

Forms and Functions of the Real Estate Market of Palermo (Italy). Science and Knowledge in the Cluster Analysis Approach

Grazia Napoli, Salvatore Giuffrida and Alberto Valenti

Abstract The analysis of the housing market of a city requires suitable approaches and tools, such as data mining models, to represent its complexity which derives on many elements, e.g. the type of capital asset-house is a common good and an investment good as well, the heterogeneity of the urban areas—each of them has own historical and representative values and different urban functions—and the variability of building quality. The housing market of the most densely populated area of Palermo (Italy), corresponding to ten districts, is analyzed to verify the degree of its inner homogeneity and the relations between the quality of the characteristics and the price of the properties. Five hundred sets of housing data have been collected and elaborated by cluster analysis with the aim of describing the structure of the housing market in each district and developing operational tools for the implementation of urban policies and public-private investments.

Keywords Housing market · Data mining · Cluster analysis · *K-means* method

1 Introduction

This paper constitutes a part of a larger research on the urban real-estate market concerned with the analysis of its general structure and the monitoring of its evolution during the last few years, when the economic and property crisis, along with the change in the fiscal system, have deeply impacted on the readiness to invest, the

G. Napoli (✉)

Department of Architecture, University of Palermo, Palermo, Italy
e-mail: grazia.napoli@unipa.it

S. Giuffrida

Department of Civil Engineering and Architecture, University of Catania, Catania, Italy
e-mail: sgiuffrida@dica.unict.it

A. Valenti

Professional, Palermo, Italy
e-mail: albvnt79@gmail.com

© Springer International Publishing AG 2017

S. Stanghellini et al. (eds.), *Appraisal: From Theory to Practice*,
Green Energy and Technology, DOI 10.1007/978-3-319-49676-4_14

competition between different types of assets, and the very idea of urban real-estate capital as well (Breuer and Nadler 2012; Gabrielli et al. 2015; Giuffrida et al. 2015). The analysis applied to the housing market in Palermo (Italy) can provide an informative base for the public investment decision processes, and the implementation of planning policy and public/private negotiations, especially in a period in which the map of the urban values is going to be reconfigured because of significant modifications of the public transport system (streetcars and subway).

Assuming the district as the minimum spatial unit, five hundreds sets of data of houses were collected in ten urban districts corresponding to a wide, densely populated, and complex area of the city. With the aim of managing and organizing the collected data, the study applies the cluster analysis approach to provide different hypotheses of articulation of the real estate market into submarkets expressing the characteristics of the properties. In such a way, the study intends to express the fluid and mutable relationship that relates the objects (the properties having their own characteristics) to the hypothetical model that is meant to represent them. The concept of homogeneity is mostly related to the topographical demarcation of the district and the peculiar characteristics of the properties that are afferent to the same submarket. This study has continued to review this concept as necessary to be employed as a unique interpretative scheme for a systematic reading of the real estate phenomenon, by the means of data-mining models or big-data management (Case et al. 2004; Fik et al. 2003).

The cluster techniques, developed in other scientific sectors, have been effectively used for the implementation of the mass appraisal in the fiscal equalization (Nesticò et al. 2014) and the fair-land planning (Giuffrida et al. 2014), where the direct or “phenomenological” approach (which is generally applied for a case-by-case valuation) must be modified to represent the structural tendencies of the market and how the real estate values react to specific or areal transformations within the city (Chan et al. 2012; Gabrielli 2013; Hepşen and Vatanserver 2012).

The study also provides the occasion for a few methodological remarks on the representation of the real estate market, at the property and urban level as well. It also shows some of the major difficulties in the development of a standardized informational support that makes possible systematic analysis of the observations and the comparisons between different urban districts.

2 The Real-Estate Market Survey

This study analyzes the area of Palermo corresponding to ten districts having varied historical, representative, and functional qualities because of the time of their establishment and the most recent urban transformations. The boundaries of the area are the Mediterranean coast on the East, Regione Siciliana Street on the South and West, and Mount Pellegrino on the North. This area covers about 48 km² (30% of the municipal land) where 55% of the population is located (about 370,000 people). The districts are the followings: Q1 Settecannoli-Brancaccio is a

working-class suburb; Q2 Oreto-Stazione, Q5 Montegrappa-S. Rosalia, Q6 Cuba-Calatafimi, Q7 Zisa-Noce, and Q10 Malaspina-Palagonia are low and medium income districts located near the city center; Q3 Tribunali-Castellamare is a part of the historic center where there is a mix of social classes; Q8 Politeama, and Q9 Libertà constitute the city “center”, that was built between the end of the 19th and the beginning of the 20th century, where high-income households live; and Q11 Resuttana-S. Lorenzo is a middle-class suburb (Fig. 1).

