

Article

# The Geography of Social Media Data in Urban Areas: Representativeness and Complementarity

Álvaro Bernabeu-Bautista , Leticia Serrano-Estrada \* , V. Raul Perez-Sanchez  and Pablo Martí 

Building Sciences and Urbanism Department, University of Alicante, Carretera de San Vicente del Raspeig s/n, San Vicente del Raspeig, 03690 Alicante, Spain; alvaro.bautista@ua.es (Á.B.-B.); raul.perez@ua.es (V.R.P.-S.); pablo.marti@ua.es (P.M.)

\* Correspondence: leticia.serrano@ua.es

**Abstract:** This research sheds light on the relationship between the presence of location-based social network (LBSN) data and other economic and demographic variables in the city of Valencia (Spain). For that purpose, a comparison is made between location patterns of geolocated data from various social networks (i.e., Google Places, Foursquare, Twitter, Airbnb and Idealista) and statistical information such as land value, average gross income, and population distribution by age range. The main findings show that there is no direct relationship between land value or age of registered population and the amount of social network data generated in a given area. However, a noteworthy coincidence was observed between Google Places data-clustering patterns, which represent the offer of economic activities, and the spatial concentration of the other LBSNs analyzed, suggesting that data from these sources are mostly generated in areas with a high density of economic activities.

**Keywords:** big data; social networks; geolocated data; land value; urban demographics



**Citation:** Bernabeu-Bautista, Á.; Serrano-Estrada, L.; Perez-Sanchez, V.R.; Martí, P. The Geography of Social Media Data in Urban Areas: Representativeness and Complementarity. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 747. <https://doi.org/10.3390/ijgi10110747>

Academic Editor: Wolfgang Kainz

Received: 20 August 2021

Accepted: 30 October 2021

Published: 3 November 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The field of research that deals with the analysis of urban dynamics through location-based social networks (LBSNs) has led to the development of new methods and techniques that provide valuable insights from a wide range of qualitative and quantitative approaches. These methods and tools aim to unlock the great potential of the information provided by these sources about urban activities and human behavior in city spaces [1].

A great deal of scholarship in the field of LBSNs data applied to the study of urban phenomena focuses on large dense urban areas with important amounts of data [2,3]. These include metropolitan areas with a high population density [4]; areas where urban activities are more likely to happen, such as commercial areas; or areas where points of interest are concentrated, as opposed to predominantly residential areas [5,6] where the amount of available data may not be considered sufficiently representative to draw meaningful conclusions. Indeed, analyzing LBSN data requires strategies to keep the size of the datasets manageable, whilst there must be sufficient gathered data to obtain rigorous findings [7]. The latter is often achieved by thickening the samples by various methods, such as combining the features of different complementary datasets to leverage the strengths of each, or by triangulating between traditional (administrative and field studies) and new data sources, which would allow for a more complete understanding of urban phenomena [8]. This study supports these two assertions and builds upon existing literature that aims to bridge the gap in the debate about whether data from LBSN is representative of the demographic and socio-economic profile in urban areas from which these data are generated and that ultimately have an impact on the reliability of these data for urban studies [9].

Specifically, this study analyses the potential relationship between the location and concentration of data from various types of LBSNs and a selection of demographic and socioeconomic factors of the geographical context in which these data are generated

(i.e., real estate value as a proxy of land value, average gross income, and population distribution by age). The hypothesis is based on the question of whether the socioeconomic aspects of a given urban context have an influence on the location and amount of LBSN data that are being generated.

While previous research was concerned with the relationship between the activities reflected in a single LBSN source such as Foursquare [10–12], Airbnb [13] or Twitter [14–17] and the social network's penetration (taking into consideration users' demographics and economic factors such as age, race, nationality or income), the contribution of this paper is twofold. Firstly, the focus of the study is on the demographic profile of the population within the city areas where data are being generated, instead of focusing on the demographics of users that generate the data. Secondly, multiple layers of information from different LBSNs are overlaid, providing a broader perspective on the opportunities and limitations with respect to the representativeness and complementarity of these data for their use in urban studies.

The paper is structured as follows: a literature review of recent research involving demographic and economic studies in urban areas through social networks is carried out (Section 2); next, the study area is defined, and the used data sources are introduced (Section 3); then, the methodology is outlined (Section 4); and finally, the obtained results are reported (Section 5), discussed (Section 6) and conclusions are drawn and summarized (Section 7).

## 2. Literature Review

### 2.1. Digital Data for Understanding Socioeconomic Urban Phenomena

Urban areas can be understood as complex systems with a wide range of overlapping and interconnected layers, such as land use, dynamics of population, economics, demography, and transportation, which, when combined, reveal a rich and realistic representation of spatial dynamics and provide insight into a wide variety of urban phenomena [18].

The understanding of this complex reality has been tackled by many disciplines. In particular, in urban research the relationship between citizens' preferences, their socioeconomic conditions and their behavior in city spaces has been approached from very different angles. An important part of the research that has aimed to relate the population's socioeconomic differences with other urban issues covers, among other topics, urban security [19], mobility and urban dynamics [20,21], urban vibrancy [22] and the cell phone usage that determines the possibility that population has to generate digital data [23–25]. More specifically, previous research focused on geolocated data has shown that the information generated through mobile devices can, to a great extent, represent a reflection of the cities' physical reality and is a complementary tool for demographic and economic studies such as those of population distribution patterns and mobility [26–30], or the analysis of housing prices [31–33].

### 2.2. Influence of Sociodemographic Characteristics on Social Media Data Generation

When addressing the use and representativeness of social media within certain urban areas, scholars are particularly interested in exploring emerging correlations between the user-generated content and population characteristics in relation to specific urban phenomena. In particular, recent research has addressed the relationships between social media data generation and ethnicity, age, gender, income and education of population in urban areas [34–36]. The study authored by Ballatore and De Sabatta [34] in the Los Angeles metropolitan area evidenced that geolocated tweets tended to be more concentrated in areas with a higher population density that often present higher poverty rates, younger population, lower income, lower education levels and higher deprivation indexes. In contrast, Foursquare venues and check-ins were more predominant in lower density areas with a white, educated and older registered population. The same study highlighted the divergence between the users' area of residence and the areas where they share tweets, and the fact that Foursquare venues tend to be present in areas with more public services

per inhabitant. In a different context, the studies by Rizwan and Gwiazdzinski [35], and Muhammad et al. [36], focused on the usage of Weibo (the Chinese social network based on check-ins) in Shanghai and Guangzhou regions, respectively, where they analyzed the demographic profile and use of registered Weibo users. Their findings showed that more than 70% of the registered users ranged between 20 to 35 years of age, that women were more likely than men to use Weibo during weekdays, and that at the weekend a similar check-in trend was observed in both groups. Moreover, men tended to be more active on activities related to professional, sales and services, but women were more active on venues registered in the platform under the residence and shopping categories. Another interesting finding was that check-in densities in activities related to sales, services and professionals were mostly concentrated in central areas of the city, whereas activities related to food, drink and residence were relevant in both the city center and suburban areas [35].

Other lines of research addressed LBSN data generation and the social differences or inequalities in certain geographic contexts [4,37,38]. For example, in the case conducted in Madrid, Spain, the majority of hotspots registered on Foursquare were located in the northern, wealthier half of the metropolitan area, characterized by low density and housing sprawl that has an important sociodemographic contrast with the poorer southern areas, with lower income levels, college education attainment and low employment rates [4]. A different case was the analysis of geotagged social media such as Twitter in the city of Louisville, US, which evidenced that despite the inequalities among different areas of the city where racial segregation was highly present, the geographical extent of these areas was not necessarily evidenced by the socio-spatial behavior of the population. However, recognizable differences were found between the areas where users lived and those from where they generated social media content [37]. Another study also conducted with geolocated Twitter posts predicted population stress level in rural and urban communities via sentiment analysis [38]. The findings connected poor health-related mentions with low socioeconomic status in rural areas, while urban communities reporting higher stress were also more likely to discuss relationships on Twitter, highlighting the social and cultural differences between regions according to the language used on social media. Twitter geolocated data have also been broadly used in studies that analyzed the population's spatiotemporal patterns in relation to land uses [39] and movement through different territorial delimitations [40].

Another research approach in this field has focused on the impact of social media data density and socioeconomic as well as demographic characteristics on urban vibrancy, e.g., population density, employment density, highly educated population density, average annual income. That is the case of the study carried out in the city of Shenzhen, China, which showed that employment density (availability) was the only demographic factor significantly associated with social media check-in based vibrancy metrics since the increase in jobs in those areas with lower employment density were shown to attract more people to perform daily activities [22].

Other researchers have pointed out the bias of LBSNs data regarding the usage differences among diverse age groups, income or educational level [9,41]. For example, a study about overall usage of Weibo in China showed that the elderly (over 65 years old) population was underrepresented because, compared with other age groups, their presence on social media in some regions of the country was very low [9]. Instead, the usage of LBSNs was found to be higher among the younger (between 18 and 29 years old) population [42]. Age, income and education levels are variables that may give rise to sampling bias over some population groups [43]. However, not every LBSN is equally influenced by the same sociodemographic determinants. For instance, a study in Great Britain found that Facebook usage was influenced by age and gender, but not income or education, while Twitter use was influenced by age and income, but not gender or education. The usage of other image-based LBSNs such as Pinterest was linked to users' age and income but not education or gender, while no demographic characteristics significantly predicted Instagram use [41].

There are many factors that may influence the representativity of LBSN data. However, the great potential of these sources is that the user-generated information is up to date and much more sizeable than the data collected from more traditional sources. This is due to the exponential increase in the amount of smartphone users, which has made it possible to implement these data to urban studies [42].

As for the aspects that have an impact on the spatial dimension of data density, research often emphasizes the relevance of city centers as nodes of higher user-generated content [4,44], especially when these become transformed by the pressure of tourism, which tends to be rather intense in central areas of historical cities and at major attractions [39,45]. The city center of Madrid, for example, in comparison with other urban nuclei of the Madrid metropolitan area, showed a great correlation between the number of registered businesses and the total number of Foursquare check-ins. This was also highlighted by hotspots analysis [4]. Indeed, Foursquare check-ins in Madrid were concentrated in areas with consumption (e.g., shopping and restaurants) [39]. However, in the case of LBSNs such as Airbnb in Barcelona, the concentration of accommodation sites was highly related to hospitality, leisure and entertainment use, but had no direct relation with the location of offices or commercial activities [45]. In a different geographical context, such as Shanghai, where the analysis included different categories from the database of the LBSN Weibo, a denser overall check-in distribution was found in the city center area, where there was a high concentration of popular places as opposed to other institutions (e.g., educational), which were scattered throughout the city [44]. Furthermore, the same study evidenced that a higher LBSN participation level was registered in venues in which the main activity was related to entertainment and shopping, especially those located in the downtown area. Nevertheless, the data density extended to suburban areas due to the presence of registered educational institutions and residential locations.

