*Article*

# Urban Vibrancy: An Emerging Factor that Spatially Influences the Real Estate Market

Alice Barreca, Rocco Curto and Diana Rolando *

Architecture and Design Department, Politecnico di Torino, Castello del Valentino, Viale Mattioli 39, 10125 Torino, Italy; alice.barreca@polito.it (A.B.); rocco.curto@polito.it (R.C.)
* Correspondence: diana.rolando@polito.it; Tel.: +39-011-090-6435

check for updates

**Abstract:** Urban vibrancy is defined and measured differently in the literature. Originally, it was described as the number of people in and around streets or neighborhoods. Now, it is commonly associated with activity intensity, the diversity of land-use configurations, and the accessibility of a place. The aim of this paper is to study urban vibrancy, its relationship with neighborhood services, and the real estate market. Firstly, it is used a set of neighborhood service variables, and a Principal Component Analysis is performed in order to create a Neighborhood Services Index (NeSI) that is able to identify the most and least vibrant urban areas of a city. Secondly, the influence of urban vibrancy on the listing prices of existing housing is analyzed by performing spatial analyses. To achieve this, the presence of spatial autocorrelation is investigated and spatial clusters are identified. Therefore, spatial autoregressive models are applied to manage spatial effects and to identify the variables that significantly influence the process of housing price determination. The results confirm that housing prices are spatially autocorrelated and highlight that housing prices and NeSI are statistically associated with each other. The identification of the urban areas characterized by different levels of vibrancy and housing prices can effectively support the revision of the urban development plan and its regulatory act, as well as strategic urban policies and actions. Such data analyses support a deep knowledge of the current status quo, which is necessary to drive important changes to develop more efficient, sustainable, and competitive cities.

**Keywords:** urban vibrancy; real estate market; housing prices; Neighborhood Services Index; spatial analyses; ESDA; LISA; Spatial Autoregressive (SAR) Models; Turin

## 1. Introduction

The recent economic-financial crisis of 2008 caused a global transformation of the economy. The massive shrinkage and the general development slowdown are still significantly influencing socio-economic dynamics and inducing a radical transformation of the classical paradigms of the real estate market. The abovementioned structural economic crisis (rather than a cyclical one) with its effects on the real estate market necessitates the study of urban development and dynamics with a new perspective in order to find new factors and rules that influence and govern classical paradigms that have changed. In particular, it is necessary to study the new phenomena that are guiding social behaviors and purchasing criteria in the housing sector, as well as the new (economic and social) spatial hierarchies that are defined by urban policies and real estate market trends. In this framework, the Italian context is rather critical. the Italian construction and real estate sectors are still in crisis and have not returned to 2008 levels, in contrast to other European countries. Therefore, different issues need to be faced with the support of indicators and indexes that are able to capture and analyze urban changes and new real estate market dynamics. Some urban structural changes have been recently studied by analyzing urban vibrancy in relation to different economic and social factors.

The concept of urban vibrancy is not new. It was introduced by Jacobs [1,2], who described urban vitality as street life over a 24-h period, and it was improved by Montgomery [3,4], who related this concept to the number of people in and around streets or neighborhoods at different times of the day and night. Assuming that "successful urban places are based predominantly on street life, and the various ways in which activity occurs in and through building and spaces" [4] (p. 6), several studies were developed to study the connection between spatial vitality (or vibrancy), urban spatial form, and urban quality of life. Although the definitions of urban vibrancy differ slightly, it is commonly recognized that it is closely associated with the attraction, diversity, and accessibility of a place [5], so that the activity intensity (in terms of neighborhood services) can be used as a proxy to measure urban vibrancy [6]. Although urban vibrancy is widely studied in the literature with several purposes, there no studies, at least to our knowledge, which have provided evidence that urban vibrancy is related to the real estate market, taking into account the spatial dimension of housing prices.

The identification of the most and the least vibrant urban areas can effectively support strategic policies for the sustainable development of a city. In particular, once the urban areas characterized by a low vibrancy or low housing prices are identified, specific actions or urban development plans can be implemented. In recent studies, we analyzed two dimensions opposed to urban vibrancy—social and housing vulnerability—in order to highlight their relationship with the real estate market. We demonstrated that the most vulnerable sectors of the population and buildings' physical features are spatially autocorrelated, and significantly and negatively influence housing prices [7,8]. Assuming these results, this study goes further by analyzing the urban vibrancy at the neighborhood scale and by investigating its influence on real estate submarkets. The analysis of these two spatial dimensions—urban vibrancy and vulnerability—and their relationship with the real estate market can therefore support the revision of an urban development plan and its regulatory acts, as well as strategic urban policies and actions, in order to increase the economic and social welfare in a city.

Therefore, the aim of this paper is twofold. Firstly, it aims to study urban vibrancy and its relationship with neighborhood services by creating an index that is able to identify the most vibrant urban areas of a city. Secondly, it aims to spatially analyze the influence of urban vibrancy in the housing price determination process, with particular reference to existing residential building stock in Turin.

In this study, the city of Turin is used as a case study to study the concept of urban vibrancy and its relationship with property prices in order to support strategic urban policies. The city of Turin represents an interesting case study since its municipality recently started a review process of their urban development plan. As a result, it is crucial to spatially analyze the new urban phenomena and the socio-economic and real estate market trends in order to support Turin's public administration in this important and delicate phase. In fact, the review of the urban development plan represents a key opportunity to develop new urban services, improve infrastructure (such as bicycle lanes and green spaces), and foster the establishment of public services, commercial activities, and cultural hubs on the fringes of the city by attracting new public and private investors.

Our pivotal questions were as follows: What role does urban vibrancy and its relationship with housing prices play in the context of urban regeneration? Can the identification of the most and the least vibrant urban areas and real estate submarkets help the municipality of Turin to address specific planning strategies?

Assuming the concept that a vibrant neighborhood is characterized by a high intensity and diversity of land-use configurations, in this study, a Neighborhood Services Index (NeSI) was created and adopted as a suitable proxy to measure urban vibrancy. Neighborhood services include, among others, accessibility to public transport, accessibility to local commercial activities, accessibility to schools and cultural buildings (such as museums and theatres), and accessibility to green areas. To create NeSI multi-source spatial data, integration and cleaning processes were addressed by using the Extract/Transform/Load (ETL) method [9,10] and a series of neighborhood services variables were defined, standardized, and clustered by using Principal Component Analysis (PCA). Exploratory Spatial Data Analyses (ESDA) were performed in this study, the presence of spatial autocorrelation

was investigated by calculating the Moran's Index, and LISA statistics were obtained to show which spatial autocorrelation types were present. Furthermore, after testing Ordinary Least Squares (OLS) models, spatial autoregressive (SAR) models were applied to study the influence of NeSI—and its Principal Components—on Existing Building Listing Prices (LP). Results showed that in central and historical areas of Turin, urban vibrancy is strictly associated with the real estate market and acts as a multiplier of housing prices. Nevertheless, in the most vulnerable areas of the city, urban vibrancy does not significantly influence the real estate market, since there are other social and housing factors that have a stronger and negative influence on prices.

The paper proceeds as follows: Section 2 introduces the background of the analysis. The methodological framework, the study area, and data are described in Section 3, while Section 4 discusses the results. The conclusions and discussion are presented in the final section.

## 2. Background

### 2.1. Urban Vibrancy

The concept of vibrancy has been widely studied and different factors have been used to evaluate it. Jacobs [1,2] was the first to introduce and describe urban vitality in terms of street life over a 24-h period. This definition was improved by Montgomery [3,4] who suggested that urban vibrancy could be described as the number of people present in all around streets or neighborhoods during the day and the night and could be related to different land uses. The interest in evaluating urban vibrancy has been progressively increased and different factors have been used to quantitatively measure it. Some recent studies, according to the definition of vibrancy by Jacobs [1,2], have analyzed people's activities at different locations and times by using social media check-in data [5,11] or numbers of mobile phone users [6] in a 24-h period as a proxy of urban vibrancy.

In addition to the number of people in a location, other aspects are also being used to evaluate urban vibrancy: accessibility and connectivity [12], night-time light data [13], housing prices [14], and built-environment attributes [15]. Therefore, in studying urban vibrancy, the first key issue is to find a suitable proxy to precisely measure it and, consequently, to select a suitable dataset. The second issue is to develop relevant variables, indicators, or indexes that affect urban vitality. In fact, the aim of several studies has been to understand which factors are associated with urban vibrancy by using different kinds of data and applying different methods.

Sharkova et al. [16] used five different data sources and applied the Ordinary Least Squares (OLS) method to measure the influences of neighborhood types, land uses, socioeconomic characteristics, and urban accessibility on vibrancy. These authors found that the concentration of civic places in neighborhoods positively influences neighborhood vitality, as represented by stable residential patterns, a higher degree of racial integration, and lower per capita crime rates. Census block groups were the units of analysis used to examine these neighborhood characteristics for 1990 and 1996. Neighborhood vibrancy may also facilitate the formation of social capital by providing opportunities for civic engagement and informal interactions and it has been concluded that land-use and density indicators are important for studying neighborhood quality. Overall, the regression equation predicting "percent of stayers" and "Change in per capita rate of crime against persons" performed reasonably well, predicting 62% and 55% of the variation in the dependent variables. The regression equation predicting "change in racial isolation" did not perform as well, predicting only 12% of the block group variation. The location of schools was the only type of civic organization that was significant in two out of the three regression models.

Yue et al. [6] used Points of Interest (POI) data from navigation databases to develop a series of mixed-use indicators that are able to reflect the multifaceted and multidimensional characteristics of mixed and multiple use neighbours at the building level. Therefore, their study used the total number of people in a neighborhood recorded by mobile phone cell towers as a proxy for neighborhood vibrancy. Therefore, they performed a series of linear regressions and demonstrated that POI agglomeration

significantly contributes to improving neighborhood vibrancy. The POI database used in this study has 274,022 POIs that belong to 64 subcategories of 15 primary categories.

POI data were also used by Wu et al. [5] and Lu et al. [11]. Wu et al. [5] investigated the quantitative relationship between land-use patterns and the vibrancy spatio-temporal distribution features by integrating kernel density estimation (KDE), geographically and temporally weighted regression (GTWR), and the Herfindahl–Hirschman index (HHI). A total of 510,635 POIs were analyzed and classified into 12 types according to the classification of the AutoNavi POI and the Code for the Classification of Urban Land Use and Planning Standards of Development Land (GB50137-2011). The categories used to measure the heterogeneity were housing-related POIs (HPOI), consumption-related POIs (CPOI), and traffic-related POIs (TPOI). The number of HPOIs, CPOIs, TPOIs, and other POIs (OPOI) were measured in each grid. Their results showed that vibrancy evolution is influenced by several factors that are heterogeneous over time and space, and that the degree of clustering of POIs has a significant influence on housing prices. Lu et al. [11] performed three regression models to analyze the relationship between neighborhood vibrancy and urban form and demonstrated that building density and functional diversity are positively correlated with neighborhood vibrancy.