The market survey collected 500 residential properties for sale located in the abovementioned districts in 2014 (Fig. 1). The data sample describes the houses by four types of characteristics (k_e location, k_i intrinsic, k_t technological, and k_a architectural) (Forte 1968) that are organized in 28 quantitative and qualitative attributes as shown in Table 1. The data sample also contains the asking prices, and the prices per square meters, as well as per room. Each attribute is expressed in a standard scale (from 1 to 5) so that the scores are aggregated at the characteristic level and, afterward, the overall quality k^* is achieved by calculating the weighted average score of each house.

Due to the generalized inconsistency of the coefficients of the four regressors calculated by the multiple linear regression for each district, the weights for the score aggregation have been calculated on the basis of the mean of the most significant coefficients in each district, having excluded the negative or the highest ones. Moreover, the weights λ_j of each district have been iteratively varied within the ranges shown below to obtain the maximum value of R^2 . Assuming $\sum_j \lambda_j = 1; j = e, i, t, a$, the ranges are: $0, 10 \leq \lambda_e \leq 0, 30$; $0, 10 \leq \lambda_i \leq 0, 30$; $0, 30 \leq \lambda_t \leq 0, 50$; $0, 10 \leq \lambda_a \leq 0, 30$.

A sample of the database for the Q3 district is shown in Table 2.

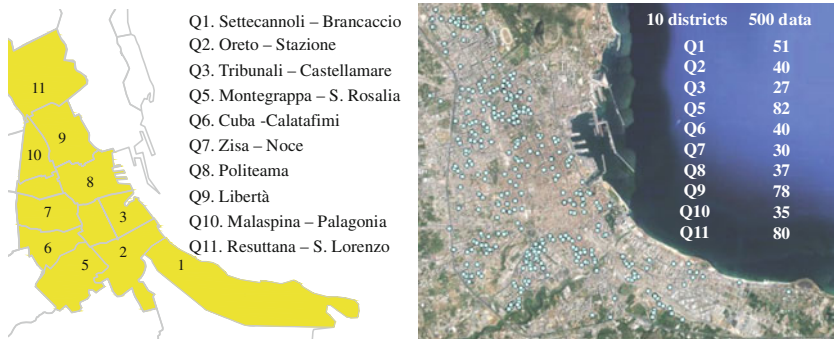


Fig. 1 Localization of the real estate data in the analyzed districts in Palermo

Table 1 Characteristics and attributes of the houses

| Characteristics | | Attributes |
|---------------------|---------------|--|
| k_e Location | Urban contest | k_{e1} Centrality k_{e2} Functional complexity k_{e3} Social complexity k_{e4} Facilities k_{e5} External accessibility with public transport k_{e6} External accessibility with private transport k_{e7} Internal accessibility |
| | Micro-context | k_{e8} Functionality k_{e9} Symbolic value k_{e10} Compositional quality k_{e11} Social mixité |
| k_i Intrinsic | | k_{i1} Adjacency k_{i2} Panorama k_{i3} Brightness k_{i4} Exposure k_{i5} Security |
| k_t Technological | | k_{t1} Age k_{t2} Building maintenance status k_{t3} Building structure k_{t4} Building finishes k_{t5} Property maintenance |
| k_a Architectural | Building | k_{a1} Architectural type k_{a2} Super elevations k_{a3} Unit size k_{a4} Terraces |
| | Property | k_{a5} Dimensions and functional layout k_{a6} Accessories k_{a7} Property finishes |

