indicator, attraction clusters can be identified based on the distance traveled to each of the consumption blocks and their adjoining blocks.

Results

First, the average distance traveled to the consumption blocks from the potential homes of the consumers is 10.7 km, with a median distance of 8.1 km. Given the expanded structure of the MZVM, the distances traveled have a strongly skewed distribution for traveling fewer kilometers than those reported by the median. For this reason, Fig. 2 shows the distribution of the distances traveled on a logarithmic scale. Distances traveled to the blocks behave in an approximately normal distribution of probability with a small bias to the left. The histogram, in turn, shows the cohorts by quartiles.

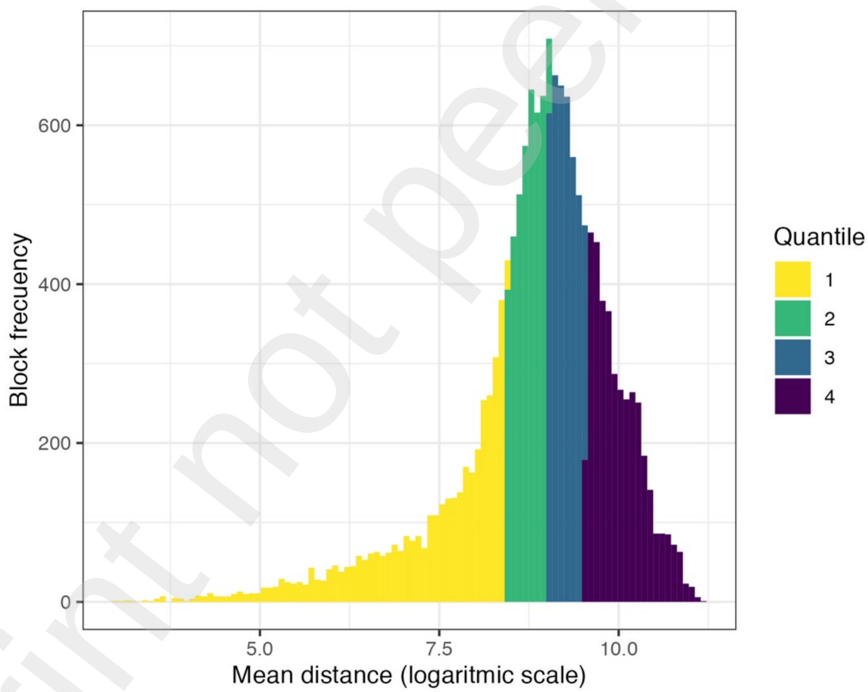

Fig. 2. Histogram of the average distances traveled to each block of consumption (logarithmic scale).

Source: author's calculations

Fig 3 shows the cartographic complement of the blocks that are located in the first quartiles of the distance distribution. In this figure, you can see two major pattern allocations of blocks. On the one hand, the blocks located in the most consolidated areas in urban terms are shown, such as the neighborhoods in central boroughs. In this pattern, the neighborhoods of the Benito Juárez borough stand out, such as the neighborhoods of Narvarte, Napoles; those of the Cuauhtémoc borough: Condesa, Hipódromo-Condesa, Roma Norte and Sur; and neighborhoods of the Miguel Hidalgo borough such as Polanco. On the other hand, blocks are identified in the municipal capitals of the municipalities of the Mexico State, as shown in the right-side map in