Other studies did not explicitly refer to the concept of vibrancy, but developed and used indicators and indexes that can also be considered suitable for the study of urban land use mix and vitality. For example, Verma et al. [17] collected POI data from different open source and online datasets and extracted the information about urban land use types and their functions by grouping the various labels into 16 classes (vehicle repair/services, night life, personal care, administrative offices, educational institutions, food, general stores and establishment, medical services, services, recreational, hotels, religion, financial, specialized stores, transportation, and general establishments). On the other hand, they classified local climate zones (LCZ) into eight classes. The results showed that a trend in POI diversity is present along the major roads of the city due to the quick access for business to other parts of the city. LCZ, on the other hand, showed a clustering pattern. LCZ diversity is concentrated in wards in the south, north, and central urban areas of Bombay, which is due to the inclusion of different built typologies such as high, mid, and low rise (slums) and water, dedicated green spaces, and the presence of small forests. Winters et al. [15] investigated the effects of different built environment features (grouped into four general categories: land use, physical environment, bicycle facilities, and the road network) on healthy transportation choices by performing multilevel logistic models. The results indicated that the built environment has a significant influence on cycling, even though different factors were important within each spatial zone.

In several studies, hedonic pricing models have been applied to analyze the effects of different neighborhood services on housing prices. For example, Geng et al. [18] showed that the establishment of a high-speed rail station positively influences housing prices, while electromagnetic radiation pollution, traffic congestion, noise, and higher crime rates negatively affect them. Also, Dai et al. [19] studied the influence of a rail station on surrounding housing prices by performing a hedonic price model and considering 21 independent variables grouped into four types (traffic service facilities, construction characteristics, location characteristics, and neighborhood characteristics). The sample contained 2964 housing units including 598 residential units around a transfer station and 2366 residential units around non-transfer stations. The results showed that, after synthesizing the positive and negative effects, the impact of a transfer station on housing prices within 200 m was negative, while impact of the non-transfer station was positive.

Jang and Kang [20] focused their attention on the retail sector and investigated spatial accessibility and proximity effects on housing prices. The multilevel hedonic price model was used to isolate differential effects of accessibility by retail type in housing price submarkets by managing housing attributes, location and transportation characteristics, and neighborhood land-use features. They compared the effects of retail accessibility and proximity between a full and a submarket model. The full model included 30,012 cases and 2465 census tract units, and the submarket model included five areas of the city. The results confirmed the non-linear effects of micro-level retail accessibility on housing

submarkets and also revealed that the accessibility of, and proximity to, different retail types has spatially differential effects (positive or negative) on housing prices. In particular, heterogeneous effects depend mainly on retail attributes, submarket socio-economical features, and the spatial distribution of retail stores.

Finally, Yuan et al. [21] performed both an OLS model and a GWR model to study how a set of variables (including structure, neighborhood, accessibility, and amenity variables) influences the spatiotemporal dynamics of housing prices in Nanjing. The results showed that the presence of high-quality schools provided by governments, proximity to parks, the presence of (sub) Central Business Districts (CBDs) and government service centers, and the locations of new cities/towns are strongly associated with housing prices.

### 2.2. Housing Prices and Spatial Analyses

Many urban housing markets are spatially and/or structurally segmented, and, over time, this has become a working assumption for many scholars [22–24]. Submarkets arise when the spatial clustering of neighboring housing units within a non-heterogeneous real estate market is combined with the demand for specific housing features that are not common to the entire urban area [25–30]. The necessity to understand which drivers have the role of changing submarket boundaries over time suggests that different aspects that commonly affect urban housing market need to be studied more deeply [24]. As confirmed by a wide range of literature, territorial segmentation in sub-markets and territorial units represent the different sets of neighborhood features well, such as housing, schools, green areas, social centers, public spaces, or police departments in each part of the urban area [7,31,32]. In real estate research, the importance of market segmentation for property price prediction is assumed to explain property prices across space. Therefore, the integration of approaches for modeling spatial effects has been fundamental [20,33].

As commonly recognized, data from one location, such as property prices, tend to exhibit similar values to those from nearby locations [34]. The nearer a house's location is to positive or negative attributes, the higher or lower positives and negative externalities should be [7,35]. In reference to building quality, properties in close proximity tend to have similar structural characteristics such as construction period, square meter living area, and design features, and this can cause spatial autocorrelation in property prices [32,36,37]. Also, citizens in the same neighborhood may follow similar commuting patterns and have similar access to urban and collective services [24,37,38].

Moreover, in the real estate market, values of properties in the same neighborhood capitalize on shared location amenities [32,37,39] and building and urban features such as buildings typologies, population density, density of commercial activities, and public services, which play crucial roles in the price determination process and affect the "location similarity" [40].

According to Anselin [40], the notion of "location similarity" is crucial in the definition of spatial autocorrelation and submarkets, defining those locations, referred to as "neighbors" by references [41,42], for which the values of the analyzed variables are correlated. In this regard, it is assumed that housing units located in one neighborhood are likely to have similar neighborhood values, and hence hedonic prices have no significant differences [43]. This implies that these dwellings can be related to the same topographically based submarket, which generally groups more neighborhoods. Therefore, it is possible that two different neighborhoods may generate the same level of neighborhood value for a housing unit.

If spatial autocorrelation is present, it affects Ordinary Least Squares (OLS) residuals which, instead of appearing randomly distributed, exhibit a regular pattern over space [37]. Spatial statistical models manage autocorrelation issues for the statistical improvements that can be gained in hedonic modeling [32,44]. Due to the enhancement of data technologies and the ever increasing availability of geographical information, the estimation of hedonic regression models has grown substantially over recent years, developing different spatial regression models [45]. In the very extensive list of spatial models, Global and Local Indicators of Spatial Autocorrelation (GISA and LISA), among ESDA

techniques, and Spatial Lag and Error Models (SLM and SEM), among Spatial Autoregressive models (SAR), have established themselves as widely used methods with lattice data [46–48].

## 3. Case Study, Materials, and Methods

According to the aims of this research, the most and least vibrant urban areas of a city (characterized by an high density of neighborhood services) and the real estate submarkets have to be identified and analyzed in order to support the Public Administration in addressing specific planning strategies, such as the revision of the urban development plan.

In this study, the city of Turin and its territorial segmentation into Statistical Zones was used as a case study (Figure 1), even if, in principle, this framework can be generalized and transferred to any other study area. The use of administrative territorial segmentation (94 Statistical Zones), as is commonly used in the literature [49,50], and the use of submarkets for studying the real estate market, as widely recognized in the literature [24,51], were chosen. In the case of Turin, the 94 Statistical Zones (SZ) can be assumed to be submarkets, as demonstrated in previous research [7,8].
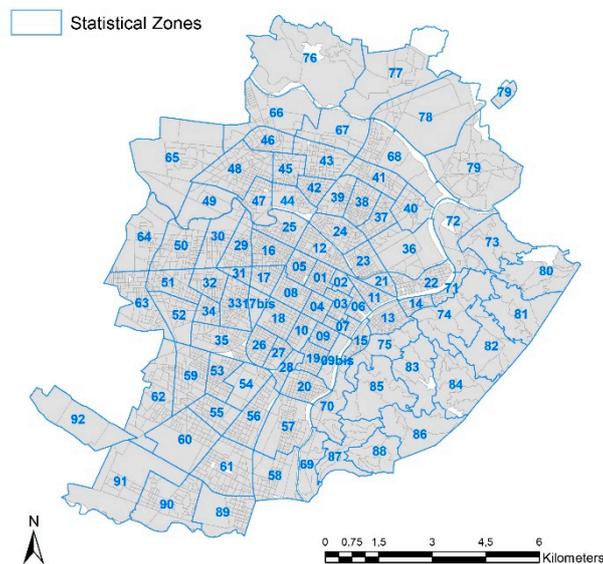


**Figure 1.** The 94 Statistical Zones of the city of Turin (Source: Authors' elaboration).

A methodological approach based on four steps was developed by applying a Principal Component Analysis and widely recognized methods of spatial analyses and spatial regression models (Figure 2).
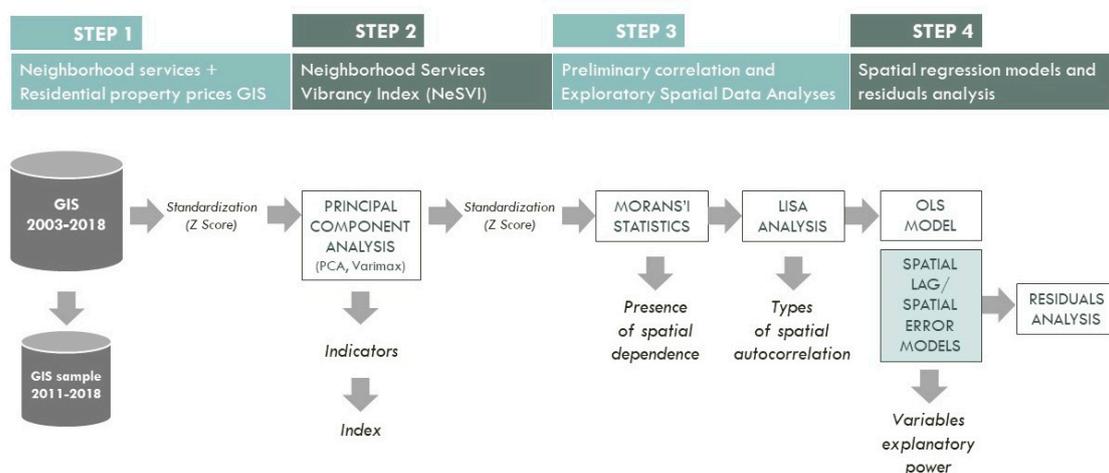


**Figure 2.** Methodological approach (Source: Authors' elaboration).

Firstly, a GIS was developed, including more than 300 urban and housing price variables. A sample of property listings published in the time period 2011–2018 was then considered in order to analyze the shrinkage period that negatively influenced the Turin real estate market in the last eight years. Secondly, a series of variables were defined, standardized, and clustered by using Principal Component Analysis (PCA) (Table 1). Then, the resulting set of factors/indicators (Principal Components) was analyzed and a Neighborhood Services Index (NeSI) was created and adopted as a suitable proxy to measure urban vibrancy.