Table 2 Database for the Q3 district

| Id | Address | Rooms | m ² | ke | ki | kt | ka | k* | €/rooms | €/m ² |
|-----|---------------------|-------|----------------|-----|-----|-----|-----|-----|---------|------------------|
| 92 | Piazza Aragona | 5.25 | 150 | 2.0 | 2.4 | 1.8 | 0.5 | 1.7 | 34,286 | 1200 |
| 93 | P.tta Ch. Cocchieri | 3.75 | 70 | 2.0 | 3.0 | 0.5 | 0.4 | 1.3 | 23,733 | 1271 |
| 94 | Via Parlamento | 4.75 | 110 | 2.0 | 3.0 | 2.5 | 0.5 | 2.1 | 25,263 | 1093 |
| 95 | Discesa dei Giudici | 7.25 | 157 | 2.0 | 2.4 | 0.5 | 0.6 | 1.2 | 31,034 | 1436 |
| 96 | Via Giardinaccio | 2.75 | 54 | 1.9 | 1.6 | 4.3 | 0.7 | 2.6 | 50,909 | 2593 |
| 97 | Via dei Cassari | 4.75 | 110 | 2.0 | 2.8 | 3.1 | 0.5 | 2.3 | 33,684 | 1461 |
| 98 | Via Immacolatella | 4.75 | 89 | 1.9 | 1.6 | 1.2 | 0.4 | 1.2 | 16,421 | 874 |
| 99 | Via N. Cervello | 3.75 | 86 | 1.9 | 1.0 | 2.5 | 0.6 | 1.7 | 23,733 | 1035 |
| 100 | Via Vetriera | 3.75 | 78 | 1.9 | 3.0 | 3.8 | 0.6 | 2.6 | 48,000 | 2300 |
| 101 | P.tta S. Sofia | 2.25 | 60 | 1.9 | 4.4 | 3.4 | 0.7 | 2.7 | 51,111 | 1931 |
| 102 | Via S. Sebastiano | 4.75 | 84 | 2.0 | 3.8 | 2.1 | 0.4 | 2.1 | 31,579 | 1777 |
| 103 | P.tta Ch. Cocchieri | 5.75 | 102 | 2.0 | 3.0 | 1.9 | 0.4 | 1.8 | 26,087 | 1467 |

(continued)

Table 2 (continued)

| Id | Address | Rooms | m ² | ke | ki | kt | ka | k* | €/rooms | €/m ² |
|-----|---------------------|-------|----------------|-----|-----|-----|-----|-----|---------|------------------|
| 104 | Via V. Emanuele | 2.25 | 50 | 2.2 | 2.2 | 4.2 | 0.7 | 2.7 | 73,333 | 3325 |
| 105 | Spasimo | 2.75 | 62 | 2.0 | 2.2 | 3.2 | 0.5 | 2.2 | 21,818 | 969 |
| 106 | Discesa dei Giudici | 7.25 | 235 | 2.0 | 5.0 | 1.8 | 0.7 | 2.2 | 31,034 | 958 |
| 107 | Paternostro | 1.75 | 46 | 2.0 | 3.0 | 1.4 | 0.7 | 1.7 | 34,286 | 1310 |
| 108 | Guascone | 5.5 | 146 | 1.9 | 3.6 | 1.2 | 0.5 | 1.7 | 14,545 | 548 |
| 109 | Via T. Mosche | 3 | 75 | 1.9 | 1.0 | 4.1 | 0.7 | 2.4 | 29,667 | 1183 |
| 110 | Via dei Frangiai | 2.75 | 75 | 1.9 | 5.0 | 3.2 | 0.7 | 2.8 | 30,909 | 1137 |
| 111 | Via della Vetriera | 2.25 | 51 | 1.9 | 3.0 | 4.2 | 0.7 | 2.8 | 53,333 | 2365 |
| 112 | Piazza Marina | 6.25 | 153 | 2.2 | 5.0 | 1.1 | 0.4 | 2.0 | 44,800 | 1827 |
| 113 | Via Credenzieri | 3 | 117 | 2.0 | 2.2 | 3.8 | 0.7 | 2.5 | 70,000 | 1803 |
| 114 | Via Garibaldi | 4.25 | 88 | 2.0 | 2.4 | 1.2 | 0.5 | 1.5 | 20,000 | 966 |
| 115 | Piazza Aragona | 3.25 | 125 | 2.0 | 3.6 | 1.8 | 0.5 | 1.9 | 46,154 | 1202 |
| 116 | Via Napoli | 6 | 172 | 2.0 | 3.6 | 3.9 | 0.5 | 2.8 | 39,167 | 1364 |
| 117 | Via Bara Olivella | 3.25 | 103 | 2.0 | 4.4 | 3.0 | 0.5 | 2.6 | 50,769 | 1603 |
| 118 | Piazza Marina | 6.5 | 135 | 2.2 | 3.4 | 1.6 | 0.6 | 1.9 | 68,462 | 3296 |