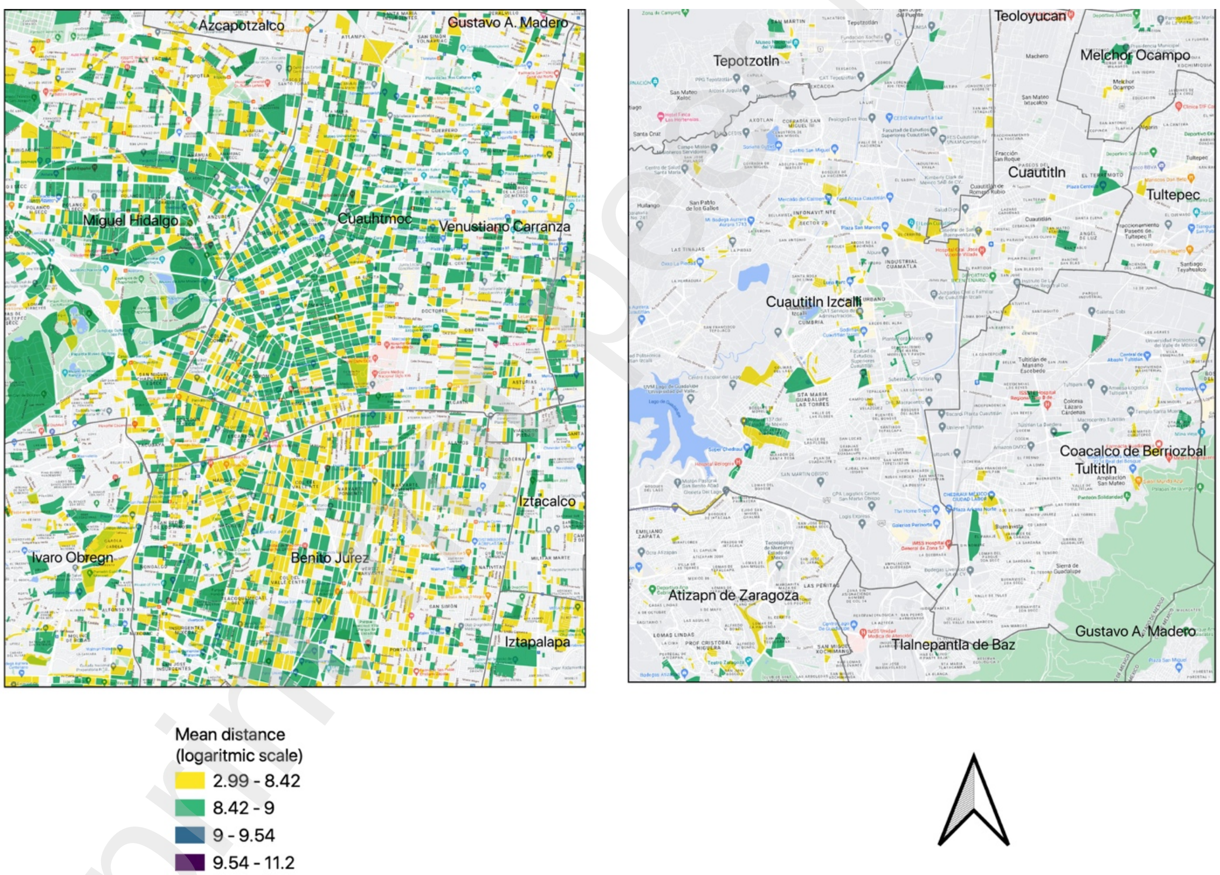

Fig. 3. Average distances traveled (logarithmic scale) quartiles 1 and 2.

Source: authors' calculations

The second group of blocks are those that report average distances traveled greater than the median of the universe according to the histogram in Fig. 2. In this regard, Fig 4 shows the blocks of the central zone of the MZVM on the left side of the image. It shows that the area surrounding the Mexico City base forms a conglomerate of blocks that attract consumers from distances greater than the metropolitan median. The image on the right side of the same figure shows the territorial pattern of the same type of blocks in the northern zone of the MZVM. This image is a sample of what happens in the peripheral areas of the metropolis, in which the empirical clusters of attraction of great distances are associated with shopping centers. In the case of the figure, the example is associated with the Punta Norte Shopping Center.

Fig 4. Average distances traveled (logarithmic scale) quartiles 3 and 4.

Source: author´s calculations

In addition to the simple distance to each block, it is relevant to know which economic specialization is associated with the range of distance to consume. Regarding the latter idea, Fig. 5 shows the distribution of the economic specialization of blocks in the MZVM. It is observed that the largest number of blocks are identified as specializing in general services, followed by professional services and, finally, financial services. The comparison of the number of blocks according to distance quartile shows that most services have the same number of blocks, except for three: personal and medical services, which have 3 percentage points and a percentage point more blocks in those that attract short distances, and retail trade, where there is a greater number of blocks that attract consumers from greater distances.