**Table 1.** Multi-source neighborhood service variables for measuring urban vibrancy with their summary statistics: (a) TREMO data, (b) ISTAT open data, (c) Municipality of Turin Geoportal Open Data, (d) Piedmont Region Geoportal Open Data (Source: Author's elaboration).

| Dimensions | Definition | Data Sources | Mean (N/ZS km$^2$) | St. Dev (N/ZS km$^2$) |
|---|---|---|---|---|
| **TOPOGRAPHY** | | | | |
| SZ_SKM | Area of each SZ in sq km | (c), (b) | 1.383 | 1.019 |
| BLD_SKM | Density of buildings | (c), (d) | 409.991 | 288.166 |
| POP_SKM | Density of residential population | (b) | 9368.274 | 7118.101 |
| **LANDMARKS** | | | | |
| NRP_SKM | Density of Religious Places | (c) | 3.312 | 4.579 |
| NHP_SKM | Density of historical places_SKM | (c) | 0.228 | 0.800 |
| **TRANSPORT** | | | | |
| NCS_SKM | Carsharing Station Density | (d) | 0.348 | 1.009 |
| NMS_SKM * | Presence of metro stations in a distance of 500 mt | (d) | 0.320 | 1.266 |
| NAS_SKM | Bus station Density | (d) | 24.185 | 13.361 |
| NXS_SKM | Taxi station Density | (d) | 1.065 | 1.767 |
| NTS_SKM * | Presence of taxi stations in a distance of 500 m | (d) | 0.393 | 1.343 |
| PAKM_SKM * | Rate of Pedestrian Areas | (c), (d) | 0.008 | 0.020 |
| **CULTURAL OFFERINGS** | | | | |
| NL | Libraries Density | (c) | 0.085 | 0.273 |
| NC * | Cinemas Density | (c) | 0.580 | 2.140 |
| NM * | Museums Density | (c) | 1.022 | 3.189 |
| NT * | Theatres Density | (c) | 0.508 | 1.300 |
| **RETAIL** | | | | |
| C_RST_SKM * | Retail_ Cafè & Restaurants | (d) | 70.953 | 86.210 |
| C_FOD_SKM * | Retail_ Grocery stores | (d) | 53.186 | 64.320 |
| C_BET_SKM * | Retail_ Beauty | (d) | 43.042 | 43.521 |
| C_MIX_SKM * | Retail_ Mix NoFood | (d) | 102.273 | 124.768 |
| C_CLT_SKM * | Retail_ Clothes | (d) | 30.562 | 57.387 |
| C_FRE_SKM * | Retail_ Freetime | (d) | 21.923 | 25.406 |
| C_JWL_SKM * | Retail_ Jewelry and antiquities | (d) | 18.145 | 25.840 |
| C_HOM_SKM * | Retail_ Home | (d) | 14.091 | 18.215 |
| C_ELC_SKM | Retail_ electronics | (d) | 11.800 | 16.762 |
| C_VEH_SKM | Reatil_cars and fuel | (d) | 6.950 | 5.590 |
| C_HEL_SKM * | Retail_ Healthcare&Pharmacies | (d) | 7.229 | 7.646 |
| C_SUP_SKM * | Retail_ Supermarkets | (d) | 2.840 | 3.408 |
| C_FUN_SKM | Retail_Funeral services | (d) | 0.827 | 1.546 |
| NCB_SKM | Commercial buildings | (c) | 15.599 | 15.129 |
| NMK | Open Air Markets | (c) | 0.489 | 0.717 |

**Table 1.** *Cont.*

| Dimensions | Definition | Data Sources | Mean (N/ZS km²) | St. Dev (N/ZS km²) |
|---|---|---|---|---|
| **RESIDENTIAL SECTOR** | | | | |
| NRB_SKM | Residential buildings | (b) | 340.402 | 264.880 |
| RESHU | Almost occupied by one resident | (b) | 4603.040 | 3716.588 |
| VHU_SKM | Vacant housing units | (b) | 471.644 | 467.029 |
| NRESHU | Occupied by only not residents | (b) | 7.187 | 21.861 |
| THU | Total housing units | (b) | 5081.871 | 4111.561 |
| **GREEN & SPORTS** | | | | |
| NSB_SKM * | Density of Sport Buildings | (c) | 6.192 | 7.735 |
| GAKM_SKM * | Rate of green areas | (c) | 1.187 | 1.037 |
| **EDUCATION** | | | | |
| NUB * | University Density | (c) | 0.542 | 2.067 |
| NNSB | Nursery School Density | (c) | 0.578 | 0.933 |
| NS | School Density | (c) | 7.002 | 6.249 |
| **HEALTHCARE** | | | | |
| NH | Hospital Density | (d) | 0.559 | 1.434 |

* variables selected and grouped by means of PCA.

Subsequently, the value for each PC, the index, and the mean housing price were calculated in each considered territorial segment, and both ESDA and Pearson correlation tests were performed. The presence of spatial autocorrelation was investigated by calculating the Moran's Index, and LISA statistics were performed to show which types of spatial autocorrelation are present.

Finally, after LM testing OLS models, SAR models were performed and, by comparing results, the influence of the spatial effects on the whole explanatory power of the model was analyzed. Furthermore, the residuals of SLM and SEM regressions were compared to verify the absence of correlation and spatial dependence in error terms.

### 3.1. GIS Creation, Data Integration, and Standardization

A GIS was developed by collecting open data from four principal sources: the Turin Real Estate Market Observatory (TREMO), the Italian National Institute of Statistics (ISTAT), and the Municipality of Turin and Piedmont Region geoportals.

To analyze housing prices in Turin, we used a data sample that belongs to a database property of TREMO [52] that was founded in 2000 [53] and that monitors and analyzes the residential real estate market of the city yearly. In particular, this study was based on a sample of 3578 property listings in Turin published on the main Italian real estate web platform in 2011–2018. Listing prices were used as a proxy of transaction prices. Even if this is justified by the literature [54,55], it is one of the key limitations of this study, due to the unavailability of public data about transaction prices in the Italian context. This sample was selected from a whole database of about 12,590 housing units located in existing buildings and listed on the market from 2003 to 2018. After the elimination of outliers and observations with missing location, the sample LP mean price was found to be 2367 Euros per square meter with a standard deviation of 1047 Euros per square meter.

Furthermore, we chose 41 neighborhood service variables from a set of more than 300 variables to study the urban vibrancy of the city of Turin (Table 1).

Multi-source spatial data integration and cleaning processes have been addressed by using the Extract/Transform/Load (ETL) method [9,10]. In particular, hierarchical data were used to even out different data types coming from different sources so that data at smaller and bigger scales could be referred to the same territorial units (94 ZS). In this study, even if the use of lattice data was justified by the literature, the lack of available geographical open data at a fine scale or point format limited the analysis, as it excluded some interesting potential variables.

### 3.2. Principal Component Analysis (PCA) for the Clustering of Neighborhood Service Variables

The application of a reductionist technique, such as Principal Component Analysis or Factor Analysis, produces a coherent and robust set of variables that can be monitored over time to evaluate possible changes and their influences on the overall vibrancy. The technique also favors variable replication at different spatial scales, thus making data compilation more efficient. PCA is commonly used to define the more explanatory variables in a sample by grouping them and transforming the values into Principal Components (PCs), often belonging to the same aspect of a phenomena [56]. Therefore, by using PCA, we reduced the dimension of the data set and identified a series of uncorrelated and ordered PCs (Table 1).

Before applying PCA, the data standardization step is highly recommended, so that the input variables have the same magnitude. PCA produces linear combinations of the original variables in the form of a set of orthogonal components. The set of components obviously changes in relation to the selected input variables and this constitutes a methodological limitation of PCA. The first component is always the linear combination that explains the major variations among original variables, while starting from the second component, the remaining variation is progressively explained. Therefore, it is necessary to select a minimum subset of components that is able to explain the maximum underlying data features and to rotate the orthogonal components by applying a *varimax* rotation. The *varimax* rotation minimizes the number of variables that are loaded on a single factor, thus increasing the percentage variation between each factor. Then, it is necessary to interpret the resulting components by how they may influence the urban vibrancy and assign signs accordingly.

Finally, the interpreted components were summed with equal weights to create a Neighborhood Services Index (NeSI) to measure urban vibrancy.

Summing up, the computations were carried out using the following steps [57]:

1. Standardization of all input variables to *z-scores* (each with a mean of 0 and a standard deviation of (1);
2. The use of standardized input variables to perform the PCA;
3. Selection of the number of components to be further used based on the not rotated solution (Kaiser criterion).
4. Rotation of the initial PCA solution (*Varimax* rotation);
5. Interpretation of the resulting components;
6. Combination of the selected component scores into a univariate score;
7. Standardization of the resulting scores to a mean of 0 and a standard deviation of 1.

### 3.3. Pearson's Correlation Test and Exploratory Spatial Data Analyses (ESDA)

Property price data are spatial data, since they are rather dependent on their locations. In fact, data from one location tend to exhibit values similar to those of other locations nearby [34]. This may cause spatial autocorrelation that, consequently, may affect the real estate market [39,58]. Spatial autocorrelation also measures spatial dependence, which arises in lattice data whereby the correlation occurs among contiguous units [32].

ESDA mainly including GISA and LISA statistics were created to manage the spatial effects that are typical of spatial data [47], allowing the decomposition of global indicators, such as Moran's Index, into the contribution of each individual observation, such as Local Moran and Local Geary [59,60].

For these reasons, in this study, LISA statistics were calculated to evaluate the spatial autocorrelation [61,62]. In particularly the standardized *z-score* of Local Moran's I provides an assessment of the similarity of each observation with those in its surroundings [63,64]. By using GeoDa software [65], that is, based on GIS infrastructure [59], starting from the typology of the sample data, a Queen Contiguity—First Order Weight matrix (W) was generated [44].

Results of Local Moran statistics were also graphically represented on the basis of the type and value of spatial autocorrelation. The Moran Scatter Plot gives no information on the significance

of Local Moran statistics but does provide a classification of spatial association into four categories. "Spatial clusters" are defined by high values of the investigated phenomenon with a high level of similarity with their surroundings (high-high), called "hot spots", and by observations with low values and a high level of similarity with their surroundings (low-low), defined as "cold spots". Moreover, observations with high values surrounded by low ones (high-low) and the inverse (low-high) are defined as "spatial outliers".

*3.4. Spatial Regression Models and Residual Analysis*

In order to measure the influence of the five principal components and NeSI on the property listing price variable [33], a traditional Ordinary Least Squares (OLS) regression was firstly applied. Then, LM spatial dependence tests were performed, and finally, SAR models were applied to correctly manage the error correlation due to spatial effects [46]. Hedonic price models and theory were introduced primarily by Rosen [66] and Lancaster [67] and have been frequently applied to analyze and measure the intrinsic non-linearity in the relationship between heterogeneous property prices, buildings features, and neighborhood characteristics, though nothing is known a priori about a specific functional form [68].