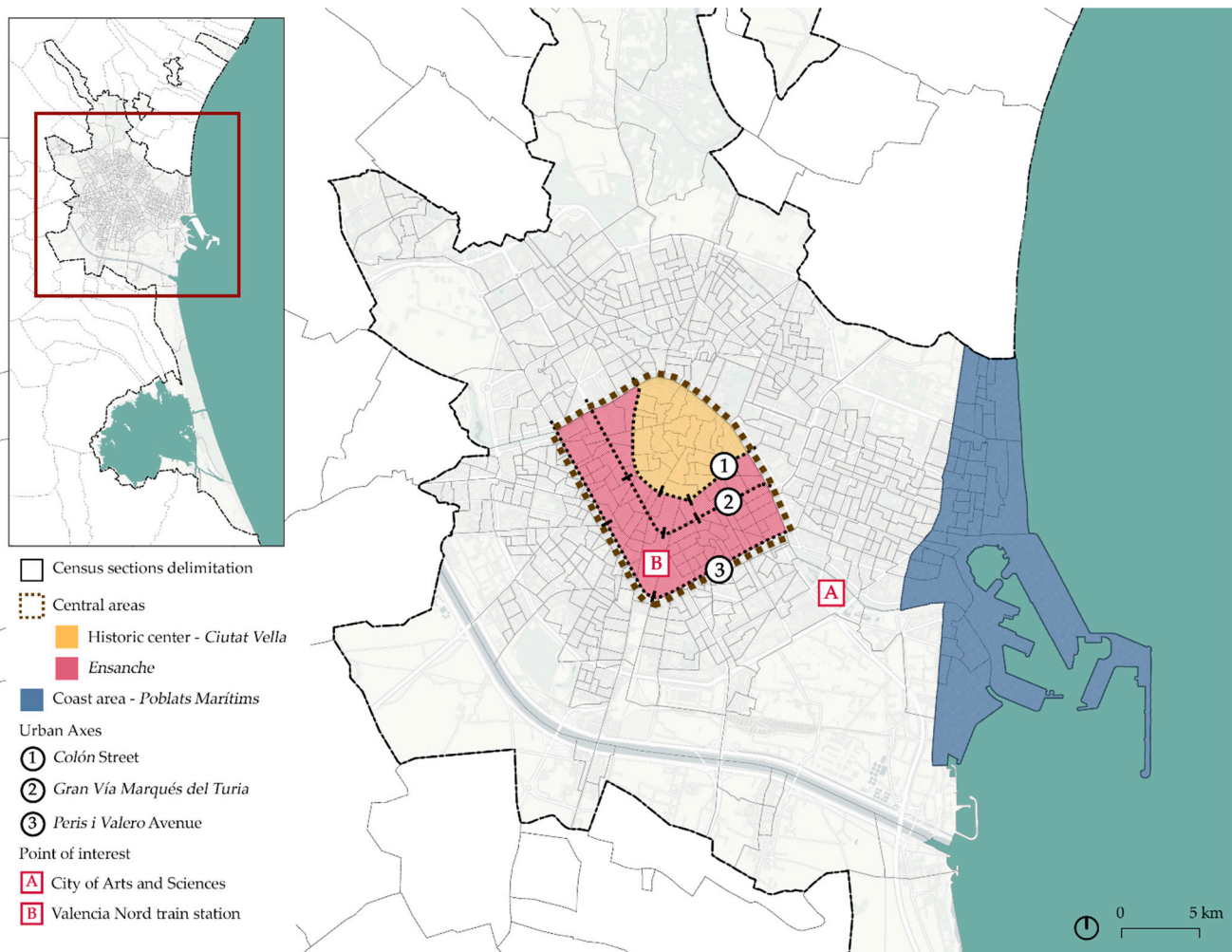
### 3. Case Study and Sources

#### 3.1. Valencia as a Case Study City

This study focuses on the Valencia municipal area, which has a census population of 794,288 inhabitants [46]. Valencia is the capital of the Valencian Community region, and is the third most populated city in Spain, the second most populated of the Spanish Mediterranean Arc (after Barcelona), and the main city of the central area in this geographical domain. Moreover, Valencia has one of Spain's biggest economies as it is the city with the third highest Gross Domestic Product (GDP) and employment rate [47]. It is considered as one of the cities that links smaller economic regions into the world's economy according to Globalization and World Cities Research Network [48]. In Spain's post-crisis period, namely from 2015 to 2019, Valencia was one of the cities that, despite losing population during the economic recession, regained it, becoming an urban area in which the periphery has grown more than the city core, with predominant suburbanization [49]. According to recent research, Valencia's central area (which includes the historic center and *Ensanche*—see Figure 1) and its surrounding districts, were the areas where prices per square meter of residential properties were higher, while the lowest prices were mainly in the north and south-west peripheral areas [50].

In the last few decades, Valencia was characterized by urban intensification, a higher presence of a young population, and coastalization as tourism intensified [51]. According to the Spanish National Statistics Institute (INE), the number of visitors and overnight stays grew exponentially in the last five years, with a significant drop in 2020 due to COVID-19 pandemic restrictions [52,53]. Tourism growth in Valencia was also reflected in the presence of Airbnb properties located, following the trend of other Spanish towns, particularly in the city center and close to tourist hotspots [54]. This was related to the presence of touristic sites and the concentration of leisure activities, but there were also other factors related to the spatial distribution of these properties, such as urban renewal strategies in certain areas or neighborhoods, or the legal framework [55]. However, the

lack of a regulatory framework means that those areas where most short-rental Airbnb properties were concentrated, were also experiencing gentrification [56].



**Figure 1.** Valencia city. Census section delimitation, main urban areas, urban axes and points of interest.

Regarding the usage of social networks, in 2020 the number of social network users in Spain between 16 and 65 years old accounted for more than 85% of the population, with little difference between gender (51% were female users and 49% male users) and age groups (21% in the range 16–24 years, 28% in 25–40 years, 29% in 41–54 years and 22% in 55–65 years) [57]. Twitter is one of the most used social networks after Facebook and Instagram. According to a recent survey, Valencia is the third city of Spain in the number of Twitter user profiles, and the second in percentage of users per population [58]. Nonetheless, in the specific case of Twitter, it is worth mentioning that there was a difference between the population age and the actual penetration of the social network. Even though Spain has a high percentage of population over 55 years of age, their usage of Twitter was lower when compared with younger groups aged between 16 and 34 [59]. The number of male users was slightly higher than the number of female users in all age groups, especially in the younger population. Furthermore, a Foursquare study showed that Valencia was the second region of Spain in terms of number of check-ins, especially at food and drink venues [60]. The penetration of other networks in Valencia is also remarkable. For instance, it was the third city of Spain in the number of registered Airbnb properties [61] and one of the cities in which the penetration of Airbnb had experienced a notorious exponential growth over the last few years [62].

In order to develop the spatial analysis, the municipality was divided in 596 spatial units based on the census sections of Valencia, which is the smallest unit of disaggregation for population data and statistical information according to the INE. Figure 1 shows the census section delimitations for the Valencia municipality. Three relevant urban areas are differentiated: the historic center—also known as the *Ciutat Vella*; the *Ensanche*—formed by those neighborhoods that resulted from the city’s growth in the 19th century; and the coastal area—also known as *Poblats Marítims*, traditionally a seafaring district that nowadays has become an important touristic area for the city. Significant urban axes and points of interest near the urban center are also included.

### 3.2. Data Sources

Various data sources have been used for this study (Figure 2). Specifically, economic data includes average gross income, while demographic data includes population age. Moreover, the data from five social networks, namely, Google Places, Foursquare, Twitter, Airbnb and Idealista, are also included.

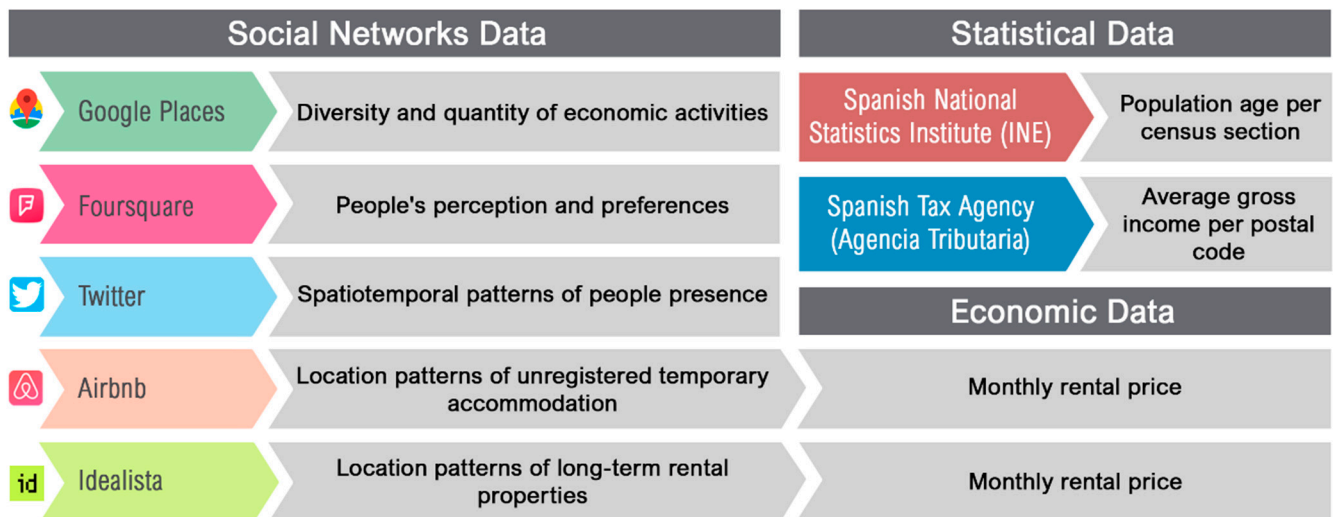


Figure 2. Data sources used for the study.

#### 3.2.1. Economic Data, Average Gross Income

The data on average gross income used to calculate the income and purchasing power distribution of the population were obtained from the 2016 Statistics of Personal Income Tax declarers by the Spanish Tributary Agency (Agencia Tributaria). The spatial units for this data are the postal code sections. For the purpose of this study, they have been reorganized into census sections.

#### 3.2.2. Population Age

The data on population age used to calculate population density by age range were obtained from the 2016 census by the INE. The census sections include the registered population by age, grouped into 5-year ranges. For this study, this information has been grouped into four intervals, the first of which included people from 0 to 19 years of age, the second from 20 to 39, the third from 40 to 64, and the fourth aged 65 years and older.

#### 3.2.3. Social Networks Data

Foursquare is a check-in based social network in which users register their presence with a check-in when they are at any given registered *venue*—a term used by the platform to refer to a place, point of interest or establishment. Since the *venues* included in the retrieved dataset have been checked into at least once, urban and geography researchers often recognize the source’s value for identifying relevant venues in the city [63] and

measuring their popularity [64]. For this study, Foursquare data were used as a proxy for the collective preference of venues and urban activities.

Google Places is a web service linked to Google Maps that aggregates and organizes all available information about *places*, as the platform refers to them, which includes geographic locations, points of interest and establishments. The datasets retrieved from Google Places contain a detailed listing of urban and economic activities registered in a given area, and are often used by scholars for analyzing the quantity, diversity and spatial clustering of business and services [1,65]. For the purpose of this research, Google Places data are used as proxy for the economic activities on offer.

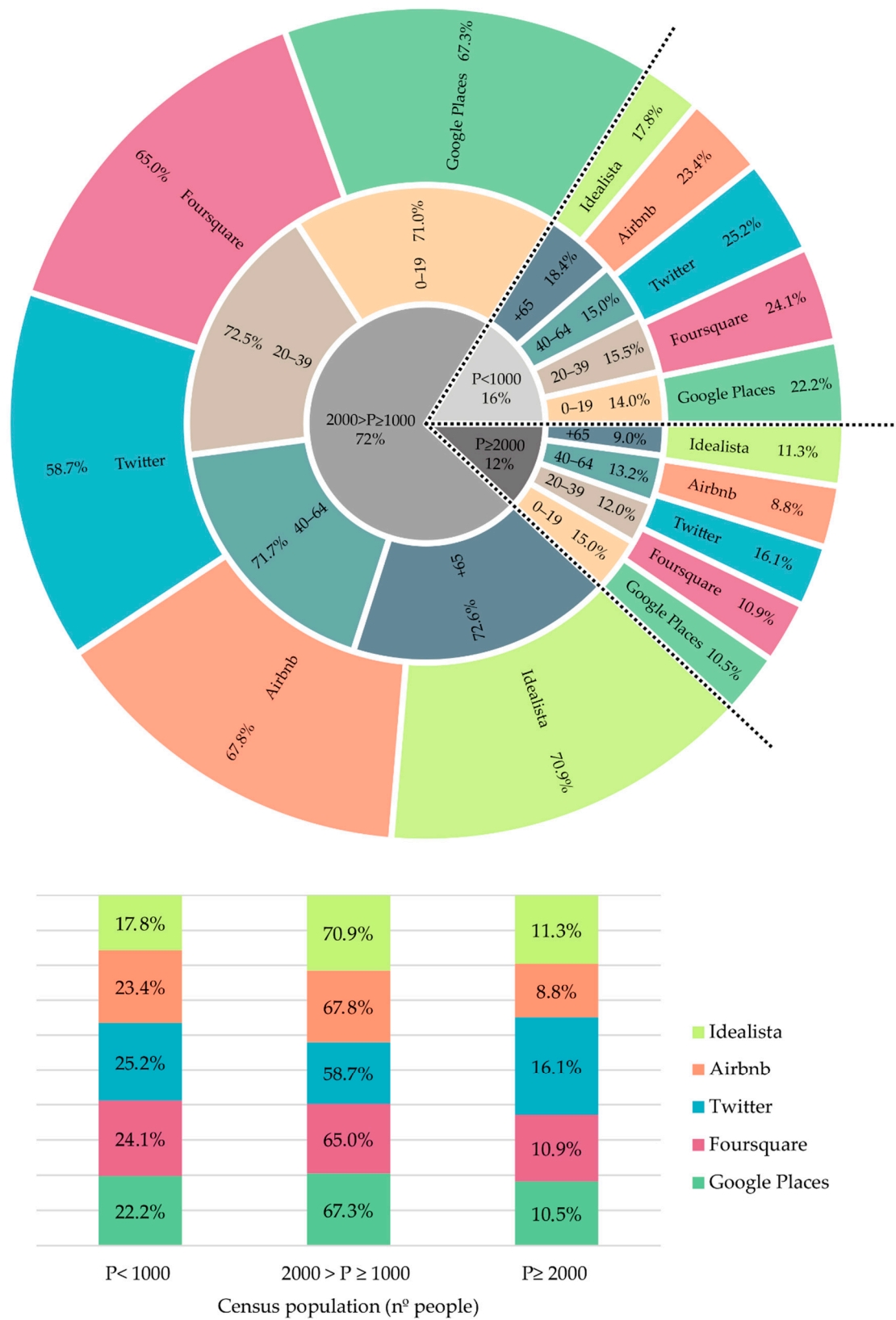
Twitter is a social network service and microblogging site on which registered users can share messages of up to 280 characters. Users decide whether to share the exact location of the tweets and, therefore, not all tweets have geo-locative information. For this study, only geolocated tweets were collected and analyzed. The datasets retrieved from this source offered valuable research opportunities to analyze, among other topics, the spatiotemporal patterns of people's location in the city [39]. Twitter data were used in this study as a proxy for the presence of people in the city.

Airbnb is a worldwide social network where short-term property rental services are listed. Users can advertise their home—or part of it—and make it available to other users of the platform. From the collected data it was possible to identify the distribution and/or concentration of this type of economic activity in a given area [66,67]. The data from Airbnb provided valuable information about the location and concentration of a very specific economic activity linked to tourism.