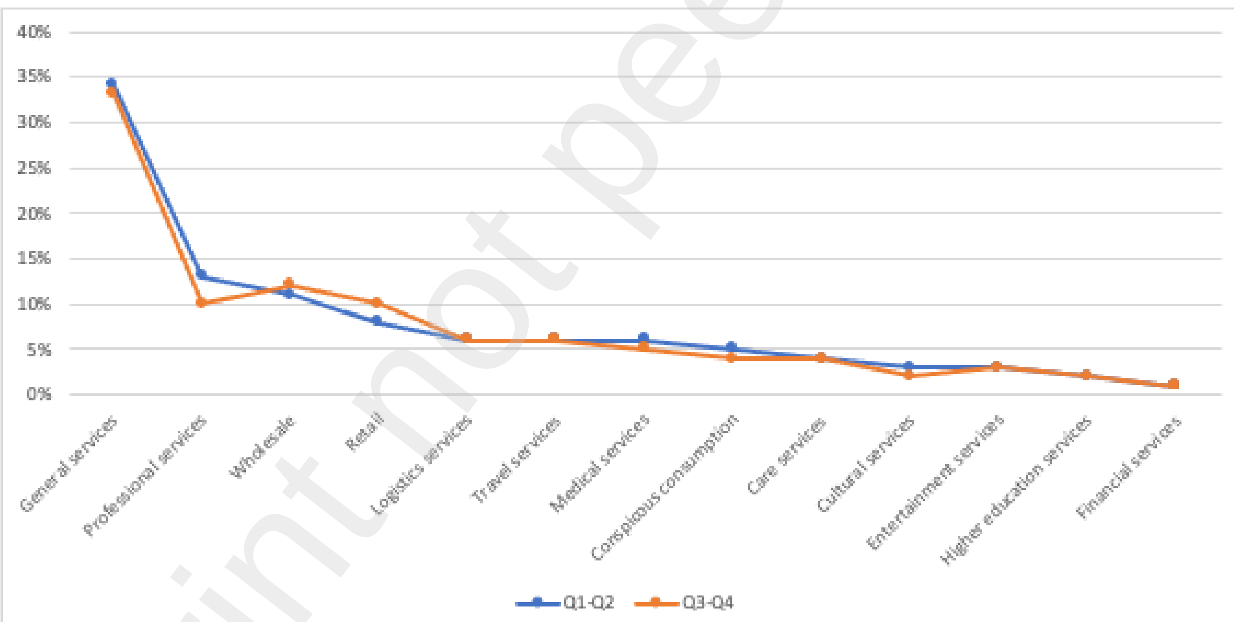

Fig. 5. Distribution of blocks by specialization and quartile of average distance to consumption. Source: authors’ calculations

Distributions of distances traveled to each type of economic specialization are shown in Fig 6. The order of the activities is based on the interquartile range, with the aim of showing the friction level of the distance from the consumer to consume according to their specialization. Activities that have the widest interquartile ranges are retail and wholesale trade, as well as care services. This amplitude in the interquartile ranges can be interpreted as a lower friction to the distance to consume these goods and services. The evidence of wider interquartile ranges in wholesale and retail commerce supports the results shown on previous maps, which show concentration patterns of attraction blocks on periphery zones of MZVM. Additionally, it is noteworthy that care services face a wide range in terms of the distances traveled to reach them. This indication could be linked to mobility patterns to work or sociodemographic structures of the city, as developed in the following section of the work. Finally, the services that face the greatest friction at a distance, or the lowest interquartile range, are financial and cultural services. In other words, consumers of this type of service are willing to travel shorter distances to consume them, as demonstrated in the next section of the work. Finally, the services that face the greatest friction at a distance, or the lowest interquartile range, are financial and cultural services. In other words, consumers of this type of service are willing to travel shorter distances to consume them.