3 Clustering Methodology

Especially when the real estate is widely heterogeneous, data-mining procedures may be applied to achieve a consistent articulation of the real estate market in submarkets that represent the similarities between objects described by a standardized set of shared characteristics. The cluster approach can be classified into these types: hierarchical, non-hierarchical (partitions), grid-based, and model-based.

The *k-mean* algorithm, which belongs to a non-hierarchical method (Jardine and Sibson 1968), has been applied to the dataset formerly described. The output of this algorithm is the optimal partition of elements that maximizes a certain objective function, and it is based on the assumption of distributing the elements of a sample over a predetermined number of groups (King 2014; Everitt et al. 2011; Kaufman and Rousseeuw 1990). The number of possible partitions p ($p = 2^{(n-1)} - 1$) can be reduced through the initial choice of the number of groups of the partition and, consequently, the optimal partition can be constructed among those partitions having the chosen number of groups, using a criterion depending on the algorithm applied.

The *k-means* algorithm forms k groups using certain values as initial centroids and placing the elements into groups on the basis of the maximum proximity to the centroids (proximity is measured using the Euclidean metric). Once the first partition has been computed, the new centers are recalculated; the previous routine is modified in the subsequent routine, until convergence is obtained (the condition is that each element is assigned to the same group as in the previous partition). When

Table 3 Steps of the *k-means* algorithm

| Steps | Description |
|-------|--|
| 0 | Initialization <i>k</i> “temporary centroids” are chosen (the choice is random or guided by an empirical criterion) |
| 1 | Partition and determination of the new centers Each element <i>n</i> is assigned to the center that minimizes its Euclidean distance from it, and <i>k</i> groups of elements are obtained. After each assignment of <i>n</i> new element to a group, the centroid of the group is calculated |
| 2 | Identification of the new partition The process of allocating elements to groups on the basis of minimum distance from the centers is repeated |
| 3 | Verification of the partition If the partition constructed in Step 1 is the first partition or if it is not equal to the previous partition, return to Step 1, otherwise proceed to Step 4 |
| 4 | Stop the procedure The optimal partition has been obtained |

this condition has been verified, the optimal partition will have been thus obtained (Steinley 2003, 2006) (Table 3).

Some problematic aspects of the iterative *k-means* procedure mainly regard the choice of both the initial centroids and the number of groups *G*. The choice of the initial centroids is the starting point from which the search for the final partition begins. If there are no specific indications regarding them, an internal algorithm of the software (IBM SPSS) will elect the centroids between the elements of the sample, so that they are well spaced. Alternatively, the analysis can be performed many times, and the final partition will be the one that is more consistent in respect to the information in the dataset resulting from cognitive domains. Regarding the number of groups *g*, if it is not available a priori on the basis of the dataset, then the procedure can be applied several times by varying *g* ($g = 2, 3, \dots$) and choosing the value of *g* according to the *CH Calinski-Harabasz index* (Milligan and Cooper 1985; Yanchi et al. 2010). The *CH index* is calculated in the following way:

$$CH(g) = \frac{B(g)/(g-1)}{W(g)/(n-g)} \quad B(g) = \sum_{i=1}^g d(\bar{x}_i, \bar{x}); \quad W(g) = \sum_{i=1}^g \sum_{j: x_j \in C_i} d(x_j, \bar{x}_i) \quad (1)$$

where: *B* is the external deviance (between the groups); *W* is the internal deviance (within the group); *g* is the number of groups; \bar{x}_i is the mean value of the observations belonging to the *i*-th cluster C_i ; \bar{x} is the mean value of the entire sample; x_j is the *j*-th observation; *d* is the Euclidean metric; and *n* is the number of observations. Obviously, the more this index increases, the more the validity of the partition improves, since it represents the ratio between the external variance and the internal variance of the partition.

4 Application of Cluster Analysis

The cluster analysis (*k-mean* algorithm) is applied to the data sample by deciding in advance that the numbers of the clusters are equal to 3, 4 and 5—because of the limited variability of the overall quality in each district—and leaving the software to make the choice of the initial centroids.