One of the main objectives of regression analysis is to explain the variation in one variable (dependent variable) based on the variation in one or more other variables (explanatory variables). Furthermore, regressions are applied both to estimate marginal prices and to discover the strength of each explanatory variable in relation to the dependent variable Y. In fact, in the studies of the real estate market, regression models are widely used both with explanatory and/or predictive purposes [69].

To find a more suitable regression model, starting from an OLS, it is necessary to test the presence of spatial dependence between the errors or the model variables [70]. Spatial dependence tests for a single variable are based on the size of an indicator that combines the observed value in each location with the average values at neighboring locations (Spatial Lags) [34]. Basically, spatial dependence tests are measures of the similarity between value associations (covariance, correlation, or difference) and associations in space (contiguity) [34]. The spatial autocorrelation statistic is considered to be significant when it assumes an extreme value compared to what would be expected from the null hypothesis (in the absence of spatial autocorrelation).

When significant "spatial effects" exist, both spatial dependence, either globally or locally, and spatial heterogeneity also exist [34,71]. The Breusch Pagan test is used to test the spatial homogeneity assumption [64], while the Moran test and the Lagrange Multiplier tests (LM-lag and LM-error) are used for testing spatial dependence [63].

In this study, two SAR models, namely the Spatial Lag Model (SLM) and the Spatial Error Model (SEM), were performed.

The weight matrix (W), as a spatial structure, is incorporated into both the dependent variable and the error term of a general SLM. A model in which the dependent variable Y is not only a function of the independent variables X but also on the Y in nearby areas, is shown in Equation (1) in which the spatial lag operator produces a weighted average of the neighboring observations [46]:

$$
\begin{aligned}
Y &= \rho W_1 Y + X\beta + u \\
u &= \lambda W_2 u + \varepsilon.
\end{aligned}
\tag{1}
$$

In (1), W defines how much a nearby (in space) observation should influence the averaging procedure. The parameter $\rho$ is the coefficient of the spatially lagged dependent variable and measures the spatial dependency between observations. The parameter $\lambda$ is the coefficient in a spatial autoregressive structure for the disturbance $\varepsilon$. If OLS estimates of $\rho$ and $\lambda$ are biased and inconsistent, they must be estimated by maximizing the likelihood function. $W_1$ and $W_2$ are the exogenously determined spatial weight matrices, and if there are no a priori reasons to assume that the spatial interaction patterns are different, then $W_1$ and $W_2$ are assumed to be identical. However, to identify Equation (1), it is necessary to have uniquely different weight matrices. A simple procedure for estimating cross-sectional models

with both a spatial lag term and a spatial autoregressive term, as the most general spatial model in Equation (1), was described by Kelejian and Prucha [72].

A spatial autoregressive lag model (SAR) means that $\lambda$ is assumed to equal zero, as in Equation (2):

$$Y = \rho W_y + X\beta + \varepsilon. \tag{2}$$

Assuming that $\rho$ is equal to zero, as in Equation (3), a Spatial Error Model (SEM) can be derived. This is the most popular model and is also widely used in real estate economics:

$$\begin{aligned} Y &= X\beta + u \\ u &= \lambda W u + \varepsilon. \end{aligned} \tag{3}$$

Summing up, Equation (1) is the most general spatial model where we have included a spatial lag term and a spatial autoregressive term. If $\rho = \lambda = 0$, then we have an OLS model. If only $\rho = 0$, then the general model is reduced to SEM, while if only $\lambda = 0$, then it is reduced to SLM, which is a restricted version of the SEM model.

## 4. Results

The methodological approach outlined in Section 3 was applied to investigate the influence of urban vibrancy in the real estate market of the city of Turin. To achieve this aim, NeSI was built to measure urban vibrancy in relation to neighborhood services, and its influence on the price determination process was analyzed by using SAR models.

### 4.1. Geographical Information System (GIS): Urban Data and Housing Prices in the City of Turin

To analyze the variability and density of neighborhood services and analyze the urban vibrancy in relation to the real estate market of the city of Turin, the GIS described in Section 3.1 was developed. After the computation and normalization of data (to percentages, per capita, or density functions), an initial set of 43 variables was selected with reference to housing prices (Figure 3) and neighborhood services (Table 2).
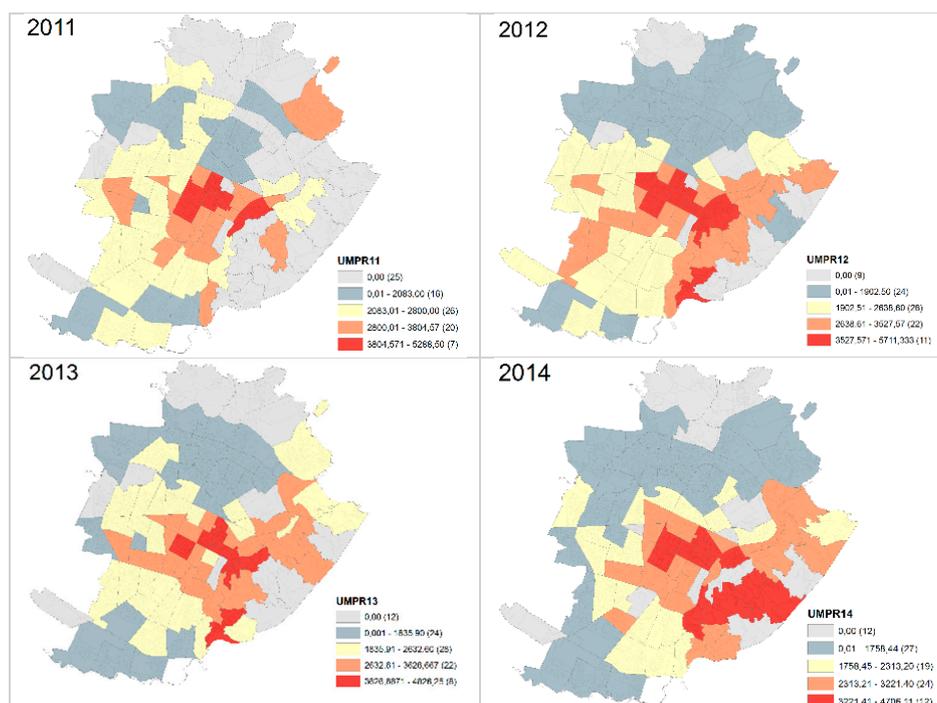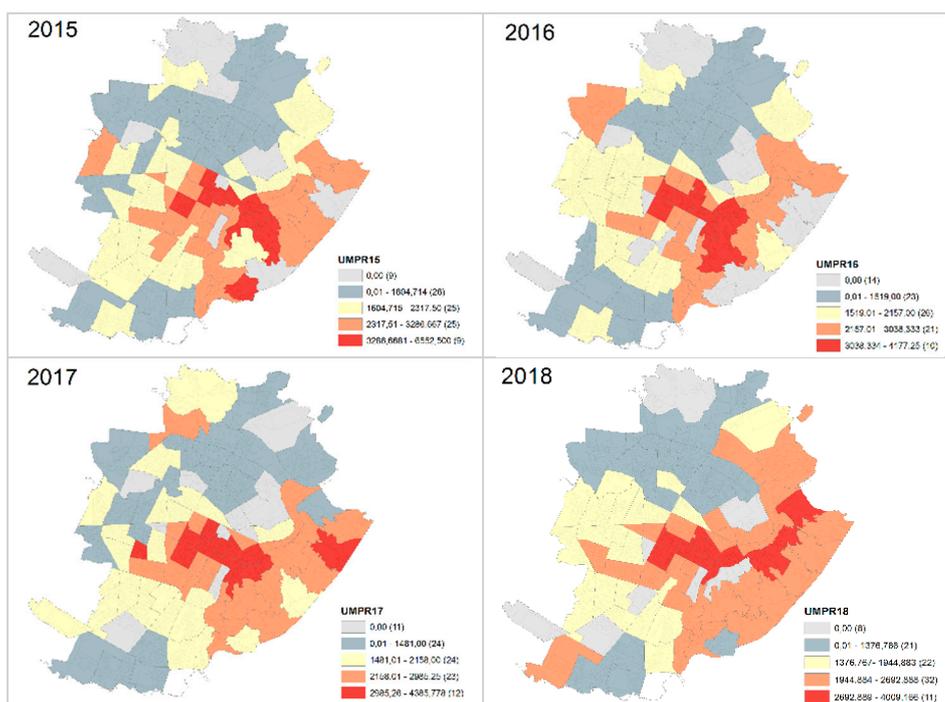


**Figure 3.** *Cont.*

**Figure 3.** Natural break maps concerning the housing prices spatial distribution in the 94 SZs in 2011–2018 (Source: Authors' elaboration).

The spatial distribution of prices in the city over an eight year time period is shown in Figure 3 by using choropleth maps, which clearly highlight five distinct price ranges in different areas of the city.

Depending on the year of the data, the price range changes, but Jenks optimization method (Jenks Natural Breaks classification method) determines the best arrangement of values into different classes. The method reduces the variance within classes and maximizes the variance between classes by minimizing each class's average deviation from the class mean and maximizing each class's deviation from the means of the other classes. In this way, colors are not comparable between years but are able to represent housing prices in more accurate classes.

The housing price spatial distribution in the 94 SZs over the considered time period showed the permanence of the highest LP in the city center, shifting from seven SZs (3800–5300 €/m$^2$) in 2011 to 11 SZs (2700–4000 €/m$^2$) in 2018. The lowest LP maintained a localization in the northern and southern fringes of the city, while the areas with average prices underwent a rotation from south in 2011 to the east area of the city (hill) in 2018. This phenomenon highlights how the housing prices and dynamics are changing, not only temporally but also spatially, reversing the previous trends.

From the whole geographical database, a set of neighborhood services variables was selected including transport stations, cultural offerings, land use (green and pedestrian areas), and commercial activities (see Table 1).

### 4.2. The Neighborhood Services Index (NeSI)

Starting from the identified neighborhood services variables, we applied a Principal Components Analysis (PCA) to reduce the number of variables on the basis of their explanation power of the global variance of the sample. To perform PCA we standardize all values of variables by using the z-scores method. From 41 input variables, PCA considered meaningful about 20 variables so that the analysis was quite efficient. To simplify the structure of the sub-dimensions and produce greater independence between the factors, a varimax rotation was used in the factor analysis.

Empirically defining the Neighborhood Services concentration to explain the urban vibrancy in each area of the city, five Principal Components (PCs) were found which explained 83.101% of the

variance among all ZSs, which were differentiated according to their relative levels of neighborhood services concentration. Each of these is briefly described in Table 2.