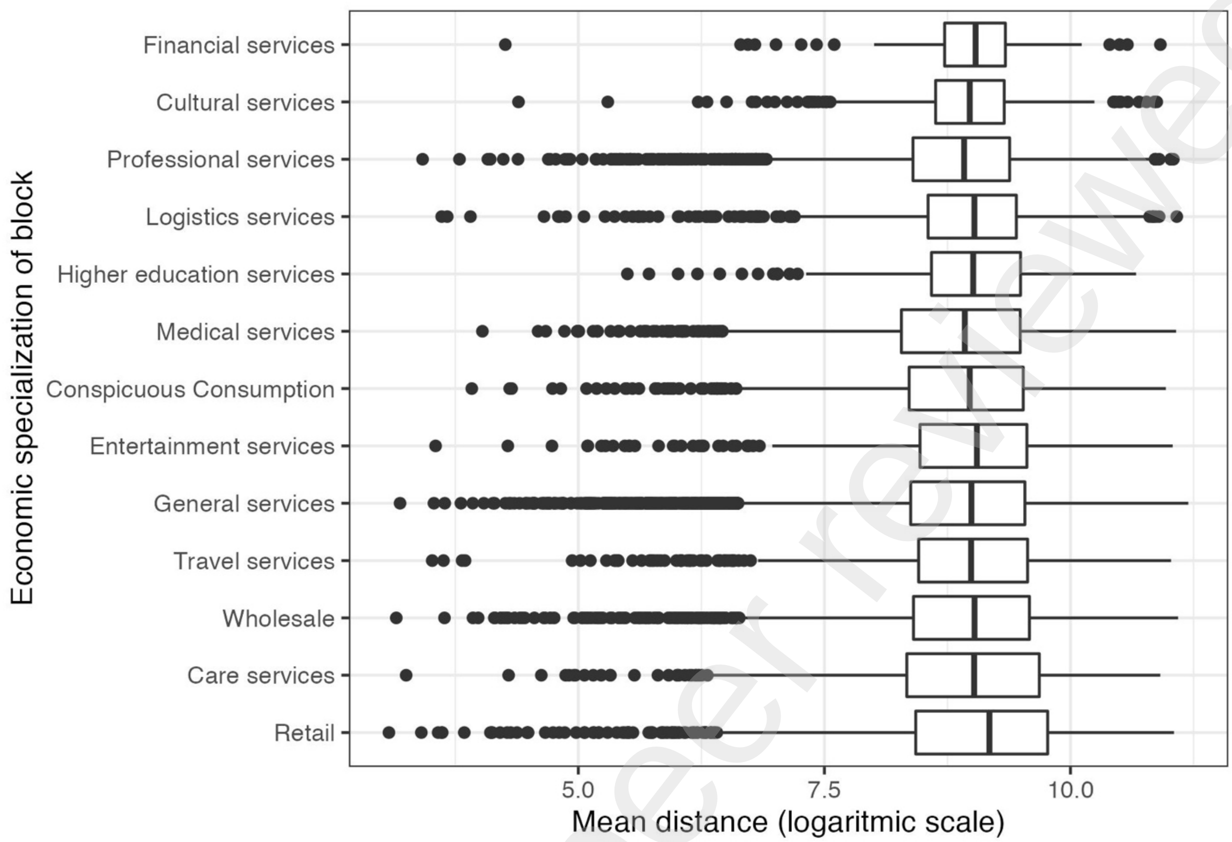

Fig. 6. Mean distances to block consumption by economic specialization

Source: own elaboration based on collected tweets and DENUE, 2021

Finally, clusters are formed that attract small and long distances to consumption. Fig 7 shows the LISA results over the logarithm of the mean distances. On the one hand, the formation of a consumption corridor with high friction at a distance in the central areas of Mexico City is shown. These areas are delimited by some of the main avenues in the center of the metropolitan area. On the other hand, long-distance attraction clusters are formed in some shopping centers on the outskirts of the metropolis, again highlighting shopping center areas.