Idealista is a platform that advertises homes for sale and rent in Spain. This real estate social network has been widely used as a source of information in recent studies of spatial distribution and housing prices [68,69]. The data from Idealista included the geolocation of the properties, as well as specific characteristics such as the floorspace, the number of bathrooms, and the availability of commodities such as air conditioning, swimming pool, etc. The data from this source allowed for real-time monitoring of the evolution of prices in the city. In addition, it was possible to identify the different market segments within the city and to analyze the relationships established between the prices on offer and the characteristics of the properties.

Indeed, Airbnb and Idealista provided different but complementary information about economic activities within the real estate sector in relation to both residential and tourist activity. For this study, the information contained in these platforms together was used as a proxy indicator for real estate profitability and, indirectly, as an approximation of land value. All social networks variables used for this study are compiled in Table 1.

For the analysis, Valencia's census sections were adopted as the geographical areas of study or spatial units. Three types of census sections were considered: those with less than 1000 inhabitants; those whose population ranged between 1000 and 2000 inhabitants; and those with more than 2000 inhabitants. Figure 3 presents, in aggregated format and according to the census population range— $p < 1000$ ,  $2000 > p \geq 1000$  and  $p \geq 2000$ —the overall percentage of datapoints collected for each age group, that is to say, the age distribution and the amount of LBSN data as per this classification. As shown in Figure 3 (upper), of the total population, more than 70% for all age group ranges and more than 50% of the social networks data were located in those census sections with a population of between 1000 and 2000 inhabitants. Figure 3 (lower) shows the proportion of each social network data per census sections type, according to the number of inhabitants.



**Figure 3.** Upper: Percentage of LBSN data and age distribution per census section population range. Lower: Proportion of LBSN data in census sections according to the population ranges.



**Table 1.** Social network variables.

	<b>Foursquare</b>	<b>Google Places</b>	<b>Twitter</b>	<b>Airbnb</b>	<b>Idealista</b>
Variables used for this study	Geographic coordinates	Geographic coordinates	Geographic coordinates	Geographic coordinates	Geographic coordinates
	Venue address	Place address	Location of tweets (if shared by users)	Listing address	Listing address
	Venue ID	Place ID	Tweet ID	Listing ID	Listing ID
	Name of <i>venue</i>	Name of <i>place</i>	Textual content	Listing price per month	Listing price per month
	Number of registered users per <i>venue</i>	-	-	-	-

Census sections with a population of less than 1000 inhabitants included 22.2% of the data from Google Places, 24.1% from Foursquare, 25.2% from Twitter, 23.4% from Airbnb, and 17.8% from Idealista. In these census sections, the distribution of data for all five social networks was rather homogeneous, with almost a quarter of the observations being grouped in them. Census sections with a population that ranged between 1000 and 2000 inhabitants had a different distribution and grouped more than half of the observed data. In this case, except for Twitter data (where the percentage did not reach 60%), the presence of data from the remaining social networks reached values that exceeded 65%—and almost 71% in the case of Idealista.

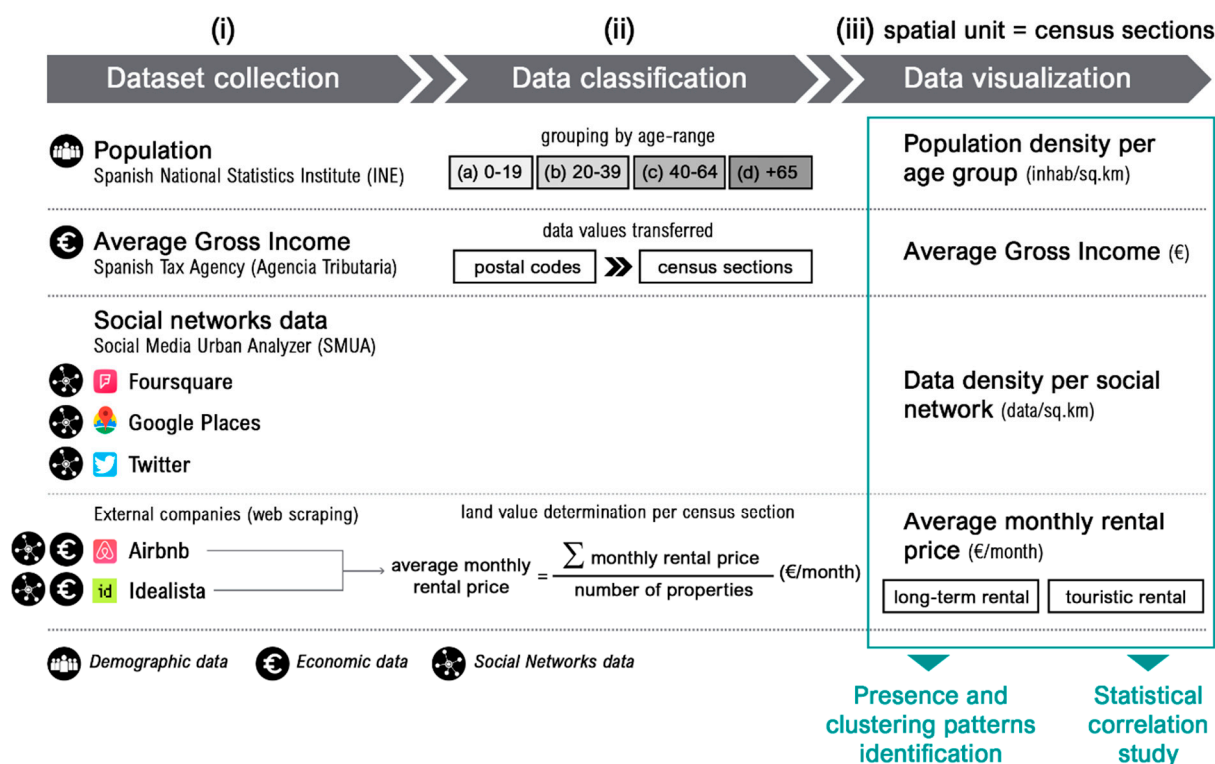
Census sections with more than 2000 inhabitants were the ones that collected the lowest number of LBSN observations. Nevertheless, Twitter data stood out, and was practically double the values of Airbnb and 1.5 times the other social networks' values.

#### 4. Method

The method mainly comprised the following steps (Figure 4).

- (i) Data were collected from diverse data sources, which included the statistical databases of the Spanish National Statistics Institute and the Spanish Tax Agency, as well as the selected social networks: Google Places, Foursquare, Twitter, Airbnb and Idealista.
- (ii) Data classification was carried out. Age groups were defined, data values were associated with their respective census section, and the land value was determined based on Airbnb and Idealista rental prices.
- (iii) Finally, two analytical and statistical techniques were used in order to find relationships between the location patterns of social network data and socioeconomic parameters. That is to say, partial methods were implemented to achieve the research objectives. First, all databases were visualized in a geographic information system and their location patterns were identified and compared; and second, a correlation study between all sources was performed.

In essence, this methodological approach was based on previous research that used a combination of social network data as layers of information, along with statistical information. For instance, data from the Spanish National Statistics Institute or the Spanish Cadastre was combined with Airbnb [45] or Twitter [70,71] in order to address short-term rental spatial patterns. Furthermore, visual representations of social media data facilitated the identification of spatial relations among different sources [72]. Similarly, overlaying the data density from various LBSN using the same spatial unit, i.e., the census sections [39], offered complex insights on urban reality [73]. Finally, correlation analysis [74] provided statistical evidence of the relationships observed in the visualizations.



**Figure 4.** Overall research methodology.

#### 4.1. Collection, Verification, and Visualization of Data Density

The Foursquare, Twitter and Google Places datasets were retrieved through their API (application programming interface) using a web-based application designed for that purpose: the SMUA—Social Media Urban Analyzer [1]. Airbnb and Idealista datasets were obtained through external companies that retrieved the information using web-scraping methods. The retrieval dates were the following: Google Places, 16 February 2018; Twitter streaming data, from 20 May 2016 to 20 May 2018; Airbnb, from 1 August 2015 to 1 January 2018; and Idealista from 30 July 2017 to 22 March 2018.

The collected LBSN datasets were cleaned following three criteria [1]: firstly, duplicate registers in Google Places, Foursquare, Airbnb and Idealista were eliminated; secondly, Google Places registers that did not represent an economic activity or a place (i.e., those that referred to street and neighborhood names, regions, postal codes, etc.) were eliminated; and lastly, tweets that were not generated by humans (tweets from ‘bots’ that were generated by automated accounts such as weather stations [2]), were discarded. It is worth noting that tweets generated from other platforms such as Instagram were considered regular tweets since they were user-generated and therefore provide evidence of human activity.

Once the data were cleaned, a visualization of data density was conducted using GIS software (QGIS). The spatial intersection between the LBSN geolocated data and the census sections allowed for the generation of themed cartographies.

#### 4.2. Data Classification

##### 4.2.1. Population Density by Age-Range

The 5-year range groups of population defined by INE and included in the census sections delimitation, were placed into broader age groups for this study. An initial grouping was suggested differentiating between active and inactive population [75]. However, in the end, a higher level of disaggregation was considered, especially in the age range corresponding to active population, due to the great variability of social media usage that can occur in this age group. Therefore, the population was finally divided into four groups

or age ranges: (i) from 0 to 19, (ii) from 20 to 39, (iii) from 40 to 64 and (iv) 65 and over. Population density was then calculated for each age group.

#### 4.2.2. Land Value and Average Income Distribution

As previously mentioned, for the purpose of this study, the information included in Airbnb and Idealista offered indications of real estate profitability and, indirectly, provided an approximation of the land value in each area. Therefore, the land value was obtained using both touristic short-term and long-term rental prices. The spatial intersection between the locations of Airbnb and Idealista properties and census sections was conducted, and the average price per spatial unit was calculated. Color gradation according to the average price was set for each land value: permanent or long-term rental (Idealista) and touristic rental (Airbnb).

The average gross income in the urban area distributed by postal code sections was transferred into census sections with the spatial intersection of both layers in order to establish a comparison between all parameters under the same spatial delimitation unit.

#### 4.3. Overlaying and Correlating LBSNs Data Layers with Economic and Demographic Determinants

After all layers of information were individually visualized considering the same spatial units, and the data had been classified, two types of analysis were carried out. The first analysis consisted of the identification of clustering patterns and correlation among LBSN data density and each of the economic and demographic parameters calculated. This was carried out by overlaying and comparing all generated cartographies. The second analysis, with a quantitative character, consisted of a correlation study (Pearson correlation) between population, amount of data obtained from each social network, average rent price for Airbnb and Idealista properties, and average income of all census sections. Finally, the results from both analyses were compared and discussed.

## 5. Results

### 5.1. Population Density by Age-Range and Social Networks Data

The visualization of population density shows that the highest of the four established age groups was generally organized radially around the historic center (*Ciutat Vella* area)—see Figure 1—and in areas near the coast, but without densifying the frontline. However, a density increase was noticed in the main axes and certain areas of the historic center near the *Ensanche*.

From the comparison between population density and LBSN data density it was found that areas with lower population density in all age groups—except for some areas of *Ensanche* (see Figure 1) where the density is slightly higher—match those with the highest data density from Foursquare, Google Places, Twitter and Airbnb (Figures 5–9). Moreover, according to the Foursquare dataset, the areas with the highest data density and the location of the top ranked ten venues with the highest number of registered users were located in the historic center, the coastline and the area surrounding the City of Arts and Sciences (Figure 10). Furthermore, the fact that these less-populated areas concentrated most short-term accommodation properties in the city, according to the Airbnb dataset, suggested that both the city center and the coast were mostly where properties were not intended as a main residence, but for touristic accommodation (Figure 11).

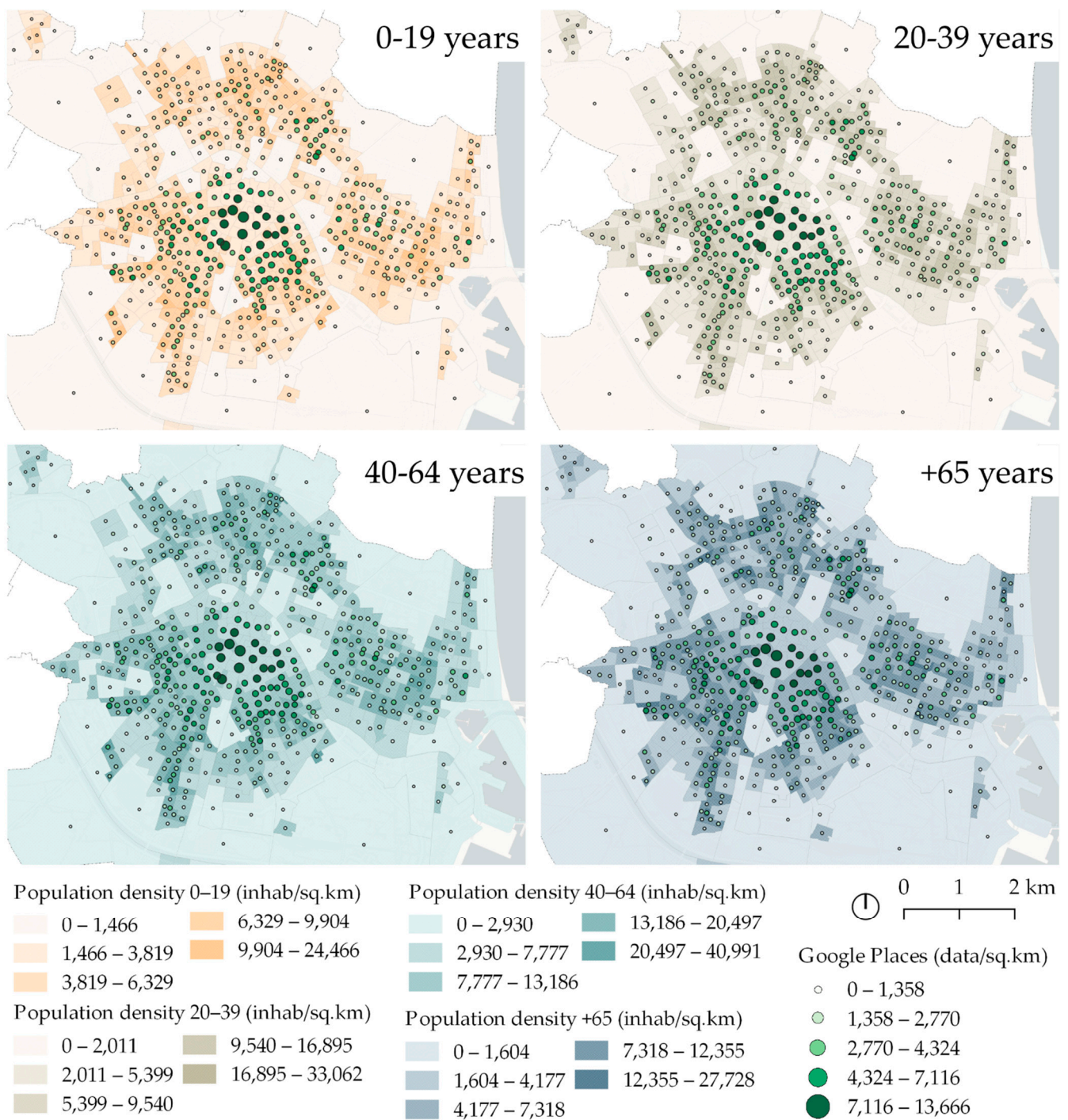


Figure 5. Population density per age group and Google Places data density.