Figure 2 shows the resulting values of the *CH index* and the number of clusters (best partitions) for which the *CH index* is maximized for each district:

- 3 clusters for the Q2, Q3, Q5, Q8, and Q9 districts;
- 4 clusters for the Q7 district;
- 5 clusters for the Q1, Q6, Q10, and Q11 districts.

In general terms, the resulting clusters are sufficiently representative of the local housing market: the suburbs, such as Q1 and Q11 districts, have a high degree of inner heterogeneity caused by various land uses (residential, industrial, and a shopping center) and by various states of maintenance of the buildings, and this complexity can be better expressed through numerous groups of properties—in this case, 5 clusters. —The central districts, such as Q3, Q8, and Q9, are instead quite homogeneous because they originate from the same period of the urban fabric and the analogous typologies of buildings, and they may be described through 3 clusters only.

However, by observing the relations between overall quality and prices in the scatter graphs (Fig. 2), we note that significant differences between the districts having the same number of clusters may occur. By comparing the Q1 and Q11 districts, for example, it can be noted that the price elasticity with respect to the overall quality is very low in the first district, whereas it is high in the second one. The low price elasticity may be explained through the fact that Q1 is a working class suburb where the lack of public facilities stops any price increase, even if the intrinsic and technological characteristics have good quality. Otherwise, by comparing the districts with three clusters, the data points in the Q2 and Q8 districts are quite close to the trend line, whereas, in the Q9 district, the data points are much more spread, so that the market prices differ greatly in correspondence to the same overall quality.

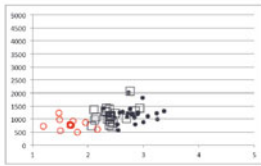
If the partition of the Q1 district (5 clusters) and the data set of the properties involved are examined with a greater detail (Fig. 3), it has been found that:

- the clusters 1 and 2 represent two groups of similar properties as all of them have the same value of k_e and k_a , whereas the first group has k_t higher and k_i lower than the second group's corresponding k (and vice versa);
- the clusters 3 and 5 are also comparable except for the k_i ;
- the properties in the cluster 4 have the lowest prices and the worst characteristics of the district.

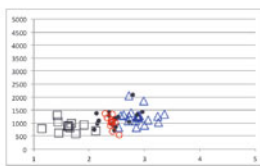
In the partition of the Q8 district into 3 clusters (Fig. 4; Table 4):