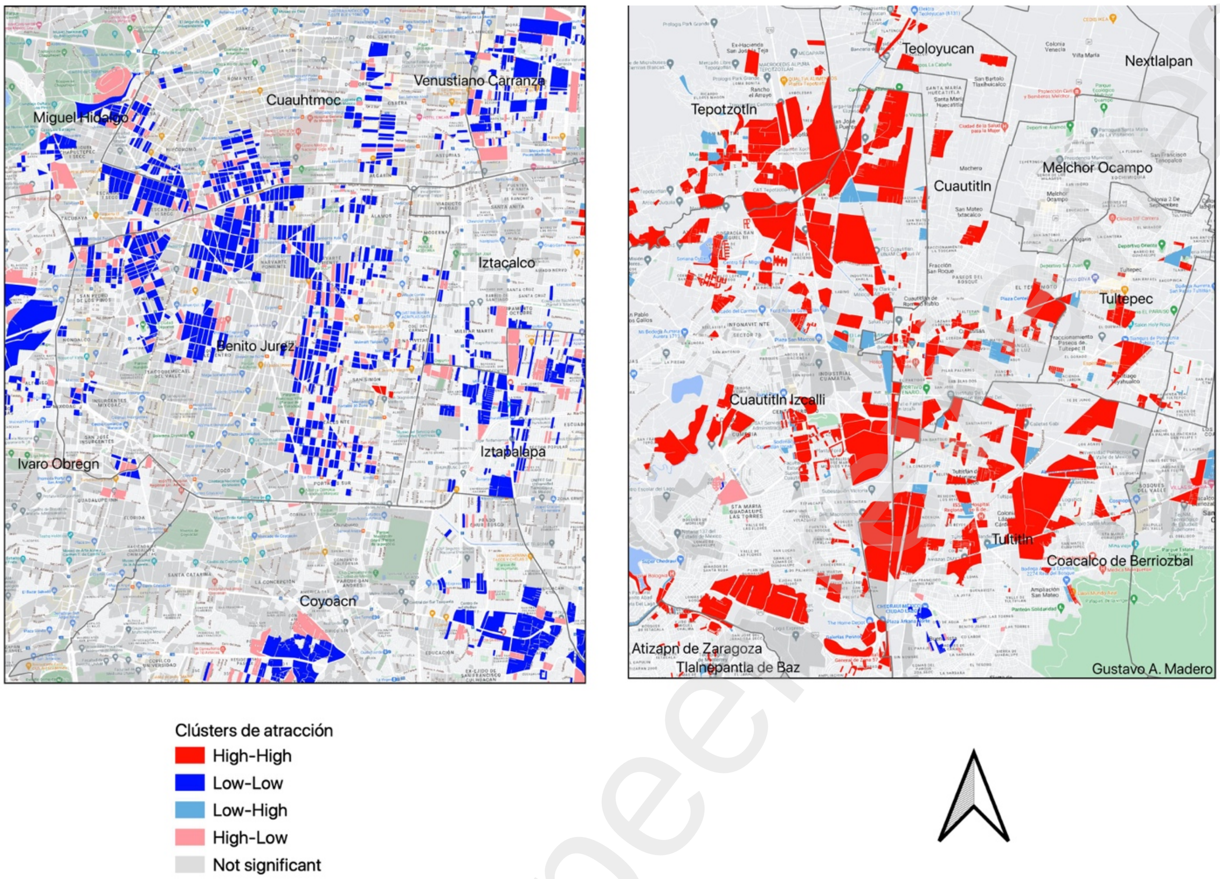

Fig 7. Conformation of clusters of attraction to consumption centers.

Source: authors' Calculations

Discussion

Central place theory points out the existence of sociospatial differentiation depending on the economic specialization and willingness to travel distances by consumers. This causes the existence of hexagons that make both the location and the distance to travel efficient. Furthermore, the location is hierarchical, that is, certain activities have less distance friction, and they are attended by consumers from greater distances. Others are located closer to the market. Although it is true that this theory was proposed at a time when activities were not so clearly differentiated, its value lies in the explanatory capacity of market spatiality. Currently, with the differentiation of consumption (thanks to supply), it is, without a doubt, necessary to reconsider explanatory models of phenomena of mobility, consumption, and sociodemographic characteristics of the population.

Specialization of certain areas of the city can generate a clearly differentiated pattern in terms of the offer and attraction of trips. Currently, it is not only the consumption of the product but also of the place that establishes patterns of attraction of trips and distance from them. The results of this work raise the question of how much the consumption of the place becomes a clearly specialized pattern that joins the consumption of the product or service.

The distribution of specialization is consistent with that reported in previous studies on specialization in the tertiary sector. Clusters that attract consumers who travel long distances have been identified, particularly in shopping malls. This pattern of consumption can be linked to Ushchev et al. (2015), in which it was proposed as a competition for the diversification of activities in shopping centers, which caused the consumer behavior of traveling long distances to take advantage of that competition. This pattern of consumption toward shopping centers has been documented in the cases of nonwork trips and for elements of certain brands located in shopping centers. Thus, cases of consumer trips to shopping centers and their socioeconomic impacts have been identified in countries such as Singapore, Ghana and the United States (Eduful, 2021; Epstein, 1961; Ponce-Lopez & Ferreira, 2021).

On the other hand, the identification of short-distance attraction clusters is linked to upper-middle class areas in which people seek consumption in the vicinity of their residence. At the same time, it opens a panorama of interpretation for the synergy between the housing structure and the availability of certain types of consumption, such as restaurants, cultural services or others that face greater friction at a distance. In this regard, it has been identified that there is a relationship between the structuring of the housing market, particularly from the perspective of hedonic prices, and the consumption of certain consumer goods and services such as restaurants, local commerce, or leisure activities (Carlino & Saiz, 2019; Kang, 2019; Kuang, 2017).

As mentioned, the configuration of the distances traveled toward each economic specialization is consistent with what is reported in the international literature. However, in the case of the MZVM, care services stand out, since they have the third widest interquartile range in all categories of services. This has been approached from the international literature as a colocation of trips to work that link the proximity of work clusters with the proximity of care services, especially in early stages of life (Aparicio Trejo, 2022; Drinkwater, 2015).

Conclusion

Consumption and mobility in cities are topics of increasing interest in urban studies. This work is inserted in an international discussion that has identified the phenomenon of consumption in cities as an defining characteristic of the daily life of the people who inhabit them, as a dimension of public policies and as a generator of urban spaces and practices in contemporary society.