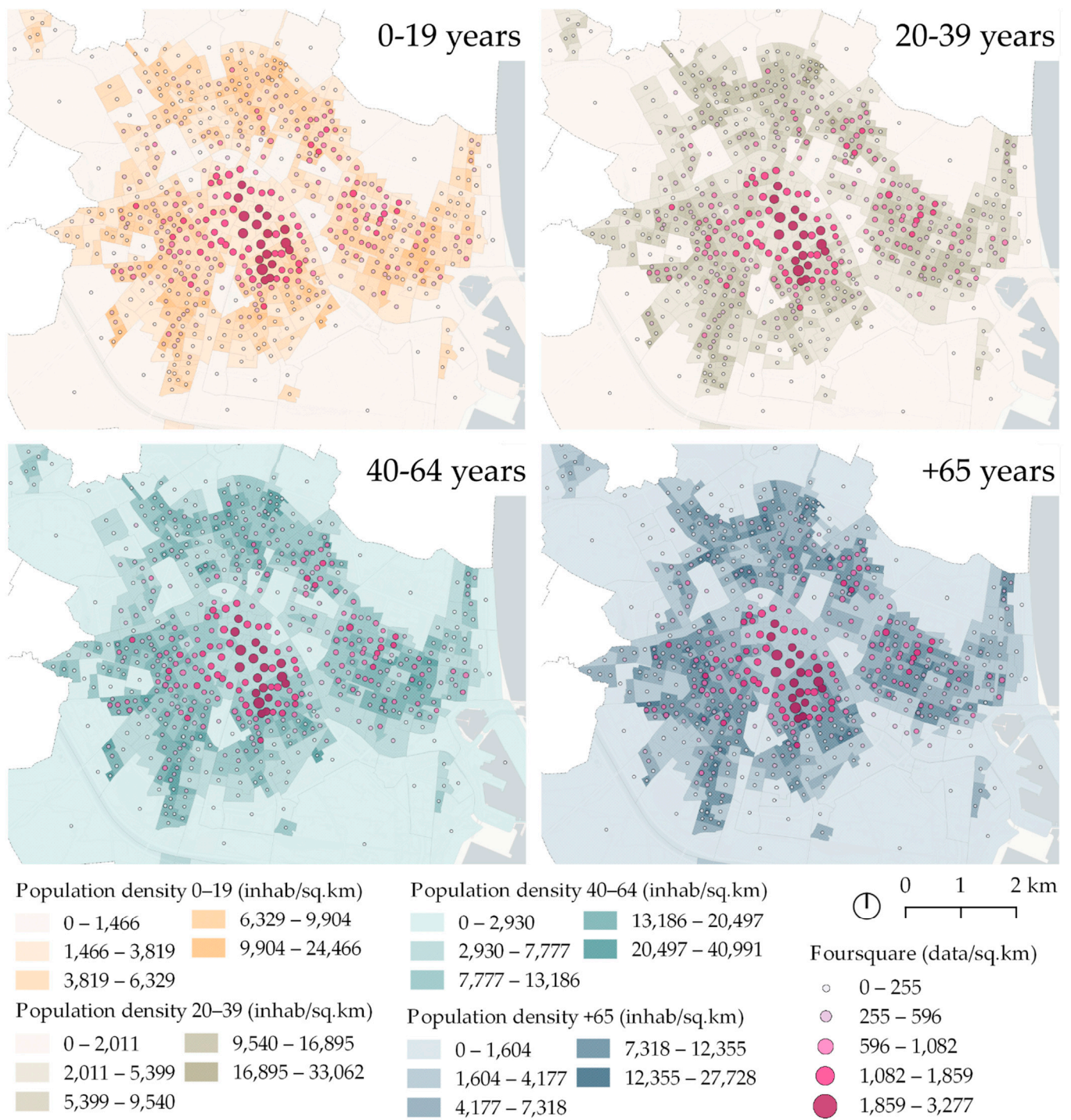


Figure 6. Population density per age group and Foursquare data density.

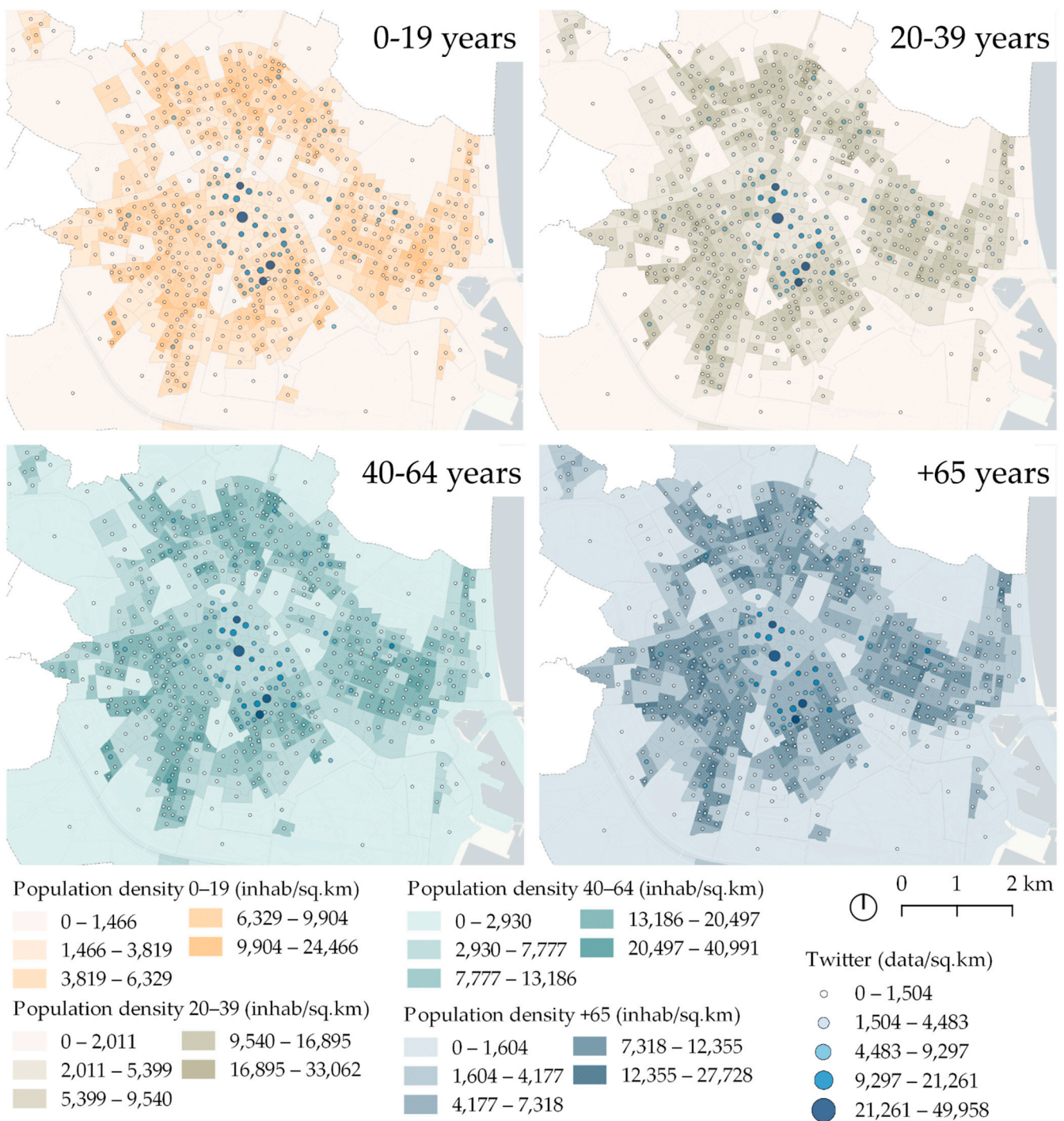


Figure 7. Population density per age group and Twitter data density.

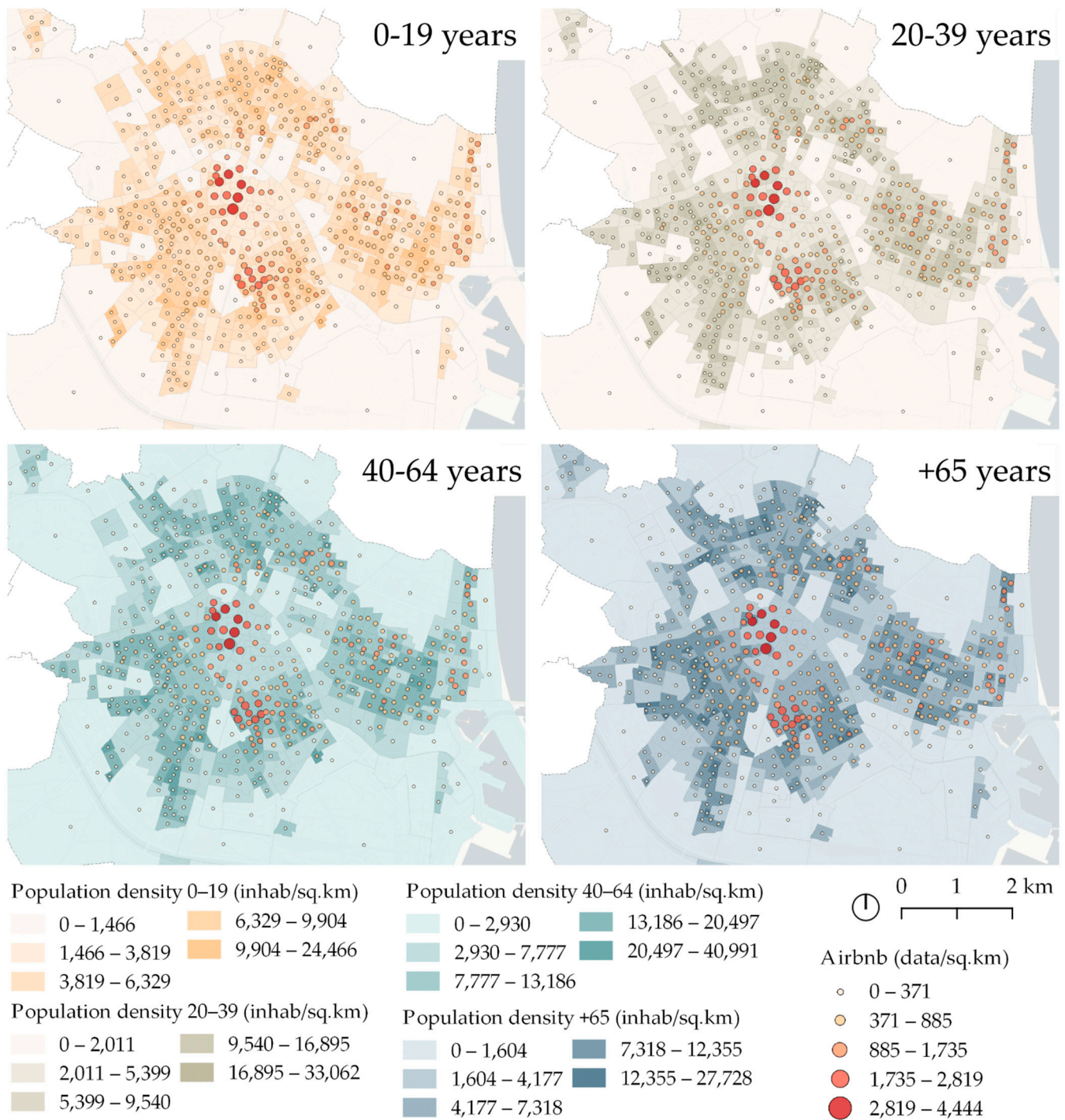


Figure 8. Population density per age group and Airbnb data density.