Q1 Settecannoli – Brancaccio

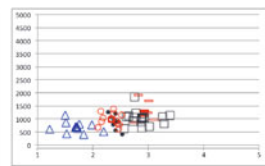
CH(3)= 660



CH(4)= 542

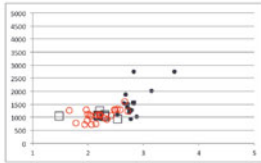


5 clusters CH(5)= 731

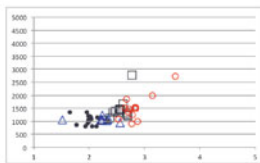


Q2 Oreto – Stazione

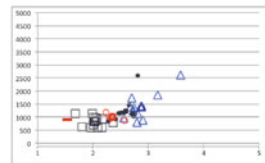
3 clusters CH(3)= 340



CH(4)= 278

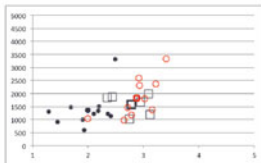


CH(5) ---

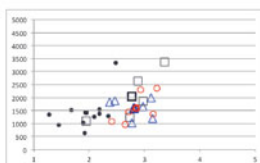


Q3 Tribunali – Castellammare

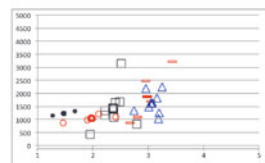
3 clusters CH(3)= 471



CH(4)= 287

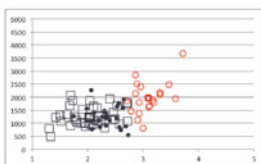


CH(5)= 214

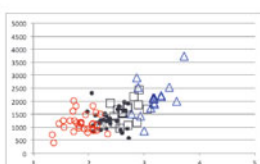


Q5 Montegrappa – S. Rosalia

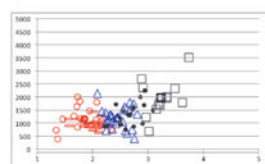
3 clusters CH(3)= 1074



CH(4)= 969

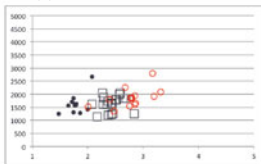


CH(5)= 882

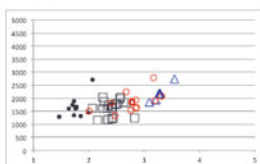


Q6 Cuba – Calatafimi

CH(3)= 429



CH(4)= 667



5 clusters CH(5)= 764

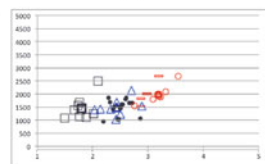


Fig. 2 Values of the CH index per district and relations between overall quality k^* (x-axis) and price €/m² (y-axis) of the houses

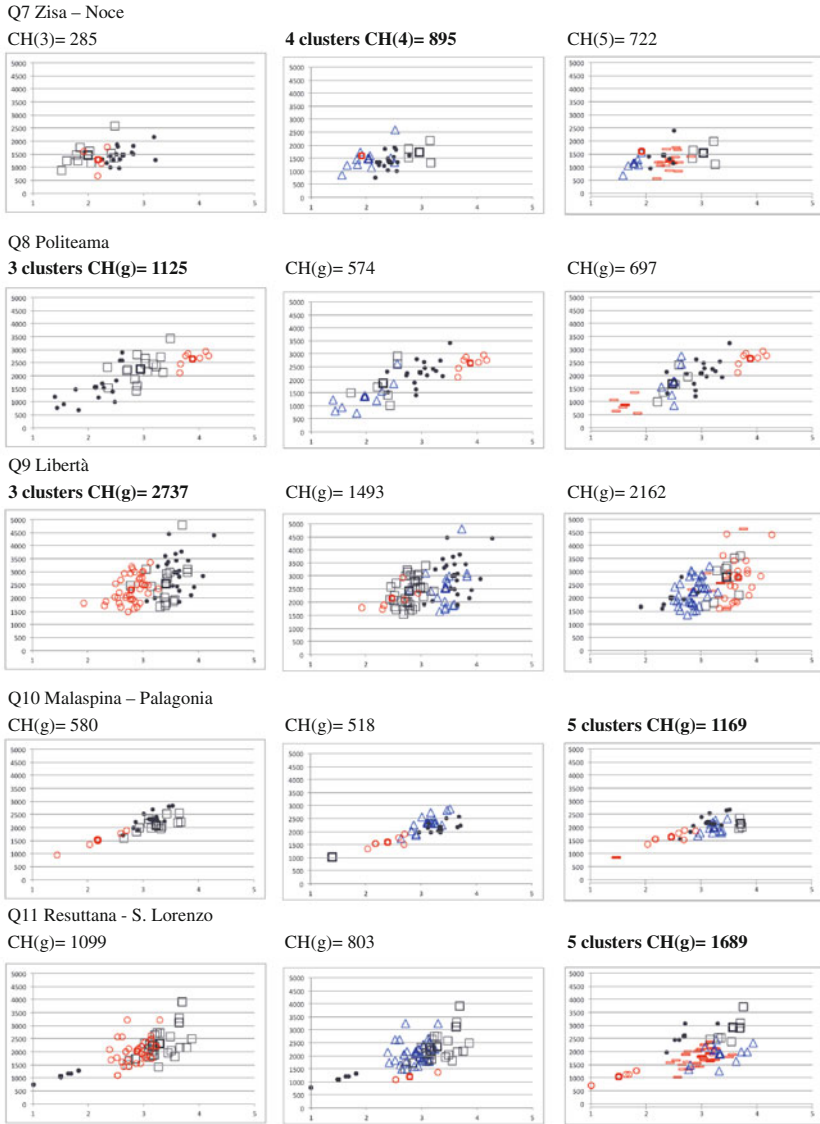


Fig. 2 (continued)