Furthermore, the use of big data and alternative sources of information from social network sites, such as Twitter, have made it possible to identify on small geographic scales, in this case urban blocks, patterns of consumption and mobility for this reason in Twitter users. In this sense, the work proposes methodologies for assembling information with official data sources to offer greater solidity to the information coming from big data.

In general, the work yields three important conclusions for the study of consumption in the MZVM. The first involves identifying activities that face different levels of friction from distance to their consumers. Regarding this point, the activities with the least friction related to retail trade, personal services and care services are highlighted. Those with the greatest distance friction on the part of consumers are specialized financial services and entertainment and cultural services.

The second, the information and the methodology used, suggests that the formation of consumption clusters characterized by the friction of distance is associated with urban dynamics such as the configuration and consolidation of homes of medium and high socioeconomic status, in the case of clusters with high distance friction and, on the other hand, the configuration of clusters with little friction is associated with consumption in peripheral areas of the metropolis, which are associated with consumption in shopping malls.

The third consideration points to the identification of the consumption of services in the city and the friction of the distance they face due to the co-location of other dynamics. In this case, it has been identified that care services have a wide interquartile range, which may suggest that people consume this type of service in different locations in the city. This is probably due to the consumption synergies of these services linked to the location of work centers.

This work opens the door for researching other aspects related to the consumption decisions of the city's inhabitants in Mexico. The first is the relationship between economic specialization and the diversification that consumers can find in relatively compact spaces. Another aspect to investigate further is the type of consumer observed through Twitter, since the information on this platform provides biased information in sociodemographic terms. Given the benefits of big data in terms of its temporal granularity, it is possible to investigate changes in consumption patterns in cities in the face of the rise of e-commerce, which increased as a result of the COVID-19 health crisis.

Finally, the work raises the need for more studies that consider three measures (consumption, mobility and big data). There are still few works in Latin American countries, but we hope that in the coming years they will grow as more information becomes available and it becomes more accessible both for research and for decision-making.

References

Alonso, W. (1970). *The economics of consumption, daily life, and urban form* [Book]. Berkeley, Calif. : Institute of Urban [and] Regional Development.

Anas, A. (2007). A unified theory of consumption, travel and trip chaining. *Journal of Urban Economics*, *62*(2), 162–186. https://doi.org/10.1016/J.JUE.2006.05.002

Anenberg, E., & Kung, E. (2015). Information technology and product variety in the city: The case of food trucks. *Journal of Urban Economics*, *90*, 60–78. https://doi.org/https://doi.org/10.1016/j.jue.2015.09.006

Anselin, L. (1995). Local Indicators of Spatial Association—LISA. *Geographical Analysis*, *27*(2), 93–115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x

Aparicio Trejo, A. F. (2022). *Distribución espacial del cuidado formal y movilidad cotidiana de las mujeres en la Zona Metropolitana del Valle de México* (L. L. Sánchez Peña & A. Escoto, Eds.).

Brzustewicz, P., & Singh, A. (2021). Sustainable consumption in consumer behavior in the time of COVI-19: Topic Modeling on Twitter Data using LDA. *Energies*, *14*, 5787.

Carlino, G. A., & Saiz, A. (2019). Beautiful city: Leisure amenities and urban growth. *Journal of Regional Science*, *59*(3), 369–408. https://doi.org/10.1111/JORS.12438

Christaller, W. (1966). *Central places in southern Germany* [Book]. Englewood Cliffs, N. J. : Prentice-Hall, 1966.

Davis, D. R., Dingel, J. I., Monras, J., & Morales, E. (2019). How segregated is urban consumption? *Journal of Political Economy*, *127*(4), 1684–1738. https://doi.org/10.1086/701680/SUPPL_FILE/2016837DATA.ZIP de Simone, R.-L. (2018). Instalando la ciudad del consumo: el palimpsesto urbano del primer shopping mall chileno en el fundo San Luis, Santiago. *Revista EURE - Revista de Estudios Urbano Regionales [on Line]*, *44*(133), 91–112.