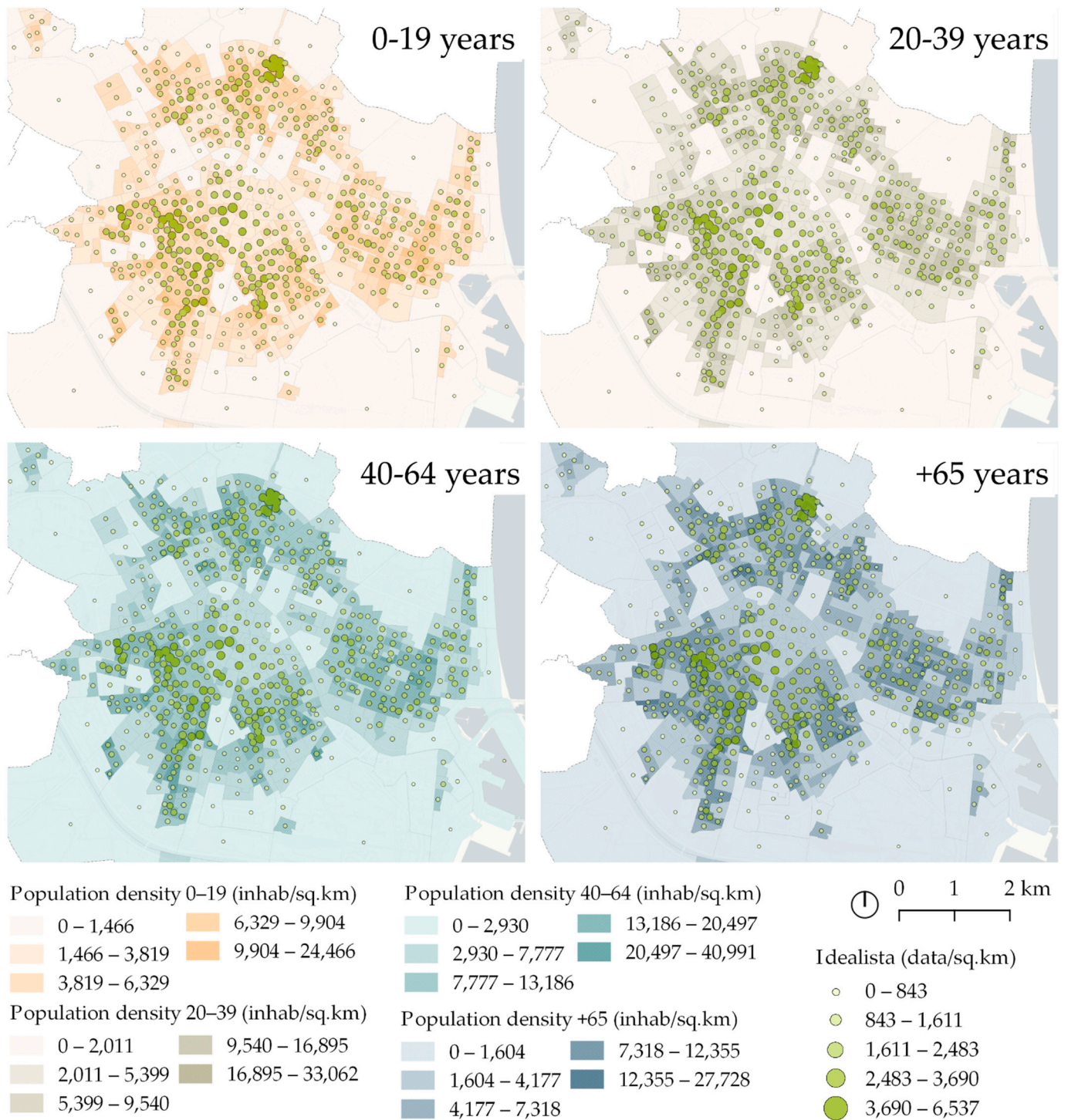
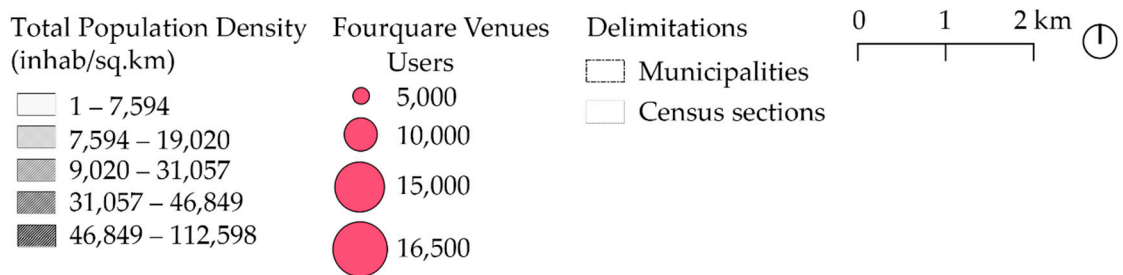
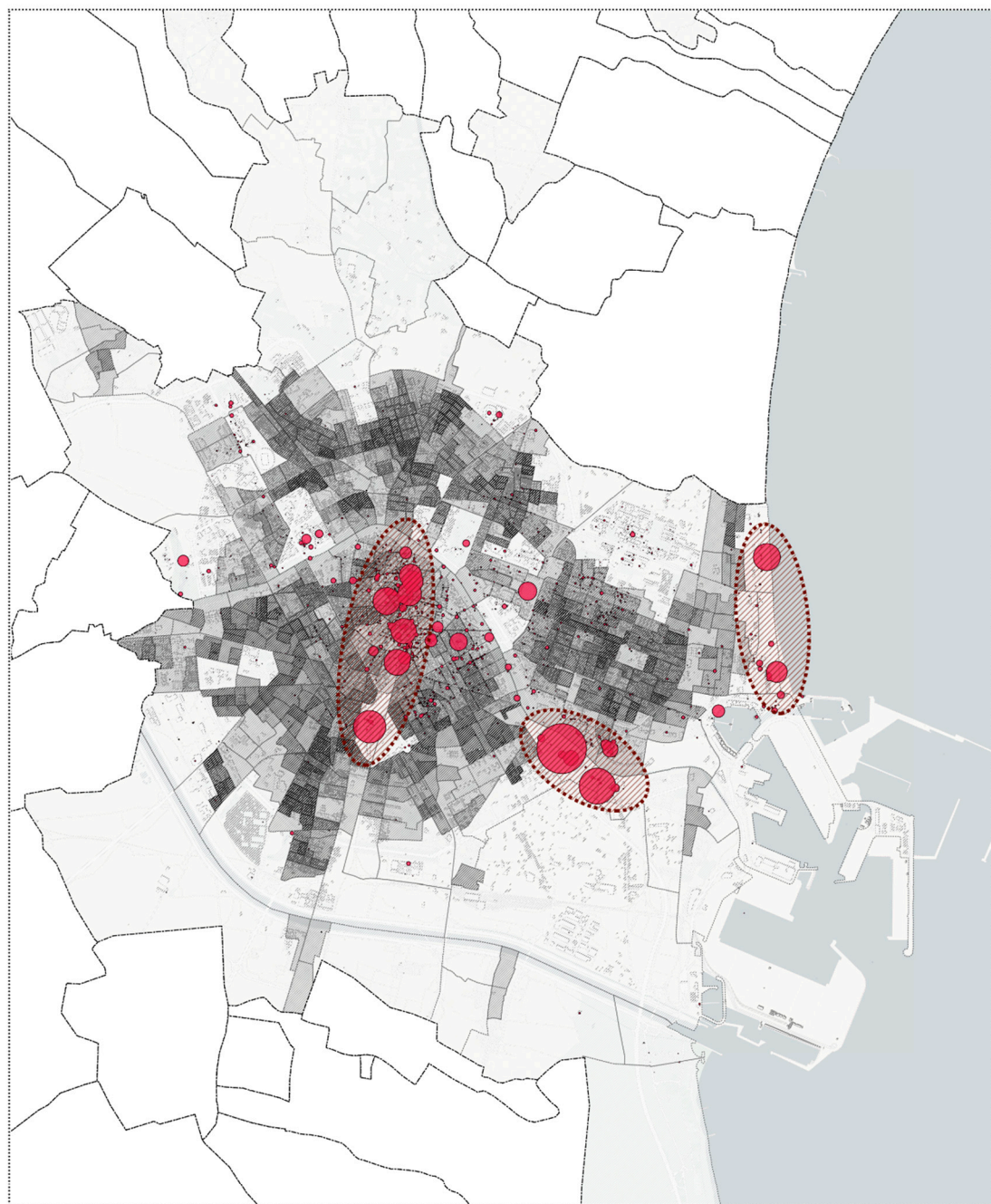
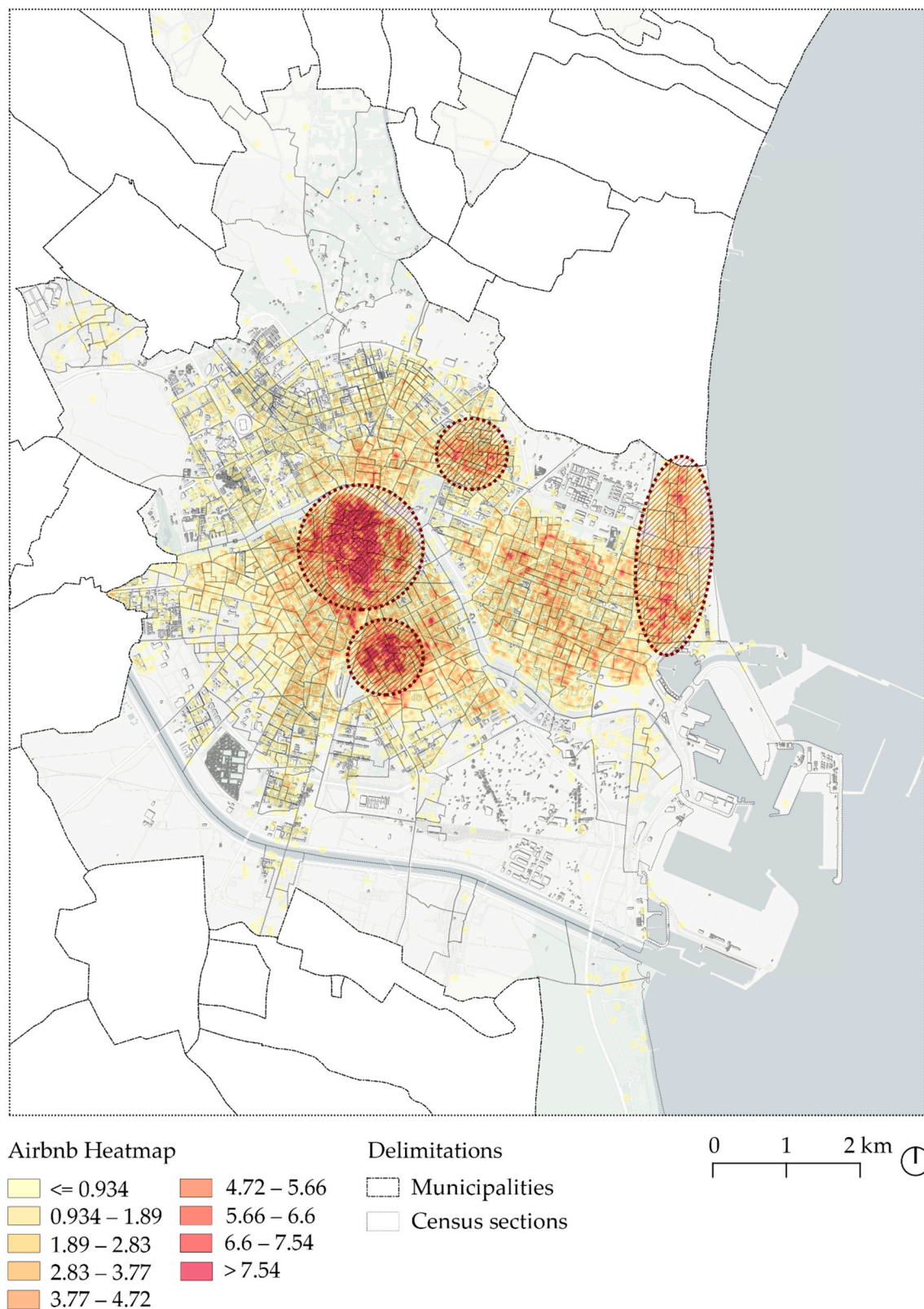


Figure 9. Population density per age group and Idealista data density.





**Figure 10.** Foursquare’s top10 venues cluster and population density per census section.



**Figure 11.** Airbnb properties density (heatmap).

As to the correlation analysis (Table 2), the results showed a positive and significant correlation between population and social networks Google Places and Idealista, remarkably so in the latter case. For the four-intervals-disaggregated population, correlation values between the first three age ranges (19 and under, 20–39 and 40–64) and Idealista

listings were similar. However, for the fourth age range (65 and over), even though the correlation was still significant, it was notably lower than that reached in the three previous cases—at 5% instead of 1%. For the remaining social networks, although weaker, a positive and significant correlation between the population aged between 20–39 years and Airbnb and Twitter was observed. The population with the higher age ranges (40 to 64 and 65 and over) correlated positively with Google Places, with a slightly higher value in the case of the older age range.