- the cluster 1 is very homogeneous, in fact the characteristics of all properties have the highest quality and the correspondent prices are higher than the mean price;
- in the cluster 2, the properties may have a low score for each k and especially k_e is very low because of their location in the blighted area of Borgo Vecchio, or they may have a high value of the location k_e and a low value of the others k ,

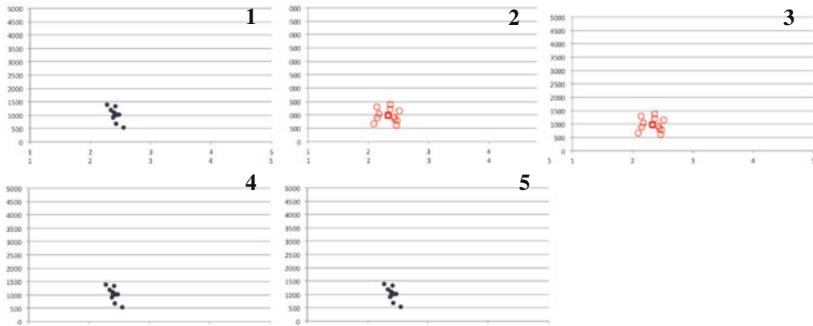


Fig. 3 The partition in the Q1 district

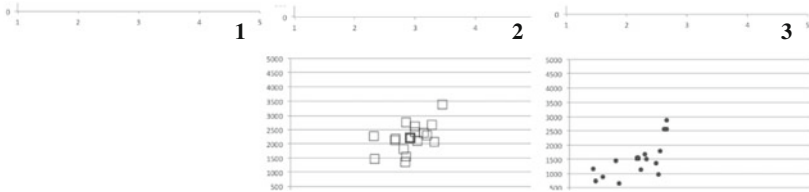


Fig. 4 The partition in the Q8 district

Table 4 Statistical results in the Q8 district

| | | Overall district Q8 | Clusters | | |
|----|---------------------------------|---------------------|----------|-------|-------|
| | | | 1 | 2 | 3 |
| ke | Mean | 2.95 | 3.87 | 2.69 | 2.78 |
| | Deviation (from mean of sample) | | 0.92 | -0.26 | -0.17 |
| | Deviation (%) | | 31 | -9 | -6 |
| | Standard deviation | 1.07 | 0.51 | 1.15 | 1.01 |
| | Standard deviation (%) | 36 | 13 | 43 | 37 |
| ki | Mean | 3.26 | 4.57 | 2.04 | 3.76 |
| | Deviation (from mean of sample) | | 1.31 | -1.22 | 0.50 |
| | Deviation (%) | | 44 | -37 | 15 |
| | Standard deviation | 1.17 | 0.57 | 0.63 | 0.57 |
| | Standard deviation (%) | 36 | 12 | 31 | 15 |
| kt | Mean | 2.81 | 4.06 | 1.96 | 3.02 |
| | Deviation (from mean of sample) | | 1.25 | -0.86 | 0.20 |
| | Deviation (%) | | 42 | -30 | 7 |
| | Standard deviation | 0.93 | 0.37 | 0.63 | 0.48 |
| | Standard deviation (%) | 33 | 9 | 32 | 16 |

(continued)

Table 4 (continued)

| | | Overall district Q8 | Clusters | | |
|--------------------------|---------------------------------|---------------------|----------|-------|-------|
| | | | 1 | 2 | 3 |
| ka | Mean | 2.31 | 2.84 | 2.02 | 2.34 |
| | Deviation (from mean of sample) | | 0.53 | -0.29 | 0.02 |
| | Deviation (%) | | 18 | -13 | 1 |
| | Standard deviation | 0.42 | 0.37 | 0.31 | 0.26 |
| | Standard deviation (%) | 18 | 13 | 15 | 11 |
| Price per m ² | Mean | 2,034 | 2,648 | 1,520 | 2,215 |
| | Deviation (from mean of sample) | | 615 | -514 | 181 |
| | Deviation (%) | | 30 | -25 | 9 |
| | Standard deviation | 706 | 285 | 709 | 521 |
| | Standard deviation (%) | 35 | 11 | 47 | 24 |

and, in this latter case, the prices rise because the market recognizes the location to produce a marginal price higher than the ones of the others features;

- the cluster 3 includes the properties with intermediate characteristics.