Dolega, L., & Lord, A. L. (2020). Exploring the geography of retail success and decline: A case study of the Liverpool City Region. *Cities*, *96*, 102456. https://doi.org/10.1016/J.CITIES.2019.102456

Dolega, L., Reynolds, J., Singleton, A., & Pavlis, M. (2019). Beyond retail: New ways of classifying UK shopping and consumption spaces: *Https://Doi.Org/10.1177/2399808319840666*, *48*(1), 132–150. https://doi.org/10.1177/2399808319840666

Drinkwater, S. (2015). Informal Caring and Labour Market Outcomes Within England and Wales. *Regional Studies*, *49*(2), 273–286. https://doi.org/10.1080/00343404.2013.766320

Duhau, E., & Giglia, A. (2016). *Metrópoli, espacio público y consumo*. Fondo de Cultura Económica.

Durand, C. (2015). Wlamart en México, una trayectoria exitosa y sus causas. En C. Alba & P. Labazée (Eds.), *Metropolización, transformaciones mercantiles y gobernanza en los países emergentes* (pp. 145–166). El Colegio de México.

Eduful, A. K. (2021). Economic impacts of shopping malls: The Accra (Ghana) case study. *Cities*, *119*. https://doi.org/10.1016/J.CITIES.2021.103371

Epstein, B. J. (1961). Evaluation of an Established Planned Shopping Center. *Economic Geography*, *37*(1), 12. https://doi.org/10.2307/142032

Hipp, J. R., Lee, S., Kim, J. H., & Forthun, B. (2022). Employment deconcentration and spatial dispersion in metropolitan areas: Consequences for commuting patterns. *Cities*, *131*, 103947. https://doi.org/https://doi.org/10.1016/j.cities.2022.103947

INEGI. (2018). *Encuesta Origen-Destino en Hogares de la Zona Metropolitana del Valle de México 2017 EOD. Documento metodológico*. www.inegi.org.mx

Jayne, M. (2006). *Cities and consumption*. London : New York, N.Y. : Routledge, 2006.

Kang, C. D. (2019). Effect of neighborhood income and consumption on retail viability: Evidence from Seoul, Korea. *Habitat International*, *94*, 102060. https://doi.org/10.1016/J.HABITATINT.2019.102060

Kuang, C. (2017). Does quality matter in local consumption amenities? An empirical investigation with Yelp. *Journal of Urban Economics*, *100*, 1–18. https://doi.org/10.1016/J.JUE.2017.02.006

Lloyd, A., Cheshire, J., & Squires, M. (2018). The Provenance of Custumer Loyalty Card Data . En P. Longley, A. Singleton, & J. Cheshire (Eds.), *Consumer Data Research* (pp. 29–39). UCL Press.

Martí, P., Serrano-Estrada, L., & Nolasco-Cirugeda, A. (2017). Using locative social media and urban cartographies to identify and locate successful urban plazas. *Cities*, *64*, 66–78. https://doi.org/https://doi.org/10.1016/j.cities.2017.02.007

Massey, J. (2005). The gentrification of consumption: A view from Manchester. *Sociological Research Online*, *10*(2). https://doi.org/10.5153/SRO.1099/FORMAT/EPUB

Milusheva, S., Marty, R., Bedoya, G., Williams, S., Resor, E., & Legovini, A. (2020). *Applying Machine Learning and Geolocation Techniques to Social Media Data (Twitter) to Develop a Resource for Urban Planning*. https://doi.org/10.1596/1813-9450-9488

Miyauchi, Y., Nakajima, K., & Redding, S. J. (2021). *Consumption Access and the Spatial Concentration of Economic Activity: Evidence from Smartphone Data* (Núm. 28497; Working Paper Series). http://www.nber.org/papers/w28497

Mulligan, G. F., Partridge, M. D., Carruthers, J. I., Mulligan, G. F., Partridge, M. D., & Carruthers, J. I. (2012). Central place theory and its reemergence in regional science. *Ann Reg Sci*, *48*, 405–431. https://doi.org/10.1007/s00168-011-0496-7

Murcio, R., Soundararaj, B., & Lugomer, K. (2018). Movements in Cities: Football and its spatio-temporal distribution. En P. Longley, J. Cheshire, & A. Singleton (Eds.), *Consumer Data Research* (pp. 85–95). UCL Press.

Olivera-Martínez, P. (2017a). Consumo y gentrificación en la Ciudad de México. En J. Gasca-Zamora (Ed.), *Espacio del consumo y el comercio en la ciudad contemporánea* (pp. 35–56). Instituto de Investigaciones Económicas-UNAM.

Olivera-Martínez, P. (2017b). *La tematización como estrategia de consumo en la ciudad* (J. Gasca-Zamora & P. Olivera-Martinez, Eds.). UNAM, Instituto de Investigaciones Económicas.