**Table 2.** Pearson correlation between population age, income and LBSNs data.

	INE Total Population	Age < 19	Age 20–39	Age 40–64	Age > 65	Google Places	Foursquare	Twitter	Airbnb	Idealista	Rental Idealista	Rental Airbnb
age < 19	0.883 **											
age 20–39	0.910 **	0.750 **										
age 40–64	0.967 **	0.881 **	0.835 **									
age > 65	0.538 **	0.188 **	0.424 **	0.412 **								
Google Places	0.100 *	0.072	0.067	0.086 *	0.129 **							
Foursquare	0.060	0.053	0.040	0.050	0.069	0.886 **						
Twitter	0.067	0.076	0.081 *	0.071	−0.027	0.603 **	0.566 **					
Airbnb	0.012	−0.039	0.083 *	0.015	−0.031	0.627 **	0.625 **	0.518 **				
Idealista	0.339 **	0.326 **	0.352 **	0.321 **	0.103 *	0.587 **	0.540 **	0.471 **	0.506 **			
Rental Idealista	−0.074	0.008	−0.114 **	−0.074	−0.070	0.489 **	0.529 **	0.273 **	0.375 **	0.231 **		
Rental Airbnb	−0.040	0.010	−0.063	−0.049	−0.026	0.346 **	0.334 **	0.253 **	0.451 **	0.264 **	0.431 **	
Income	−0.167 **	−0.130 **	−0.192 **	−0.186 **	−0.008	0.576 **	0.585 **	0.276 **	0.360 **	0.203 **	0.684 **	0.334 **

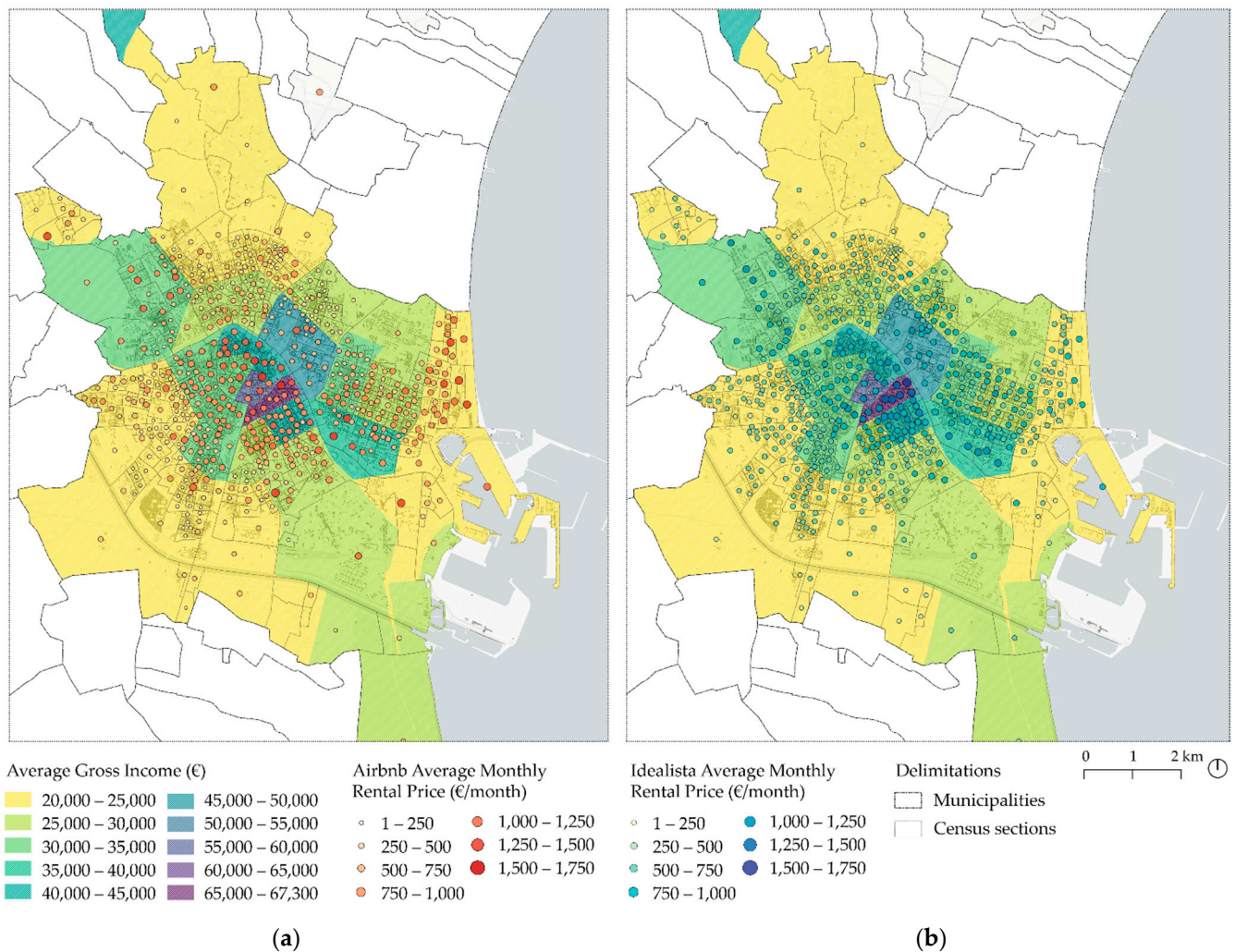
\*\* The correlation is significant at the 0.01 level (bilateral). \* The correlation is significant at the 0.05 level (bilateral).

## 5.2. Land Value, Average Income Distribution and Social Networks Data

Unlike the case of population density distribution, Airbnb properties with the highest monthly rental price were located in the historic center, the surroundings of the City of Arts and Sciences and the coastline, a finding that was not surprising given the tourist nature of these areas. In the case of Idealista, although the price distribution was more homogeneous than the touristic properties, the *Ensanche* area close to *Ciutat Vella* stood out as the location where properties with the highest rental prices were located. The historic center and coastal areas maintained a similar land value, which was distributed homogeneously throughout the central strip of the city from east to west.

In relation to the distribution of the average gross income, the *Ensanche* area between the historic center and the *Gran Vía del Marqués del Túria* axis was the area where this parameter was significantly higher than other nearby areas in the city.

Through the comparison of the different economic parameters, certain coincidences were found between the average income distribution and the Idealista properties rental price distribution, corresponding to permanent or long-term rental accommodation (Figure 12 right). However, the rental price distribution of non-regulated touristic properties registered on Airbnb did not follow the same patterns as the income distribution. Indeed, it is worth highlighting that some of the most expensive Airbnb properties were located in areas where the lowest income was registered, such as the coastal zone (Figure 12 left).



**Figure 12.** Average gross income and properties' monthly rental price for (a) Airbnb and (b) Idealista.

In the comparison between average income distribution and the LBSNs' data density distribution, it was observed that the income level of the population does not necessarily have an influence on the social networks data density. Indeed, in areas of the historic center, where data density values were similar, the average gross income values were different. Moreover, in the case of Twitter and Airbnb social networks, the amount of data diminished significantly in those census sections with higher incomes.

The results of the correlation analysis showed a significant and positive correlation of almost 0.6 between the average gross income and the social networks data presence from Google Places and Foursquare. There was also a positive and significant correlation, although with lower values, in the case of the other three social networks: Twitter, Airbnb and Idealista. A positive correlation of 0.7 was also observed between the average gross income and average rental prices on Idealista. However, the correlation value obtained for Airbnb rental prices and income was somewhat lower. These results showed, on the one hand, a strong relationship between Google Places and Foursquare data presence in those census sections with higher income; and, on the other hand, an equally strong relationship between Idealista rental prices and the average gross income. This correlation was not as significant in the case of Airbnb prices, which may indicate the different market segmentation generated by these two platforms. Idealista properties are generally used for medium/long-term rentals and are usually occupied by residents. However, Airbnb properties encourage short-term rentals, primarily intended for tourism and therefore rental to non-local residents, whose income is not collected in the census sections analyzed.

In relation to the average rental prices for both Idealista and Airbnb and the number of observations on each of these two networks per census section, it was noted that the correlation was significant and positive. This evidenced that the areas where a greater number of properties from both social networks were clustered tended to have higher average monthly prices. It must be borne in mind that these values focused on the offer of both markets, disregarding whether the properties were occupied or not. A future line of research could include demand data values that would allow for an examination of whether the increase in property availability and price values was related to an increase in demand. Nevertheless, the results obtained from this study may be showing the effect that the offered properties generate on attracting new properties, which take existing prices as a guide in setting their own prices.

### 5.3. Social Networks Data Density Comparison

A higher density of data distribution was noted in central areas of the city, especially in the historic center (*Ciutat Vella*) and in the *Ensanche*. In the case of Google Places and Foursquare, the transition between these two areas tended to be continuous, while in the case of Twitter and Airbnb the distribution was more polarized. In addition, Airbnb had a remarkable relevance in coastal areas in comparison with the other social networks analyzed.

As shown in Figure 13, the density distribution patterns of the social networks data in relation to that of Google Places data (which represented the offer of economic activity in the city) showed a strong coincidence, especially with Foursquare and Twitter data (which represented preference of activities and people presence, respectively). Consequently, those areas where the density of economic activities (according to Google Places data) was higher, was where citizens' preferred places were located (according to Foursquare data); the presence of users was higher (according to Twitter data); and more Airbnb accommodations were available for rent, at a higher monthly rental price than in other areas of the city. As mentioned earlier, these observations were found to be unrelated to the citizens' income, the long-term rental properties' monthly price, and the age of the census population in these areas.

The correlation analysis relating to social network data density showed a positive and strong correlation among all social networks. Data densities from Google Places and Foursquare were strongly correlated, having obtained a significant value at 1% of 0.9. Furthermore, there was also a positive correlation reaching a value of 0.6 between Google Places, Twitter, Idealista and Airbnb. These results suggested that, at least in the case study adopted, the information from one of these sources could be used as an indicator of a degree of social activity since the activity observed on each of them correlated positively and strongly with the others.

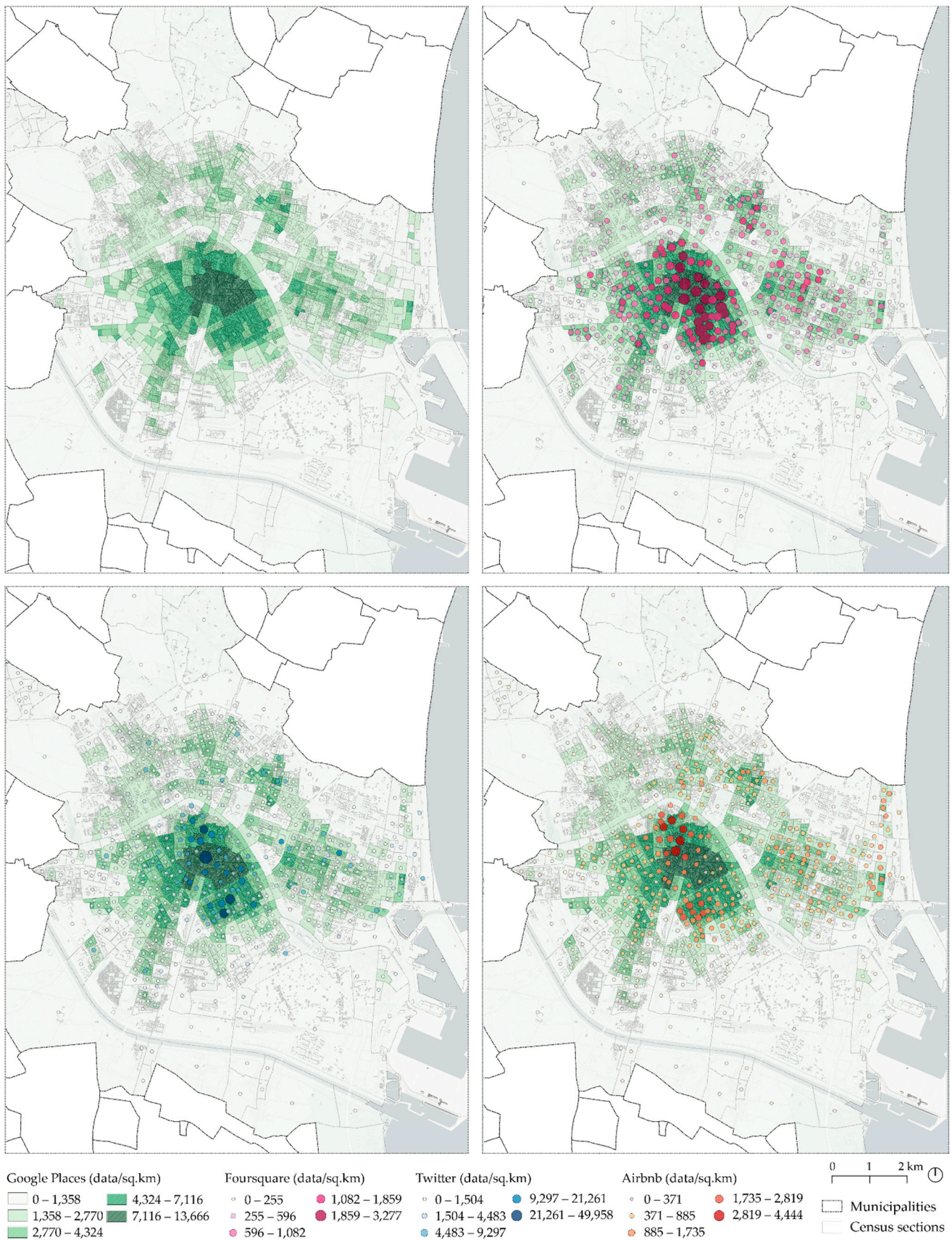


Figure 13. Google Places (economic activities) data density and Foursquare, Twitter and Airbnb data density.