**Table 2.** Principal Component Analysis (PCA) results (Source: Authors' elaboration).

| Dimension and Variance Explained (%) | Cumulative Variance Explained (%) | Indicators | | Components | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | PC1 | PC2 | PC3 | PC4 | PC5 |
| **RETAIL** (50.246) + | 50.246 | Retail_ Miscellaneous | C_MIX_SKM | **0.943** | 0.248 | 0.058 | −0.099 | 0.040 |
| | | Retail_ Food | C_FOD_SKM | **0.941** | 0.155 | −0.031 | −0.051 | 0.118 |
| | | Retail_ Healthcare | C_HEL_SKM | **0.926** | 0.096 | 0.064 | −0.029 | 0.155 |
| | | Retail_ Freetime | C_FRE_SKM | **0.925** | 0.247 | 0.200 | −0.076 | −0.021 |
| | | Retail_ Home | C_HOM_SKM | **0.920** | 0.068 | 0.031 | −0.067 | −0.019 |
| | | Retail_ Beauty | C_BET_SKM | **0.918** | 0.016 | 0.174 | −0.025 | 0.120 |
| | | Retail_ Café Restaurants | C_RST_SKM | **0.895** | 0.315 | 0.025 | −0.088 | 0.096 |
| | | Retail_ Jewelry | C_JWL_SKM | **0.868** | 0.393 | 0.019 | −0.129 | −0.115 |
| | | Retail_ Clothes | C_CLT_SKM | **0.864** | 0.335 | 0.024 | −0.141 | −0.131 |
| | | Retail_ Supermarkets | C_SUP_SKM | **0.804** | 0.102 | 0.251 | −0.027 | −0.063 |
| **CULTURAL OFFERING** (11.992) + | 62.239 | Museums | NM_SKM | 0.076 | **0.880** | 0.032 | 0.040 | 0.132 |
| | | Theatres | NT_SKM | 0.111 | **0.779** | 0.014 | 0.078 | 0.139 |
| | | PedestrianAreas | PAKM_SKM | 0.359 | **0.751** | −0.105 | 0.029 | −0.093 |
| | | Universities | NUB_SKM | 0.274 | **0.737** | 0.007 | −0.117 | −0.135 |
| | | Cinemas | NC_SKM | 0.519 | **0.521** | 0.130 | −0.140 | −0.287 |
| **CONNECTIVITY** (8.912) + | 71.151 | Metro_stations | NMS_SKM | 0.117 | −0.030 | **0.968** | −0.035 | −0.002 |
| | | Train_stations | NTS_SKM | 0.188 | 0.020 | **0.960** | −0.025 | 0.001 |
| **GREEN&SPORTS** (6.703) + | 77.854 | Sport_Buildings and Areas | NSB_SKM | −0.017 | −0.083 | −0.096 | **0.840** | −0.155 |
| | | GreenAreas | GAKM_SKM | −0.215 | 0.103 | 0.042 | **0.801** | 0.067 |
| **HEALTHCARE** (5.247) + | 83.101 | Hospitals | NH_SKM | 0.119 | 0.035 | 0.002 | −0.089 | **0.892** |
| Kaiser–Meyer–Olkin Measure of Sampling Adequacy | 0.799 | Extraction method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization Rotation converged in six iterations | | | | | | |

### 4.2.1. Retail

The first factor identified the retail sector as measured by the percentage of shops selling miscellaneous goods, food, healthcare, and free−times goods, house furniture, beauty products, restaurants and café, jewelry, clothes, and supermarkets. The retail factor explained 50.246% of the variance. Retail and commercial streets enable communities to exchange products and knowhow and allow people to frequent the public parts of the city both by day and by night. On the other hand, it is evident that a lack of commerce can consistently reduce urban vibrancy. A lower shop concentration/density attracts fewer people, the streets are less crowded, and therefore, they are less secure and the local communities cannot find the products suited to their lifestyle and culture.

### 4.2.2. Cultural Offerings

The second factor, which explains 11.992% of the variation among ZSs, included museums, theatres, pedestrian areas, universities, and cinemas. The preponderance of buildings voted as cultural offerings and cultural amenities of the city load this dimension positively, as commonly recognized in the literature [73].

### 4.2.3. Connectivity

The third factor describes the degree of accessibility to public transport, measuring the density of metro stations and train stations. It explains 8.912% of the variation in ZS. This factor confirms findings in the literature: a high concentration of transport services is able to positively influence the vibrancy of urban areas [19,20].

### 4.2.4. Green and Sports

Green and open−air sports areas contribute to the vibrancy of the city in different ways. Parks and green areas are the lungs of urban areas and provide landscapes and natural environments to

citizens. Sports and leisure areas are places where citizens gather in their free-time and visitors meet for some sport events. This fourth factor explains 6.703% of the variation.

### 4.2.5. Healthcare

The fifth factor, measuring the density of hospitals, explains 5.247% of the variance. The nature of hospitals in Italy creates a spillover of uses all around their location, so the presence of healthcare can be considered to be an important component of vibrancy. Apartments and hotels around hospitals are often temporarily used to host relatives of ill people, offices of voluntary associations, and co−housing for families of people undergoing long−term hospital stays, or alternatively for cafes and restaurants.

### 4.2.6. The Neighborhood Services Index (NeSI)

The construction of the Neighborhood Services Index (NeSI) was based on the PCA, where the five identified factors explained over 83% of the variance. The sign attributed to each component was based on the literature review and on particular conditions that may change the way dimensions operate in different urban contexts [74]. By making no a priori assumptions about the importance of each dimension in the overall sum, the five principal components were aggregated by using a simple additive model. In this way, each factor was viewed as having an equal contribution to the NeSI formation and the overall vibrancy identification.

Therefore, we used the NeSI to measure the overall urban vibrancy in each SZ, where positive values of NeSI indicated higher levels of vibrancy, and negative values indicated lower levels of vibrancy. To determine the most and least vibrant ZS, NeSI scores were mapped based on standard deviations from the mean into five categories ranging from <–0.05 at the lower end to >2.5 at the upper end (Figure 4).
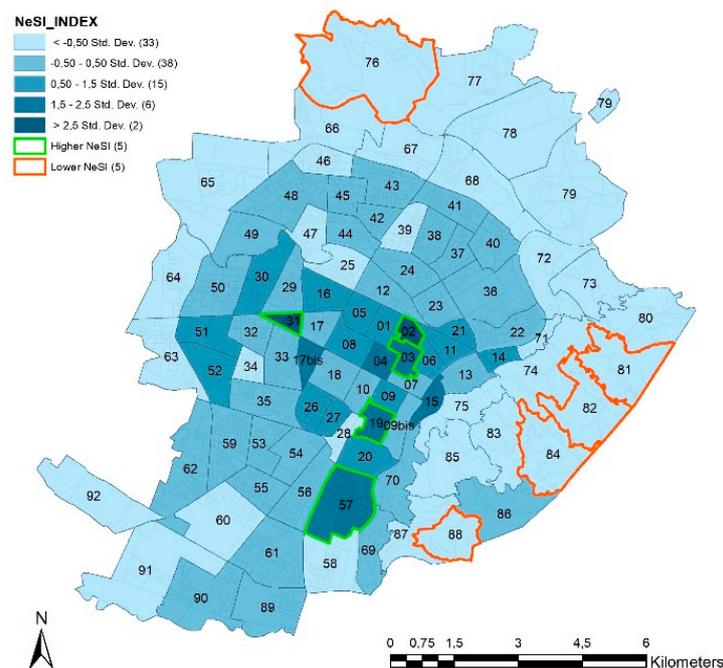


**Figure 4.** Neighborhood Services Index (NeSI) Standard deviation map (Source: Authors' elaboration).

### 4.2.7. The Geography of Urban Vibrancy

The NeSI ranged from –3.148 (low neighborhood services concentration) to 7.602 (high neighborhood services concentration) As shown in Figure 3, SZs with NeSI scores greater than +2.5 standard deviations were labeled as the most vibrant. This only included two SZs (green boundaries 31, 02): the first and central zone is Piazza Castello, the administrative, commercial,

and historical center of the city, which is often crowded by citizens, visitors, and tourists. The second zone is a part of Pozzo Strada on the subway route of Corso Francia area that is characterized by a strong housing and commercial presence, even recently, that has good urban and building quality. In the top five more vibrant SZs (green boundaries 03, 19, 57), three more SZs were shown to have an NeSI score of +1.5 standard deviations, and these were all along via Nizza, an important vehicle transport route which is the north–south subway axis. A total of 33 SZs (35% of the total) were classified into the less vibrant category. The top five least vibrant SZs (red boundaries 88, 84, 82, and 81) are located in the Hill Area of the city (East bank of River Po), largely based on the absence of commercial activities, public services (such as libraries), and public transport. One other SZ rounded out the top five least vibrant areas (76), but its absence of vibrancy was derived partially from different indicators, such as the absence of a transport station, a lack of commercial activity, a lack of schools and cultural offering, and, in general, the presence of a urban environment with a low quality level.

### 4.3. Pearson's Correlation Test and Exploratory Spatial Data Analyses (ESDA)

Exploratory Spatial Data Analyses were performed to focus on the spatial effects of the five PCs, NeSI, and housing Listing Prices (LP). Initially, the five PCs and NeSI were taken into account and their correlations with the LP variable were tested by means of a Pearson correlation test. Results showed, as expected, the absence of a correlation between a component (PC) and each other, but also showed the absence of correlations between components and LP, allowing us to perform regression analyses (Table 3). The absence of a strong correlation between NeSI and LP (0.487) was also verified.

**Table 3.** The Pearson correlation test (Source: Authors' elaboration).

|      | LP  | PC1   | PC2   | PC3   | PC4   | PC5   | NeSI  |
|------|-----|-------|-------|-------|-------|-------|-------|
| LP   | 1   | 0.312 | 0.339 | 0.220 | 0.014 | 0.187 | 0.487 |

Subsequently, the presence of local spatial autocorrelation was verified by means of ESDA statistics. The local spatial autocorrelations between different SZs were measured to understand how values are distributed in each territorial unit and their contiguous ones. In particular, in order to consider the dimensions of the geo−spatial clustering of LP, NeSI, PC1−Retail, and PC2−Cultural offer, the Moran's I scatterplot and LISA were calculated with a significance calculation (99 permutations) processed by GeoDa on the basis of Monte Carlo statistics. The significance of the clusters is guaranteed with a $p$−value of between 0.001 and 0.05. The results are shown in Figure 5.