Orozco-Hernández, M. E., & García-Luna-Villagrán, G. A. (2014). Walmart en áreas periurbanas de la ciudad de Toluca, México: efectos sociales, económicos y territoriales. *Convergencia*, *64*(1), 93–116.

Ponce-Lopez, R., & Ferreira, J. (2021). Identifying and characterizing popular non-work destinations by clustering cellphone and point-of-interest data. *Cities*, *113*(September 2019), 103158. https://doi.org/10.1016/j.cities.2021.103158

Salganik, M. (2018). *Bit by bit. Social research in the digital age*. Princeton University Press.

Tabuchi, T., & Thisse, J. F. (2011). A new economic geography model of central places. *Journal of Urban Economics*, *69*(2), 240–252. https://doi.org/10.1016/J.JUE.2010.11.001

Tang, S., Hao, P., & Feng, J. (2020). Consumer behavior of rural migrant workers in urban China. *Cities*, *106*. https://doi.org/10.1016/J.CITIES.2020.102856

Ushchev, P., Sloev, I., & Thisse, J. F. (2015). Do we go shopping downtown or in the 'burbs? *Journal of Urban Economics*, *85*, 1–15. https://doi.org/10.1016/J.JUE.2014.10.001

van Weerdenburg, D., Scheider, S., Adams, B., Spierings, B., & van der Zee, E. (2019). Where to go and what to do: Extracting leisure activity potentials from Web data on urban space. *Computers, Environment and Urban Systems*, *73*, 143–156. https://doi.org/10.1016/J.COMPENVURBSYS.2018.09.005

Vargas Magaña, G. (2022). *"El jazz viene de la noche" : vida nocturna, ocio y consumo desde la escena de jazz y la música de improvisación en la Ciudad de México* (T. B. Hayden, Ed.).

Xiao, R., Yu, X., Zhang, Z., & Wang, X. (2021). Built-up land expansion simulation with combination of naive Bayes and cellular automaton model—A case study of the Shanghai-Hangzhou Bay agglomeration. *Growth and Change*, *52*(3), 1804–1825. https://doi.org/10.1111/GROW.12489

Xu, S., Li, S., & Wen, R. (2018). Sensing and detecting traffic events using geosocial media data: A review. *Computers, Environment and Urban Systems*, *72*, 146–160. https://doi.org/10.1016/J.COMPENVURBSYS.2018.06.006

Yao, Z., Yang, J., Liu, J., Keith, M., & Guan, C. H. (2021). Comparing tweet sentiments in megacities using machine learning techniques: In the midst of COVID-19. *Cities*, *116*. https://doi.org/10.1016/J.CITIES.2021.103273

Consumer attraction centers in the Metropolitan Zone of the Valley of Mexico.

Approach based on data from social networking sites (Twitter)

Alejandro Sánchez-Zárate[a]

[a] Centro de Estudios Internacionales, El Colegio de México Carretera Picacho Ajusco 20, Col. Ampliación Fuentes del Pedregal, CP.14110. Ciudad de México, México

alsanchez@colmex.mx

Aurora A. Ramírez-Álvarez[b]

[b] Centro de Estudios Económicos, El Colegio de México Carretera Picacho Ajusco 20, Col. Ampliación Fuentes del Pedregal, CP.14110. Ciudad de México, México

aurora.ramirez@colmex.mx

Enrique Pérez-Campuzano[c]

[c]Instituto de Geografía, Universidad Nacional Autónoma de México, Circuito Exterior, Col. Universidad Nacional Autónoma de México, CP. 04510, Mexico

eperez@geografia.unam.mx

Corresponding author

Alejandro Sánchez-Zárate

Centro de Estudios Internacionales, El Colegio de México Carretera Picacho Ajusco 20, Col. Ampliación Fuentes del Pedregal, CP.14110. Ciudad de México, México

alsanchez@colmex.mx

Acknowledgements

The first author wants to thank to El Colegio de México for postdoctoral fellowship.

Funding

This work was supported by UNAM-PAPIIT IN304222

Role of the funding source

There is not any role of funding source on this paper, design or results

Data availability

Available upon request

Declaration of competing interest

The authors declare that this research was conducted in the absence of any relationships that could be construed as a potential conflict of interest in this manuscript.

Author contribution

All study participant provided informed consent. All authors have agreed with this submission. Alejandro Sánchez-Zárate contributes to research design, writing, editing, Aurora A. Ramírez-Álvarez contributes to research design, draft writing, editing, interpretation and revising. Enrique Pérez-Campuzano contributes to literature review research design, writing and revising.