## 6. Discussion

User-generated urban data brings new possibilities for data analytics. In particular, integrating data from different sources enables the understanding of urban environments by introducing a wide variety of viewpoints [76,77]. However, the constant changes in functionality, data accessibility and penetration of sources, such as location based social networks, raise new research questions and debates on their reliability and applicability for urban studies. One of the most significant debates regarding the representativeness of LBSN data in urban areas is related to the demographic determinants of the users generating the content, that is, their age, gender, or income status in relation to the actual demographic profile of the area where the content is being generated. This study aimed to build upon prior research that aimed to shed light on the representativeness of five different LBSN sources—Google Places, Foursquare, Twitter, Airbnb and Idealista—establishing relationships between statistical data from administrative sources and the location and density of the user-generated LBSN data. In this regard, it should be stated that this research was specifically concerned with the demographic determinants of the areas where LBSN data were shared, but not those of the actual population that shared the data. Although it is not within the scope of this research, future work may consider a deeper examination of the relationship between the demographics of users who share LBSN data in specific areas of the city, and that of the residents.

Methodologically, the research proposed a twofold analysis: on the one hand, that of the relationship between the presence of LBSN data and the socio-economic conditions; and on the other, the relationship between the data location and concentration of the networks themselves. These relationships were analyzed graphically through visualizations, as well as through a correlational analysis, as a means by which the observed relationships in the visualizations could be statistically characterized.

According to the results, the hypothesis was not entirely supported. Findings proved that each type of data analyzed presented a different relationship with the other and with the socio-economic and demographics of the area in which they are generated. However, it was proved that regardless of the socioeconomic status of an urban area, the location and density of economic activity—represented in this case by Google Places data—had a strong influence on where data from other sources were shared. This is probably the reason why several studies have found these specific social networks to be complementary for characterizing different urban phenomena [1,78,79].

Specifically for the case of Valencia, the central areas of the city had greater relevance than other areas in the city, as that was where the higher density of data points were found, especially in the historic center and the *Ensanche* areas (see Figure 1). These findings concurred with previous studies that suggested that city centers concentrated higher user-generated content density [44]. In line with this fact, the most popular places and points of interest of the city were also located in these areas, as reflected by the number of Foursquare check-ins, especially in public spaces and food, shopping, and entertainment-related venues. Indeed, Foursquare data density was also highest in the historic center, where the majority of the most frequently visited places were clustered, despite having the lowest population density for all age groups. This finding was consistent with previous research [34] and evidenced the touristic character of Valencia's city center. Areas with high Twitter data density did not necessarily coincide with more densely populated areas. Even though the distribution of tweets was spread throughout the city, a higher concentration was found in the historic center, which, as pointed out earlier, had a low population density. Findings also showed that locations with high Twitter data density did not follow a single income distribution pattern, because differences between wealthier and poorer areas in terms of data density were not significant in the case study analyzed. However, the areas where the population were aged between 20–39 years old had positive correlation with Twitter data.

Moreover, a consideration that should be taken into account in regard to the data points' spatial distribution is the fact that social networks are used differently depending,

among other factors, on the geographical context. For instance, in a certain geographic region the registers may be predominantly related to the residential category, whereas in Valencia, the Foursquare venues under the residential category were not significant. This could be one of the reasons why Foursquare data density in Valencia did not extend to suburban areas. However, entertainment and shopping areas stand out in the number of check-ins, especially in the city center area. Therefore, there was a variability in the number of registers per category, the total number of check-ins, and where they were located. In this respect it is worth noting that Airbnb was the only social network with relevant data density outside the central areas of the city—more specifically in the coastal areas, where Twitter, Foursquare and Google Places data density was lower.

The location of Airbnb and Idealista registers had an important correlation with monthly rental prices, but the results suggested that in both networks the dynamics to determine the value of a property were different. For instance, the highest prices for Airbnb were located in the city center and near the coast, where the number of most frequently visited places (according to Foursquare check-ins) and the number of services (according to Google Places urban activities) were higher, even though the number of touristic properties was also significant. Conversely, the properties with the highest monthly rental prices on Idealista were located in the *Ensanche*, the area with the highest income level of the city. Moreover, although the number of properties in this case was lower, the number of services was significantly large in this area.

## 7. Concluding Remarks

A series of conclusions can be drawn regarding the relationship found between the spatial distribution and concentration of various LBSNs (specifically, Google Places, Foursquare, Twitter, Airbnb and Idealista) and the socio-economic and demographic determinants (population age and average gross income) selected for this study. The overlaying of these data has allowed interesting insights to be obtained that suggest that, at least in the case study city, data from the analyzed LBSNs may not be entirely representative of the local socio-economic and demographic profile. Concentrations of these data were located in areas of low population density, with moderate income levels and a significant number of points of interest and economic activities on offer, some of which were the most visited places in the city. In addition, rental prices in these areas tended to be higher than in other neighborhoods, for both touristic and long-term rental properties. However, there was a slight correlation between land value (namely, Airbnb and Idealista rental prices) and the amount of data from social networks. Higher data density could be found in areas where the land value or rents were not at the highest rates, and lower data density was found in areas where land value and rent prices increased considerably—such as areas of the *Ensanche*, adjacent to the historic center. There was no significant relationship between the age of the population and the amount of data from social networks that were being generated. The only clear tendency observed was that in areas with lower population densities in all age groups, the LBSN data density was higher. Finally, the results showed a remarkable coincidence between Google Places data density patterns—which had been used as a proxy indicator of the offer of economic and urban activities in the city—and the presence and density of Foursquare, Twitter and Airbnb data. That is to say, the location patterns of social networks data in an urban area did not depend on local demographics or land values as much as on the existence of economic activities.

**Author Contributions:** Conceptualization, Pablo Martí, Leticia Serrano-Estrada, Álvaro Bernabeu-Bautista and V. Raul Perez-Sanchez; methodology, Pablo Martí, Leticia Serrano-Estrada, Álvaro Bernabeu-Bautista, V. Raul Perez-Sanchez; validation, Pablo Martí, Leticia Serrano-Estrada, Álvaro Bernabeu-Bautista and V. Raul Perez-Sanchez; formal analysis, V. Raul Perez-Sanchez; investigation, Pablo Martí, Leticia Serrano-Estrada, Álvaro Bernabeu-Bautista and V. Raul Perez-Sanchez; data curation, Álvaro Bernabeu-Bautista; writing—original draft preparation, Álvaro Bernabeu-Bautista and Leticia Serrano-Estrada; writing—review and editing, Pablo Martí, Leticia Serrano-Estrada, Álvaro Bernabeu-Bautista. and V. Raul Perez-Sanchez; visualization, Álvaro Bernabeu-Bautista;



Supervision, Pablo Martí, Leticia Serrano-Estrada and V. Raul Perez-Sanchez; Funding Acquisition, Pablo Martí, Leticia Serrano-Estrada, Álvaro Bernabeu-Bautista and V. Raul Perez-Sanchez. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was co-funded by (i) Conselleria de Innovación, Universidades, Ciencia y Sociedad Digital, Generalitat Valenciana and European Social Fund (ACIF/2020/173) and (ii) Conselleria de Innovación, Universidades, Ciencia y Sociedad Digital, Generalitat Valenciana (GV/2021/177).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Martí, P.; Serrano-Estrada, L.; Nolasco-Cirugeda, A. Social Media data: Challenges, opportunities and limitations in urban studies. *Comput. Environ. Urban Syst.* **2019**, *74*, 161–174. [[CrossRef](#)]
- Martin, M.E.; Schuurman, N. Social Media Big Data Acquisition and Analysis for Qualitative GIScience: Challenges and Opportunities. *Ann. Am. Assoc. Geogr.* **2020**, *110*, 1335–1352. [[CrossRef](#)]
- Yang, D.; Qu, B.; Cudre-Mauroux, P. Location-Centric Social Media Analytics: Challenges and Opportunities for Smart Cities. *IEEE Intell. Syst.* **2020**, *36*, 3–10. [[CrossRef](#)]
- Carpio-Pinedo, J.; Gutiérrez, J. Consumption and symbolic capital in the metropolitan space: Integrating ‘old’ retail data sources with social big data. *Cities* **2020**, *106*, 102859. [[CrossRef](#)]
- Chen, T.; Hui, E.C.M.; Wu, J.; Lang, W.; Li, X. Identifying urban spatial structure and urban vibrancy in highly dense cities using georeferenced social media data. *Habitat Int.* **2019**, *89*, 102005. [[CrossRef](#)]
- Ebrahimpour, Z.; Wan, W.; Velázquez García, J.L.; Cervantes, O.; Hou, L. Analyzing Social-Geographic Human Mobility Patterns Using Large-Scale Social Media Data. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 125. [[CrossRef](#)]
- Latzko-Toth, G.; Bonneau, C.; Millette, M. Small Data, Thick Data: Thickening Strategies for Research, Trace-based Social Media. In *The SAGE Handbook of Social Media Research Methods*; Sloan, L., Quan-Haase, A., Eds.; SAGE Publications Ltd.: London, UK, 2017; pp. 199–212, ISBN 1473916321.
- Folch, D.C.; Spielman, S.E.; Manduca, R. Fast Food Data: Where User-Generated Content Works and Where It Does Not. *Geogr. Anal.* **2018**, *50*, 125–140. [[CrossRef](#)]
- Yuan, Y.; Lu, Y.; Chow, T.E.; Ye, C.; Alyaqout, A.; Liu, Y. The Missing Parts from Social Media-Enabled Smart Cities: Who, Where, When, and What? *Ann. Am. Assoc. Geogr.* **2020**, *110*, 462–475. [[CrossRef](#)]
- Fekete, E. Foursquare in the city of fountains: Using Kansas city as a case study for combining demographic and social media data. In *Thinking Big Data in Geography: New Regimes, New Research*; University of Nebraska Press: Lincoln, NE, USA, 2017; pp. 165–188, ISBN 9781496205377.
- Santani, D.; Gatica-Perez, D. Speaking swiss: Languages and venues in foursquare. In Proceedings of the MM 2013—2013 ACM Multimedia Conference, Barcelona, Spain, 22 October 2013; ACM Press: New York, NY, USA, 2013; pp. 501–504.
- Li, D.; Liu, J. Uncovering the relationship between point-of-interests-related human mobility and socioeconomic status. *Telemat. Inform.* **2019**, *39*, 49–63. [[CrossRef](#)]
- Dudás, G.; Vida, G.; Kovalcsik, T.; Boros, L. A socio-economic analysis of Airbnb in New York City. *Reg. Stat.* **2017**, *7*, 135–151. [[CrossRef](#)]
- Longley, P.A.; Adnan, M.; Lansley, G. The geotemporal demographics of Twitter usage. *Environ. Plan. A* **2015**, *47*, 465–484. [[CrossRef](#)]
- Murthy, D.; Gross, A.; Pensavalle, A. Urban Social Media Demographics: An Exploration of Twitter Use in Major American Cities. *J. Comput. Commun.* **2016**, *21*, 33–49. [[CrossRef](#)]
- Jiang, Y.; Li, Z.; Ye, X. Understanding demographic and socioeconomic biases of geotagged Twitter users at the county level. *Cartogr. Geogr. Inf. Sci.* **2019**, *46*, 228–242. [[CrossRef](#)]
- Lerman, K.; Arora, M.; Gallegos, L.; Kumaraguru, P.; Garcia, D. Emotions, demographics and sociability in Twitter interactions. In Proceedings of the 10th International Conference on Web and Social Media, ICWSM 2016, Cologne, Germany, 17–20 May 2016; pp. 201–210.
- White, R.; Engelen, G.; Uljee, I. *Modeling Cities and Regions as Complex Systems*; MIT Press: Cambridge, MA, USA, 2015, ISBN 9780262029568.
- Pinheiro Gawryszewski, V.; Scarlazzari Costa, L. Homicídios e desigualdades sociais no Município de São Paulo. *Rev. Saude Publica* **2005**, *39*, 191–197. [[CrossRef](#)] [[PubMed](#)]
- Marquet, O.; Miralles-Guasch, C. Walking short distances. The socioeconomic drivers for the use of proximity in everyday mobility in Barcelona. *Transp. Res. Part A Policy Pract.* **2014**, *70*, 210–222. [[CrossRef](#)]