The Moran scatter plots showed that observations fall mainly in the II and IV quadrants. This implies the presence of a positive autocorrelation across the SZ. The highest autocorrelation value was observed for the LP variable (Moran's I = 0.605), followed by NeSI (Moran's I = 0.432).

LISA identified spatial clusters that represented the highest concentrations of the highest and the lowest values of LP, NeSI, PC1−Retail, and PC2−Cultural offerings. The LISA results suggested the presence of striking geographic clustering of housing prices in the central, northern, and southern urban areas (a). A first cluster, located in the northern and southern fringes of the city, represents the urban areas characterized by a positive autocorrelation of low values (Low−Low). A second cluster, located in the city center and on the hillside, represents the urban areas characterized by a positive autocorrelation of high values (High−High). Comparing the LISA results for LP (a) and for NeSI (b), it is possible to notice a similar clustering in the northern urban area (Low−Low) and in the city center (High−High). In contrast, on the hillside, those areas with a positive autocorrelation of high property prices (High−High) correspond to areas characterized by a positive autocorrelation of a low density of neighborhood services (Low−Low).

Since the main part of the variance of NeSI is explained by PC1—Retail (c) followed by PC2—Cultural offerings (d), we decided to also verify the presence of spatial autocorrelation for these two factors. Comparing LISA results, it is possible to notice that the clustering of PC1 is quite similar

both to the NeSI one and to LP. On the other hand, the clustering of PC2 is located in the city center and four "potential spatial outliers" with high values are principally located in the south and west parts of the city, where the cultural offerings are present but there is a low level of similarity with the surrounding areas.
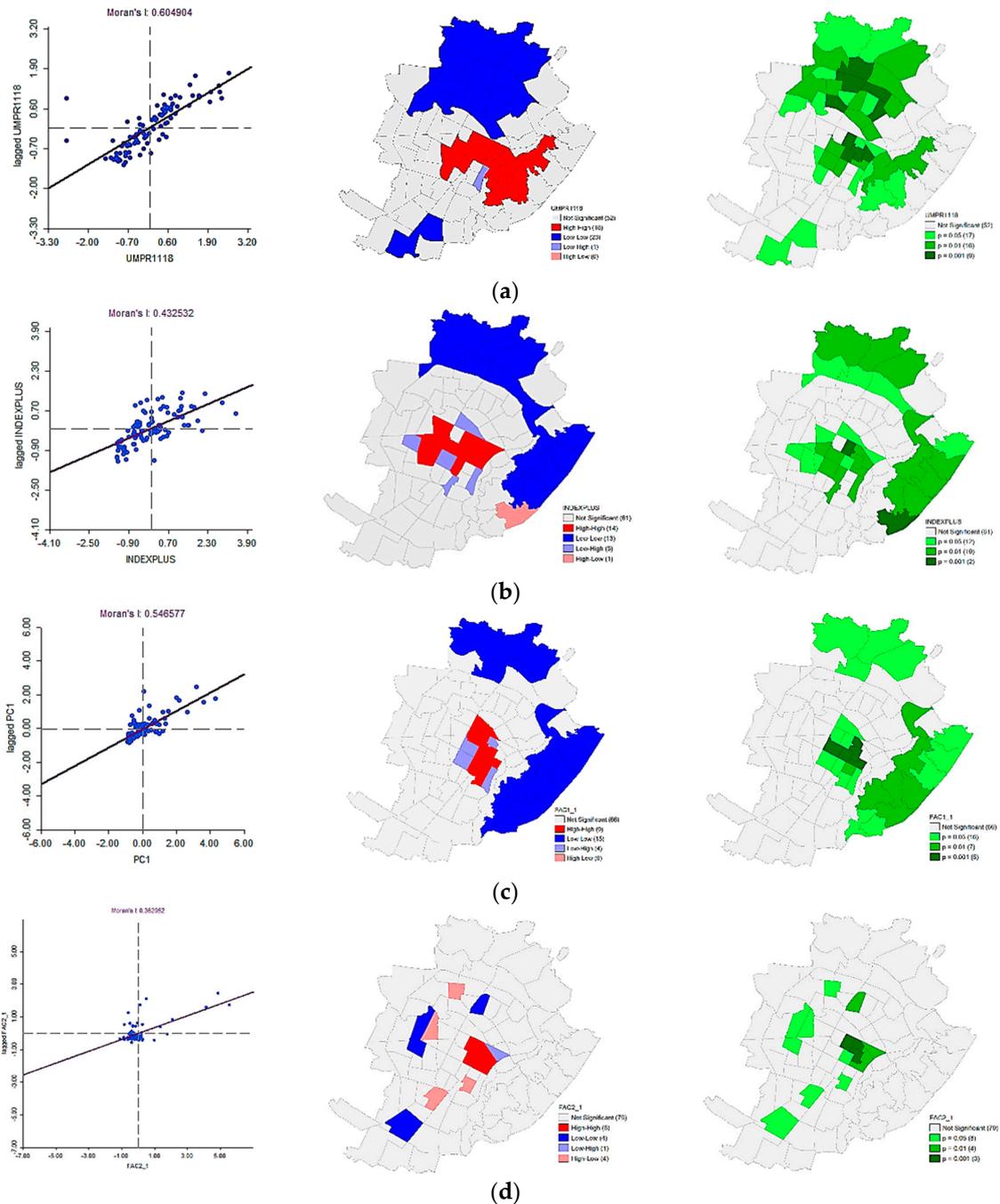


**Figure 5.** Univariate Moran's I and Moran scatter plots, Local Indicator of Spatial Association (LISA) (cluster and significance) maps depicting spatial clustering and spatial outliers in the 94 SZs of Turin: (**a**) Listing Prices (LP, Euro/m$^2$); (**b**) NeSI rate; (**c**) PC1—Retail rate; (**d**) PC2—Cultural offerings rate (Source: Authors' elaboration).

### 4.4. Spatial Regression and Residuals Analysis

In the set of Spatial Autoregressive (SAR) models, Spatial Lag and/or Spatial Error models were performed on the basis of preliminary spatial econometric model tests [75]. On the basis of the Lagrange multiplier (LM) principle, several diagnostics for spatial econometric models were performed. In particular, model misspecification due to spatial dependence (in the form of an omitted spatially lagged dependent variable and spatial residual autocorrelation) as well as spatial heterogeneity (in the form of heteroskedasticity) was detected [64].

The spatial regressions outlined in Section 3 were performed to assess the influence of the density of neighborhood services on property prices. The dependent variable was LP, while the independent variables were NeSI, in the first model (Table 4), and PC1 (Retail), PC2 (Cultural offer), PC3 (Transport), PC4 (Green and Sports), and PC5 (Healthcare) in the second one (Table 5). Since we found considerable autocorrelation in the OLS regression, we performed SLM and SEM.

To assess the influence of NeSI on LP, SEM resulted in the best model, with a higher $R^2$ value, a lower AIC value, and a non−significant Breush–Pagan test.

**Table 4.** Spatial Error Model (SEM) to assess the relationship between LP and NeSI (Source: Authors' elaboration).

| Variable | Spatial Error Model (SEM) | |
|---|---|---|
| | Coefficients | Probability |
| Spatial coefficient (λ) | 0.815 | 0.000 |
| NeSI | 90.907 | 0.002 |
| Constant | 2263.63 | 0.000 |
| Number of observations | 94 | |
| Log likelihood | −725.053 | |
| AIC | 1454.11 | |
| R square | 0.636 | |
| Breush–Pagan test for diagnosing Heteroskedasticity | 0.062 | 0.804 |
| Likelihood ratio test | 64.187 | 0.000 |

The results of the Spatial Error model (SEM) in Table 4 show a better fit of the model when the spatial dependence was managed with the use of the spatial weight matrix and the introduction of a coefficient, the disturbance (λ), in the explanatory variables. The spatial autoregressive model with spatial error dependence consists of a linear relationship between a conditional expectation of the dependent variable and its values with spatial dependent error terms in the rest of the system [76]. The R−squared value of 0.64 highlights that NeSI had a significant and positive influence on the LP variance.

To assess the influence of PCs on LP, an SLM was performed. The explanatory variables included a spatial lag for the dependent variable by using a spatial contiguity matrix (W), which is a symmetric matrix generated from topological information [77]. The findings revealed the model with the introduction of the spatial variable (W) has an R−squared value of 0.66 (Table 5). In both models, the R−squared values were not particularly high. Nevertheless, it must be taken into account that the aim of this analysis was not to implement a model for predictive purposes, including for apartment and building characteristics. This study aimed to investigate the influences of only a few factors (neighborhood services) on property prices. Therefore, the explanatory power of the model, outlined by the R−squared values, can be considered to be acceptable.

The findings of the SLM reveal that PC1 is not significant, while PC2 has a significant and positive influence on LP, and the other PCs have non−significant *p*-values. This is probably due to the concentration of museums, theaters, cinemas, and universities in the city center, which, in Turin, represents the zone with the highest housing values.

**Table 5.** Spatial Lag Model (SLM) to assess the relationship between LP and the five principal components (Source: Authors' elaboration).

| Variable | Spatial Lag Model (SLM) | |
|---|---|---|
| | Coefficients | Probability |
| Spatial coefficient (W) | 0.721 | 0.000 |
| PC1 | 43.963 | 0.403 |
| PC2 | 236.021 | 0.000 |
| PC3 | 66.606 | 0.178 |
| PC4 | −9.923 | 0.840 |
| PC5 | 47.780 | 0.336 |
| Constant | 626.079 | 0.001 |
| Number of observations | 94 | |
| Log likelihood | −718.867 | |
| AIC | 1451.73 | |
| R square | 0.665 | |
| Breush–Pagan test for diagnosing heteroskedasticity | 10.165 | 0.070 |
| Likelihood ratio test | 52.640 | 0.000 |

Furthermore, the residual correlation analyses are illustrated by two scatterplots and Moran's Index. The plots illustrate that, in both models, the use of a spatial variable ($\lambda$ or W) guarantees the absence of autocorrelation between residuals (Figure 6).