21. Grujic, N.; Brdar, S.; Novovic, O.; Govedarica, M.; Crnojevic, V. Evidence of urban segregation from mobile phone data: A case study of Novi Sad. In Proceedings of the 27th Telecommunications Forum, TELFOR 2019, Belgrade, Serbia, 26–27 November 2019; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2019.
22. Tu, W.; Zhu, T.; Xia, J.; Zhou, Y.; Lai, Y.; Jiang, J.; Li, Q. Portraying the spatial dynamics of urban vibrancy using multisource urban big data. *Comput. Environ. Urban Syst.* **2020**, *80*, 101428. [[CrossRef](#)]
23. Frias-Martinez, V.; Virseda, J. On the relationship between socio-economic factors and cell phone usage. In *ACM International Conference Proceeding Series*; ACM Press: New York, NY, USA, 2012; pp. 76–84.
24. Yardi, S.; Bruckman, A. Income, race, and class: Exploring socioeconomic differences in family technology use. In *Conference on Human Factors in Computing Systems—Proceedings*; ACM Press: New York, NY, USA, 2012; pp. 3041–3050.
25. Penard, T.; Poussing, N.; Yebe, G.Z. Comparing the Determinants of Internet and Cell Phone Use in Africa: Evidence from Gabon. *Commun. Strateg.* **2012**, *1*, 65–83.
26. Deville, P.; Linard, C.; Martin, S.; Gilbert, M.; Stevens, F.R.; Gaughan, A.E.; Blondel, V.D.; Tatem, A.J. Dynamic population mapping using mobile phone data. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 15888–15893. [[CrossRef](#)]
27. Bayir, M.A.; Demirbas, M.; Eagle, N. Mobility profiler: A framework for discovering mobility profiles of cell phone users. *Pervasive Mob. Comput.* **2010**, *6*, 435–454. [[CrossRef](#)]
28. Demissie, M.G.; Phithakkitnukoon, S.; Sukhvilul, T.; Antunes, F.; Gomes, R.; Bento, C. Inferring Passenger Travel Demand to Improve Urban Mobility in Developing Countries Using Cell Phone Data: A Case Study of Senegal. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 2466–2478. [[CrossRef](#)]
29. Smoreda, Z.; Olteanu-Raimond, A.-M.; Couronné, T. Spatiotemporal Data from Mobile Phones for Personal Mobility Assessment. In *Transport Survey Methods: Best Practice for Decision Making*; Emerald Group Publishing Ltd.: Bingley, UK, 2013; pp. 745–768.
30. Osorio Arjona, J.; García Palomares, J.C. Nuevas fuentes y retos para el estudio de la movilidad urbana. *Cuad. Geográficos* **2017**, *56*, 247–267.
31. Li, H.; Wei, Y.D.; Wu, Y.; Tian, G. Analyzing housing prices in Shanghai with open data: Amenity, accessibility and urban structure. *Cities* **2019**, *91*, 165–179. [[CrossRef](#)]
32. Ye, Y.; Xie, H.; Fang, J.; Jiang, H.; Wang, D. Daily Accessed Street Greenery and Housing Price: Measuring Economic Performance of Human-Scale Streetscapes via New Urban Data. *Sustainability* **2019**, *11*, 1741. [[CrossRef](#)]
33. Shen, Y.; Karimi, K. Urban function connectivity: Characterisation of functional urban streets with social media check-in data. *Cities* **2016**, *55*, 9–21. [[CrossRef](#)]
34. Ballatore, A.; De Sabbata, S. Los Angeles as a digital place: The geographies of user-generated content. *Trans. GIS* **2020**, *24*, 880–902. [[CrossRef](#)]
35. Rizwan, M.; Wanggen, W.; Gwiazdzinski, L. Visualization, Spatiotemporal Patterns, and Directional Analysis of Urban Activities Using Geolocation Data Extracted from LBSN. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 137. [[CrossRef](#)]
36. Muhammad, R.; Zhao, Y.; Liu, F. Spatiotemporal Analysis to Observe Gender Based Check-In Behavior by Using Social Media Big Data: A Case Study of Guangzhou, China. *Sustainability* **2019**, *11*, 2822. [[CrossRef](#)]
37. Shelton, T.; Poorthuis, A.; Zook, M. Social media and the city: Rethinking urban socio-spatial inequality using user-generated geographic information. *Landsc. Urban Plan.* **2015**, *142*, 198–211. [[CrossRef](#)]
38. Jaidka, K.; Guntuku, S.C.; Lee, J.H.; Luo, Z.; Buffone, A.; Ungar, L.H. The rural–urban stress divide: Obtaining geographical insights through Twitter. *Comput. Human Behav.* **2020**, *114*, 106544. [[CrossRef](#)]
39. Salas-Olmedo, M.H.; Moya-Gómez, B.; García-Palomares, J.C.; Gutiérrez, J. Tourists’ digital footprint in cities: Comparing Big Data sources. *Tour. Manag.* **2018**, *66*, 13–25. [[CrossRef](#)]
40. Blanford, J.L.; Huang, Z.; Savelyev, A.; MacEachren, A.M. Geo-located tweets. Enhancing mobility maps and capturing cross-border movement. *PLoS ONE* **2015**, *10*, e0129202. [[CrossRef](#)]
41. Blank, G.; Lutz, C. Representativeness of Social Media in Great Britain: Investigating Facebook, LinkedIn, Twitter, Pinterest, Google+, and Instagram. *Am. Behav. Sci.* **2017**, *61*, 741–756. [[CrossRef](#)]
42. Laman, H.; Eluru, N.; Yasmin, S. Using location-based social network data for activity intensity analysis. *J. Transp. Land Use* **2019**, *12*, 723–740. [[CrossRef](#)]
43. Hu, W. *Dynamic Origin Destination Estimation with Location-Based*; The state University of New Jersey: New Brunswick, NJ, USA, 2019. [[CrossRef](#)]
44. Khan, N.U.; Wan, W.; Yu, S. Location-Based Social Network’s Data Analysis and Spatio-Temporal Modeling for the Mega City of Shanghai, China. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 76. [[CrossRef](#)]
45. Gutiérrez, J.; García-Palomares, J.C.; Romanillos, G.; Salas-Olmedo, M.H. The eruption of Airbnb in tourist cities: Comparing spatial patterns of hotels and peer-to-peer accommodation in Barcelona. *Tour. Manag.* **2017**, *62*, 278–291. [[CrossRef](#)]
46. Instituto Nacional de Estadística Padrón. Población por Municipios. Available online: [https://www.ine.es/dyngs/INEbase/categoria.htm?c=Estadistica\\_P&cid=1254734710990](https://www.ine.es/dyngs/INEbase/categoria.htm?c=Estadistica_P&cid=1254734710990) (accessed on 28 September 2021).
47. OECD. Stat- Metropolitan Areas. Available online: <https://stats.oecd.org/> (accessed on 11 August 2020).
48. GaWC—Globalization and World Cities Research Network the World According to GaWC. Available online: <https://www.lboro.ac.uk/gawc/> (accessed on 28 September 2021).
49. Gil-Alonso, F.; López-Villanueva, C.; Bayona-i-Carrasco, J.; Pujadas, I. Towards an Even More Spatially Diversified City? New Metropolitan Population Trends in the Post-Economic Crisis Period. *Urban Sci.* **2021**, *5*, 41. [[CrossRef](#)]

50. Guadalajara, N.; López, M.Á.; Iftimi, A.; Usai, A. Influence of the cadastral value of the urban land and neighborhood characteristics on the mean house mortgage appraisal. *Land* **2021**, *10*, 250. [CrossRef]
51. Serra, P.; Vera, A.; Tulla, A.F.; Salvati, L. Beyond urban–rural dichotomy: Exploring socioeconomic and land-use processes of change in Spain (1991–2011). *Appl. Geogr.* **2014**, *55*, 71–81. [CrossRef]
52. Instituto Nacional de Estadística. Encuesta de Ocupación Hotelera. Available online: <https://www.ine.es/dynt3/inebase/es/index.htm?padre=238&dh=1> (accessed on 28 September 2021).
53. Instituto Nacional de Estadística. Encuesta de Ocupación de Alojamientos Turísticos. Available online: [https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica\\_C&cid=1254736176962&menu=ultiDatos&idp=1254735576863](https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176962&menu=ultiDatos&idp=1254735576863) (accessed on 28 September 2021).
54. Adamiak, C.; Szyda, B.; Dubownik, A.; García-Álvarez, D. Airbnb Offer in Spain—Spatial Analysis of the Pattern and Determinants of Its Distribution. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 155. [CrossRef]
55. García-Amaya, A.M.; Temes-Cordovez, R.; Simancas-Cruz, M.; Peñarrubia-Zaragoza, M.P. The Airbnb effect on areas subject to urban renewal in Valencia (Spain). *Int. J. Tour. Cities* **2021**, *7*, 361–390. [CrossRef]
56. Nieuwland, S.; del Romero Renau, L. The impact and regulatory issues of Airbnb in a mid-sized city: Valencia (Spain). In *A Modern Guide to the Urban Sharing Economy*; Edward Elgar Publishing: Northampton, MA, USA, 2021; pp. 103–118. [CrossRef]
57. IAB. *Estudio Anual Redes Sociales 2020*; IAB Spain: Madrid, Spain, 2020; p. 71.
58. The Social Media Family. VI Estudio sobre los Usuarios de Facebook, Twitter e Instagram en España. Available online: [https://www.amic.media/media/files/file\\_352\\_2282.pdf](https://www.amic.media/media/files/file_352_2282.pdf) (accessed on 28 September 2021).
59. Gallego, C.Z.; Carrasco, J.S. Geolocalized tweets for assessing daily mobility: Methodology to analyse and detect homelocation in the urban area of Valencia. *Bol. Asoc. Geogr. Esp.* **2018**, *79*, 2464. [CrossRef]
60. E. Life Monitor. Uso de Foursquare en España. Available online: <https://t-position.com/el-uso-de-foursquare-en-espana-infografia/> (accessed on 28 September 2021).
61. AirDNA | Short-Term Rental Data & Analytics | Vrbo & Airbnb Data. Available online: <https://www.airdna.co/> (accessed on 10 June 2021).
62. Hosteltur. Las 8 Ciudades de España Donde más ha Crecido Airbnb | Hoteles y Alojamientos. Available online: [https://www.hosteltur.com/127379\\_8-ciudades-espana-donde-ha-crecido-airbnb.html](https://www.hosteltur.com/127379_8-ciudades-espana-donde-ha-crecido-airbnb.html) (accessed on 28 September 2021).
63. Martí, P.; Serrano-Estrada, L.; Nolasco-Cirugeda, A. Using locative social media and urban cartographies to identify and locate successful urban plazas. *Cities* **2017**, *64*, 66–78. [CrossRef]
64. Arribas-Bel, D.; Bakens, J. Use and validation of location-based services in urban research: An example with Dutch restaurants. *Urban Stud.* **2019**, *56*, 868–884. [CrossRef]
65. Martí, P.; García-Mayor, C.; Serrano-Estrada, L. Identifying opportunity places for urban regeneration through LBSNs. *Cities* **2019**, *90*, 191–206. [CrossRef]
66. Zervas, G.; Proserpio, D.; Byers, J.W. The Rise of the Sharing Economy: Estimating the Impact of Airbnb on the Hotel Industry. *J. Mark. Res.* **2017**, *54*, 687–705. [CrossRef]
67. Sans, A.A.; Quaglieri, A. Unravelling Airbnb: Urban perspectives from Barcelona. In *Reinventing the Local in Tourism: Producing, Consuming and Negotiating Place*; Channel View Publications: Bristol, UK, 2016; pp. 209–228.
68. Domínguez, M.; Vilà, G. Los centros urbanos de las grandes ciudades españolas: Una perspectiva demográfica y urbanística. Los casos de de Madrid y Barcelona. In *Actas del XI Congreso Español de Sociología*; Universidad Complutense de Madrid: Madrid, España, 2013; pp. 166–181.
69. Donat, C. El impacto desigual de la crisis inmobiliaria en los barrios de Barcelona: Un análisis a partir de los precios de las viviendas. In *Nuevos Escenarios Urbanos Nuevos Conflictos y Nuevas Políticas. XIII Coloquio de Geografía Urbana*; Documenta Universitaria: Girona, España, 2017; pp. 14–28.
70. García-Palomares, J.C.; Salas-Olmedo, M.H.; Moya-Gómez, B.; Condeço-Melhorado, A.; Gutiérrez, J. City dynamics through Twitter: Relationships between land use and spatiotemporal demographics. *Cities* **2018**, *72*, 310–319. [CrossRef]
71. Osorio-Arjona, J.; García-Palomares, J.C. Social media and urban mobility: Using twitter to calculate home-work travel matrices. *Cities* **2019**, *89*, 268–280. [CrossRef]
72. Schreck, T.; Keim, D. Visual analysis of social media data. *Computer* **2013**, *46*, 68–75. [CrossRef]
73. Bernabeu-Bautista, Á.; Serrano-Estrada, L.; Martí, P. Social Media Data Collection and Quality for Urban Studies. In *Principles of Social Networking. The New Horizon and Emerging Challenges*; Springer: Singapore, 2021; pp. 229–248.
74. Huang, Q.; Wong, D.W.S. Activity patterns, socioeconomic status and urban spatial structure: What can social media data tell us? *Int. J. Geogr. Inf. Sci.* **2016**, *30*, 1873–1898. [CrossRef]
75. Instituto Geográfico Nacional. Capítulo 9. Sociedad. In *España en Mapas. Una Síntesis Geográfica*; Centro Nacional de Información Geográfica: Madrid, España, 2019; pp. 228–247, ISBN 9788441648456.
76. McMillan, D.; Engström, A.; Lampinen, A.; Brown, B. Data and the city. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, San Jose, CA, USA, 7–12 May 2016; pp. 2933–2944. [CrossRef]
77. Offenhuber, D.; Ratti, C. *Decoding the City: Urbanism in the Age of Big Data*; Birkhäuser: Basilea, Switzerland, 2015, ISBN 978-3038215974.

- 
78. Kovacs-györi, A.; Ristea, A.; Havas, C.; Mehaffy, M.; Hochmair, H.H.; Resch, B.; Juhasz, L.; Lehner, A.; Ramasubramanian, L.; Blaschke, T. Opportunities and Challenges of Geospatial Analysis for Promoting Urban Livability in the Era of Big Data and Machine Learning. *Int. J. Geo-Inf.* **2020**, *9*, 752. [[CrossRef](#)]
  79. Miliadis, V.; Psyllidis, A. Assessing the influence of point-of-interest features on the classification of place categories. *Comput. Environ. Urban Syst.* **2021**, *86*, 101597. [[CrossRef](#)]