**(a)**                                      **(b)**
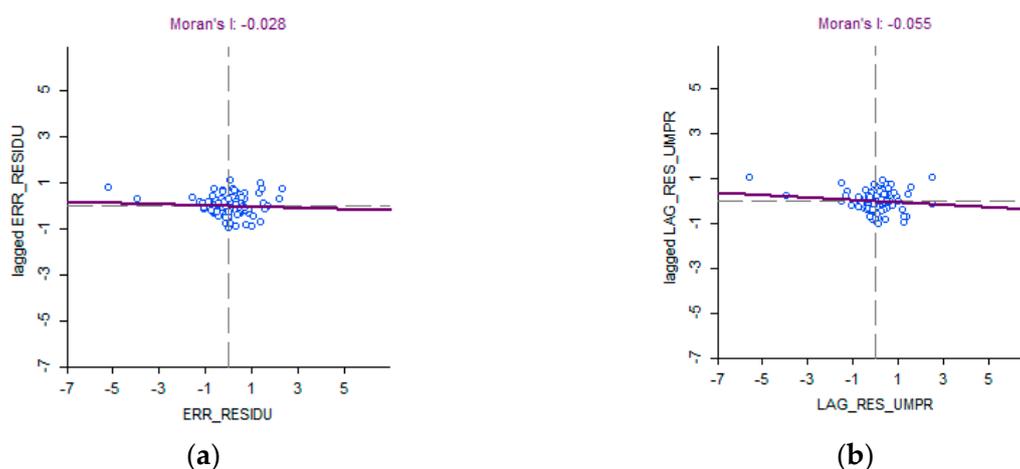
**Figure 6.** SEM residual Moran Scatter plots (**a**); SLM residual Moran Scatter plots (**b**) (Source: Authors' elaboration).

## 5. Conclusions

As part of the broader framework of research aimed at studying the vibrancy of cities and urban areas, a PCA and spatial analyses were proposed in this paper to specifically analyze the urban vibrancy and its relationship with the real estate market in the city of Turin.

The first step of the proposed methodological approach, starting from multi−source urban data integration, developed a set of 41 neighborhood services variables that cover almost all factors that are able to generate urban vibrancy. Those factors were clustered by means of a PCA (second step), which grouped 20 of the initial variables and returned five Principal Components (PCs). The resulted PCs were further aggregated and a Neighborhood Services Index (NeSI) was created. In the third step, the ESDA results highlighted a high positive correlation both in housing prices (LP) and in NeSI, as well as a certain correspondence between their spatial clusters. Therefore, spatial autoregressive models were applied to manage the spatial dimension. In particular, a Spatial Error Model (SEM) was performed and its results highlighted that NeSI had a partial but positive influence on housing

prices. Moreover, the influence of each PC on housing prices was also investigated. The outcomes of a Spatial Lag Model (SLM) highlighted that only PC2—Cultural Offerings had a significant and positive influence on property prices.

Therefore, it is possible to conclude that NeSI represents a good proxy to measure urban vibrancy. In fact, our results allow us to also identify the least vibrant areas of the city (33%) and the most vibrant ones (29%). The results show, as expected, that the northern part of the city is characterized by spatial clusters of low housing prices and low vibrancy values (with some exceptions in the northeastern part where, in recent years, a gentrification process has started) and that the central part of the city presents more vibrant areas characterized by high real estate values. Therefore, it is possible to conclude that in the central and historical areas of the city, urban vibrancy is strictly associated with the real estate market and acts as a multiplier of housing prices.

In contrast, it is interesting to focus on the hill side of the city on the right−bank of the river Po. Although this is one of the most luxurious areas of the city (with the main concentration of detached houses, villas, and private parks, inhabited by a rich population), it is also one of the least vibrant areas of the city. This because there is an almost total absence of urban services, a lack of public transport connections with the city center, a lack of shops for basic necessities, and a general isolation from urban cultural activities. Furthermore, we noted another a specific neighborhood in the city center, where there are vibrant areas characterized by low housing prices (this is the case of Porta Palazzo neighborhood, where there is an high concentration of people with a low education level, a large population of foreigners and temporary residents, and where the buildings' physical features are rather low), denoting that different kinds of vibrancy can coexist and differently influence the real estate market.

Therefore, it is possible to conclude that in the most vulnerable areas of the city, urban vibrancy does not significantly influence the real estate market, since there are other social and housing factors that have a stronger and negative influence on prices.

The analysis of the most and least vibrant urban areas has therefore to be integrated with other analyses performed, such as those related to the social and housing vulnerability, since often but not always a reverse correspondence between vibrant and vulnerable areas exists. This indicates that the presence and absence of vibrancy (and/or vulnerability) is linked with different segments of the population, whose influence and behavior are not fully absorbed by the market but that determine differences in the economic and urban development of the city.

At present, the results achieved are concretely useful for the Municipality of Turin, which started a revision process of their urban development plan, in order to implement spatially differentiated strategic policies for the sustainable development of the city. The municipality of Turin should also address specific goals in order to promote the attractiveness of the least vibrant urban areas by involving new public and private investors. Moreover, a series of challenges could be overcome by improving and fostering the neighborhood services development in the transition to a sustainable city of the future. Balancing and integrating social, cultural, economic, and environmental perspectives, in the field of urban services, the enhancement of accessibility to infrastructures and green areas, such as bicycle lanes, should therefore be strengthened. Moreover, initiatives should be promoted for the establishment of public services and commercial activities, even in the areas where they are currently absent. Hedges of cultural offerings should be rethought, also spreading cultural hubs to the fringes of the city to reactivate the dynamism of the real estate market. In this way, municipal policies could be effectively oriented towards the sustainable development of urban areas, fostering integrated social and economic welfare, shifting from the enhancement of the physical urban environment to the improvement of the current and future inhabitants' quality of life.

On the basis of these first results, further research can be developed to analyze other urban dimensions by using the GIS created that has many potentialities both in terms of the huge number of variables in it and the historical data series that can be analyzed. In fact, the housing price data analyzed in this study refer only to the existing building stock market. Another different interesting

real estate market sector to be studied could be new building and construction sites. This study could be also developed by considering other municipalities in the area surrounding Turin, since different accessibility and neighborhood service dynamics may emerge. Moreover, assuming the nature of the real estate market is influenced more by social phenomena than construction ones, the reverse correspondence between vibrancy and vulnerability will be further investigated, in order to support the municipality not only in defining its regulatory act for Turin's urban development plan but also in planning other social and economic policies. To achieve this aim, level data connected to social dimensions and to the physical features of buildings and residential units will be analyzed through the application of GWR. Finally, in further research, another aspect will be analyzed in relation to the spatial effects of data: variations of real estate values over time will be investigated by managing the time−varying spatial autoregressive coefficients as well as the time−varying regressor coefficients and cross−sectional standard deviations by means of another Spatial Autoregressive Model (SAR).

*Limitations of the Study*

One of the key limitations of this study, even if justified by the literature, was the necessity of using listing prices as a proxy for the actual transaction prices, due to the unavailability of public data about transaction prices in the Italian context.

Moreover, even if the use of lattice data is justified by the literature, the lack of available geographical open data at a fine scale or point format limited the analysis, as it excluded some interesting potential variables.

## References

1. Jacobs, J. *The Death and Life of Great American Cities. The Failure of Town Planning*; Random House: New York, NY, USA, 1961.
2. Jacobs, J. *The Life of Cities*; Random House: New York, NY, USA, 1969.
3. Montgomery, J. Urban Vitality and the Culture of Cities. *Plan. Pract. Res.* **1995**, *10*, 101–110. [CrossRef]
4. Montgomery, J. Making a city: Urbanity, vitality and urban design. *J. Urban Des.* **1998**, *3*, 93–116. [CrossRef]
5. Wu, C.; Ye, X.; Ren, F.; Du, Q. Check-in behaviour and spatio-temporal vibrancy: An exploratory analysis in Shenzhen, China. *Cities* **2018**, *77*, 104–116. [CrossRef]
6. Yue, Y.; Zhuang, Y.; Yeh, A.G.O.; Xie, J.Y.; Ma, C.L.; Li, Q.Q. Measurements of POI-based mixed use and their relationships with neighbourhood vibrancy. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 658–675. [CrossRef]
7. Barreca, A.; Curto, R.; Rolando, D. Assessing social and territorial vulnerability on real estate submarkets. *Buildings* **2017**, *7*, 94. [CrossRef]
8. Barreca, A.; Curto, R.; Rolando, D. Housing vulnerability and property prices: Spatial analyses in the turin real estate market. *Sustainability* **2018**, *10*, 3068. [CrossRef]
9. Smith, J.M.; Bernstein, P.A.; Dayal, U.; Goodman, N.; Landers, T.; Lin, K.W.T.; Wong, E. Multibase integrating heterogeneous distributed database systems. In Proceedings of the AFIPS Conference, Chicago, IL, USA, 4–7 May 1981.
10. Vassiliadis, P. A survey of extract-transform-load technology. *Int. J. Data Warehous. Min.* **2009**, *5*, 1–27. [CrossRef]
11. Lu, S.; Huang, Y.; Shi, C.; Yang, X. Exploring the Associations Between Urban Form and Neighborhood Vibrancy: A Case Study of Chengdu, China. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 165. [CrossRef]
12. Braun, L.M.; Malizia, E. Downtown vibrancy influences public health and safety outcomes in urban counties. *J. Transp. Health* **2015**, *2*, 540–548. [CrossRef]
13. Mellander, C.; Lobo, J.; Stolarick, K.; Matheson, Z. Night-time light data: A good proxy measure for economic activity? *PLoS ONE* **2015**, *10*, 1–10. [CrossRef]

14. Wu, C.; Ye, X.; Ren, F.; Wan, Y.; Ning, P.; Du, Q. Spatial and social media data analytics of housing prices in Shenzhen, China. *PLoS ONE* **2016**, *11*, 1–10. [CrossRef] [PubMed]

15. Winters, M.; Brauer, M.; Setton, E.M.; Teschke, K. Built environment influences on healthy transportation choices: Bicycling versus driving. *J. Urban Health* **2010**, *87*, 969–993. [CrossRef] [PubMed]

16. Sharkova, I.V.; Sanchez, T.W. *An Analysis of Neighborhood Vitality: The Role of Local Civic Organizations*; Portland State University: Portland, OR, USA, 2017.

17. Verma, D.; Jana, A. How diverse are the neighbourhoods? A diversity index to assess land use mix through open source and online datasets. In Proceedings of the 12th World Congress of RSAI, Goa, India, 29 May–1 June 2018.

18. Geng, B.; Bao, H.; Liang, Y. A study of the effect of a high-speed rail station on spatial variations in housing price based on the hedonic model. *Habitat Int.* **2015**, *49*, 333–339. [CrossRef]

19. Dai, X.; Bai, X.; Xu, M. The influence of Beijing rail transfer stations on surrounding housing prices. *Habitat Int.* **2016**, *55*, 79–88. [CrossRef]

20. Jang, M.; Kang, C.D. Retail accessibility and proximity effects on housing prices in Seoul, Korea: A retail type and housing submarket approach. *Habitat Int.* **2015**, *49*, 516–528. [CrossRef]

21. Yuan, F.; Wu, J.; Wei, Y.D.; Wang, L. Policy change, amenity, and spatiotemporal dynamics of housing prices in Nanjing, China. *Land Use Policy* **2018**, *75*, 225–236. [CrossRef]

22. Kyriakidis, D. Housing submarkets and future demographic developments: A review on the literature. In *Handbook of Research on Policies and Practices for Sustainable Economic Growth and Regional Development*; IGI Global: Hershey, PA, USA, 2017.

23. Rae, A.; Sener, E. How website users segment a city: The geography of housing search in London. *Cities* **2016**, *52*, 140–147. [CrossRef]

24. Costello, G.; Leishman, C.; Rowley, S.; Watkins, C. Drivers of spatial change in urban housing submarkets. *Geogr. J.* **2019**, *1*, 1–10. [CrossRef]

25. Megbolugbe, I.F.; Hoek-Smit, M.C.; Linneman, P.D. Understanding neighbourhood dynamics: A review of the contributions of William, G. Grigsby. *Urban Stud.* **1996**, *33*, 1779–1795. [CrossRef]

26. Leishman, C. Spatial change and the structure of urban housing sub-markets. *Hous. Stud.* **2009**, *24*, 563–585. [CrossRef]

27. Maclennan, D.; Tu, Y. Economic perspectives on the structure of local housing systems. *Hous. Stud.* **1996**, *1*, 387–406. [CrossRef]

28. Watkins, C.A. The definition and identification of housing submarkets. *Environ. Plan. A* **2001**, *33*, 2235–2253. [CrossRef]

29. Gabrielli, L.; Giuffrida, S.; Trovato, M.R. Gaps and overlaps of urban housing sub-market: Hard clustering and fuzzy clustering approaches. *Green Energy Technol.* **2017**, 203–219.

30. Napoli, G.; Giuffrida, S.; Trovato, M.R.; Valenti, A. Cap rate as the interpretative variable of the urban real estate capital asset: A comparison of different sub-market definitions in Palermo, Italy. *Buildings* **2017**, *7*, 80. [CrossRef]

31. Can, A. The measurement of neighborhood dynamics in urban house prices. *Econ. Geogr.* **1990**, *66*, 254–272. [CrossRef]

32. Ismail, S. Spatial Autocorrelation and Real Estate Studies: A Literature Review. *Malays. J. Real Estate* **2006**, *1*, 1–13.

33. Bourassa, S.C.; Cantoni, E.; Hoesli, M. Predicting House Prices with Spatial Dependence: A Comparison of Alternative Methods. *J. Real Estate Res.* **2010**, *32*, 139–159.

34. LeSage, J.; Pace, R.K. *Introduction to Spatial Econometrics*; CRC: Boca Raton, FL, USA, 2009.

35. Morano, P.; Locurcio, M.; Tajani, F.; Guarini, M.R. Fuzzy logic and coherence control in multi-criteria evaluation of urban redevelopment projects. *Int. J. Bus. Intell. Data Min.* **2015**, *10*, 73–93. [CrossRef]

36. Gillen, K.; Thibodeau, T.; Wachter, S. Anisotropic Autocorrelation in House Prices. *J. Real Estate Financ. Econ.* **2001**, *23*, 5–30. [CrossRef]

37. Cartern, C.C.; Haloupek, W.J. Spatial Autocorrelation in a Retail Context. *Int. Real Estate Rev.* **2000**, *3*, 34–48.

38. Tajani, F.; Morano, P. Concession and lease or sale? A model for the enhancement of public properties in disuse or underutilized. *WSEAS Trans. Bus. Econ.* **2014**, *17*, 18.

39. Basu, S.; Thibodeau, T.G. Analysis of Spatial Autocorrelation in House Prices. *J. Real Estate Financ. Econ.* **1998**, *17*, 61–85. [CrossRef]

40. Anselin, L.; Bera, A. Spatial dependence in linear regression models with an introduction to spatial econometrics. In *Handbook of Applied Economic Statistics*; CRC: Boca Raton, FL, USA, 1998.

41. Anselin, L. Lagrange Multiplier Test Diagnostics for Spatial Dependence and Spatial Heterogeneity. *Geogr. Anal.* **1988**, *20*, 1–17. [CrossRef]

42. Cressie, N. Statistics for Spatial Data. *Terra Nova* **1992**, *4*, 613–617. [CrossRef]

43. Tu, Y.; Sun, H.; Yu, S.M. Spatial autocorrelations and urban housing market segmentation. *J. Real Estate Financ. Econ.* **2007**, *34*, 385–406. [CrossRef]

44. Anselin, L. Local Indicators of Spatial Association-LISA. *Geogr. Anal.* **1995**, *27*, 93–116. [CrossRef]

45. Fernández-Delgado, M.; Sirsat, M.S.; Cernadas, E.; Alawadi, S.; Barro, S.; Febrero-Bande, M. An extensive experimental survey of regression methods. *Neural Netw.* **2019**, *111*, 11–34. [CrossRef]

46. Wilhelmsson, M. Spatial models in real estate economics. *Hous. Theory Soc.* **2002**, *19*, 92–101. [CrossRef]

47. Cajias, M.; Ertl, S. Spatial effects and non-linearity in hedonic modeling. *J. Prop. Invest. Financ.* **2018**, *36*, 32–49. [CrossRef]

48. Bourassa, S.C.; Cantoni, E.; Hoesli, M. Spatial dependence, housing submarkets, and house price prediction. *J. Real Estate Financ. Econ.* **2007**, *35*, 143–160. [CrossRef]

49. Armaş, I.; Gavriş, A. Social vulnerability assessment using spatial multi-criteria analysis (SEVI model) and the Social Vulnerability Index (SoVI model)—A case study for Bucharest, Romania. *Nat. Hazards Earth Syst. Sci.* **2013**, *13*, 1481–1499. [CrossRef]

50. Scardaccione, G.; Scorza, F.; las Casas, G.; Murgante, B. Spatial autocorrelation analysis for the evaluation of migration flows: The Italian case. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Proceedings of the International Conference on Computational Science and Its Applications, Fukuoka, Japan, 23–26 March 2010*; Springer: Berlin/Heidelberg, Germany, 2010.

51. Barreca, A.; Curto, R.; Rolando, D. Location and real estate values: A study of the segmentation of the Microzones of Turin. *Territ. Ital.* **2017**, *2*, 49–71.

52. Politecnico di Torino. TREMO Turin Real Estate Market Observatory. 2000. Available online: http://www.oict.polito.it/en/ (accessed on 1 October 2019).

53. Curto, R.; Fregonara, E. Un sistema informativo territoriale per l'osservazione del mercato immobiliare a supporto dei catasti urbani e della gestione del territorio (24–42). *Aestimum* **2002**, *8*, 42–60.

54. Curto, R.; Fregonara, E.; Semeraro, P. Asking prices vs. Market Prices: An Empirical Analysis. *Territ. Ital.* **2012**, *1*, 53–72.

55. Curto, R.; Fregonara, E.; Semeraro, P. Listing behaviour in the Italian real estate market. *Int. J. Hous. Mark. Anal.* **2015**, *8*, 97–117. [CrossRef]

56. Cutter, S.L.; Boruff, B.J.; Shirley, W.L. Social Vulnerability. *Soc. Sci. Q.* **2003**, *84*, 242–261. [CrossRef]

57. Schmidtlein, M.C.; Deutsch, R.C.; Piegorsch, W.W.; Cutter, S.L. A sensitivity analysis of the social vulnerability index. *Risk Anal.* **2008**, *28*, 1099–1114. [CrossRef]

58. Goodchild, M.F. *Spatial Autocorrelation. Concepts and Techniques in Modern Geography*; Geo Books: Norwich, UK, 1986; Volume 47, p. 56.

59. Anselin, L. GIS Research Infrastructure for Spatial Analysis of Real Estate Markets. *J. Hous. Res.* **1998**, *9*, 113–133.

60. Anselin, L. Global spatial autocorrelation. *Spat. Anal. Course* **2008**, 255–278.

61. Moran, P.A. Notes on continuous stochastic phenomena. *Biometrika* **1950**, *37*, 17–23. [CrossRef]

62. Cliff, A.D.; Ord, J.K. Spatial and temporal analysis: Autocorrelation in space and time. *Quant. Geogr. Br. View* **1981**, *5*, 1–23.

63. Haining, R.; Wise, S.; Ma, J. Exploratory spatial data analysis in a geographic information system environment. *J. R. Stat. Soc. Ser. D Stat.* **1998**, *47*, 457–469. [CrossRef]

64. Breusch, T.S.; Pagan, A.R. A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica* **1979**, 1287–1294. [CrossRef]

65. Anselin, L.; Syabri, I.; Kho, Y. GeoDa: An introduction to spatial data analysis. *Geogr. Anal.* **2006**, *38*, 5–22. [CrossRef]

66. Rosen, S. Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *J. Political Econ.* **1974**, *82*, 34–55. [CrossRef]

67. Lancaster, K.J. A New Approach to Consumer Theory. *J. Political Econ.* **1966**, *74*, 132–157. [CrossRef]

68. Haupt, H.; Schnurbus, J.; Tschernig, R. On nonparametric estimation of a hedonic price function. *J. Appl. Econom.* **2010**, *22*, 695–699. [CrossRef]

69. Morano, P.; Tajani, F. Estimative analysis of a segment of the bare ownership market of residential property. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Proceedings of the International Conference on Computational Science and Its Applications, Ho Chi Minh, Vietnam, 24–27 June 2013*; Springer: Berlin/Heidelberg, Germany, 2013.

70. Seymour, L. Spatial Data Analysis: Theory and Practice. *J. Am. Stat. Assoc.* **2005**. [CrossRef]

71. Anselin, L.; Griffith, D.A. Do spatial effect really matter in regression analysis? *Pap. Reg. Sci.* **1988**, *65*, 11–34. [CrossRef]

72. Kelejian, H.H.; Prucha, I.R. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *J. Econom.* **2010**, *157*, 53–67. [CrossRef]

73. Nicodemus, A.G. Fuzzy vibrancy: Creative placemaking as ascendant US cultural policy. *Cult. Trends* **2013**, *22*, 213–222. [CrossRef]

74. Armaş, I.; Gavriş, A. Census-based Social Vulnerability Assessment for Bucharest. *Procedia Environ. Sci.* **2016**, *32*, 138–146. [CrossRef]

75. Anselin, L.; Bera, A.K.; Florax, R.; Yoon, M.J. Simple diagnostic tests for spatial dependence. *Reg. Sci. Urban Econ.* **1996**, *26*, 77–104. [CrossRef]

76. Catania, L.; Billé, A.G. Dynamic spatial autoregressive models with autoregressive and heteroskedastic disturbances. *J. Appl. Econom.* **2017**, *32*, 1178–1196. [CrossRef]

77. Anselin, L.; Getis, A. Spatial statistical analysis and geographic information systems. *Adv. Spat. Sci.* **2010**, *26*, 19–33.