5 Conclusions

The results of the cluster analysis revealed that the housing market in each district has its own degree of complexity and peculiar relations between the market prices and the clusters representing the housing characteristics. The best number of clusters, chosen on the basis of *Calinski-Harabasz index*, expressed the inner variable heterogeneity of each district and represented the urban complexity.

The relationships between asking price and characteristics can significantly vary within the same cluster even when the characteristic quality is almost equivalent, and this fact is indicative of the typical information asymmetry and opacity of the real estate market and, moreover, of the current uncertainty and instability of the social and economic system, so that the owners of the real estate capitals express dissimilar expectations of the capital gains or losses (plus-minus valorization), translating them into different bid prices (Rizzo 1999).

Cluster analysis may be a useful tool to manage and analyze big data for describing, even in not exhaustive way, the structure of the real estate market, because this approach can select homogeneous groups of properties, reduce the degree of intrinsic complexity of the urban property data, and build a knowledge system to support the implementation of urban policy.

Acknowledgments The work was discussed collectively by the authors in all its parts. However Grazia Napoli has edited paragraphs 2, 4 and 5; Salvo Giuffrida, paragraphs 1 and 5; and Alberto Valenti, paragraph 3.

References

- Breuer W, Nadler C (2012) Real estate and real estate finance as a research field. an international overview. *J Bus Econ* 82:5–52
- Case B, Clapp J, Dubin R, Rodrigues M (2004) Modeling spatial and temporal house price patterns: a comparison of four models. *J Real Estate Finance Econ* 29:167–191
- Chan L, Ng HT, Ramchand R (2012) A cluster analysis approach to examining Singapore's property market. property markets and financial stability. *BIS Pap Bank Int Settlements* 64: 43–53
- Everitt B, Landau S, Morvene L, Stahl D (2011) *Cluster analysis*, 5th edn. Wiley, New York
- Fik TJ, Ling DC, Mulligan GF (2003) Modeling spatial variation in housing prices: a variable interaction approach. *Real Estate Econ* 31:623–646
- Forte C (1968) *Elementi di estimo urbano*. Napoli
- Gabrielli L (2013) Cluster analysis and Italian real estate market analysis during the downturn. In: 20th Annual European Real Estate Society Conference, ERES European Real Estate Society
- Gabrielli L, Giuffrida S, Trovato MR (2015) From surface to core: a multi-layer approach for the real estate market analysis of a Central Area in Catania. In Gervasi O et al. (eds) *Computational science and its applications—ICCSA 2015*, vol III. Springer, Heidelberg, pp 284–300
- Giuffrida S, Ferluga G, Valenti A (2014) Clustering analysis in a complex real estate market: the case of Ortigia (Italy). In Murgante B et al. (eds) *Computational science and its applications—ICCSA 2014*, vol III. Springer, Heidelberg
- Giuffrida S, Ferluga G, Valenti A (2015) Capitalization rates and “real estate semantic chains”. An application of clustering analysis. *Int J Bus Intell Data Min* 10(2):174–199
- Hepşen A, Vatansever M (2012) Using hierarchical clustering algorithms for turkish residential market. *Int J Econ Finan* 4:138–150 [online]. Doi: [10.5539/ijef.v4n1p138](https://doi.org/10.5539/ijef.v4n1p138)
- Jardine N, Sibson R (1968) The construction of hierarchic and non-hierarchic classifications. *Comput J* 1:177–184
- Kaufman L, Rousseeuw P (1990) *Finding Groups in Data: an introduction to clustering analysis*. Wiley, New York
- King R (2014) *Cluster analysis and data mining: an introduction*. Mercury Learning & Information, Herndon
- Milligan G, Cooper M (1985) An examination of procedures for determining the number of clusters in a data set. *Psychometrika* 50:159–179
- Nesticò A, De Mare G, Galante M (2014) The sustainable limit of the real estate tax: an urban-scale estimation model. In Murgante B et al. (eds) *Computational science and its applications—ICCSA 2014*, vol III. Springer, Heidelberg
- Rizzo F (1999) *Valore e valutazioni. La scienza dell'economia o l'economia della scienza*. FrancoAngeli, Milano
- Steinley D (2003) Local optima in k-means clustering: what you don't know may hurt you. *Psychol Methods* 8:294–304
- Steinley D (2006) Profiling local optima in K-means clustering: developing a diagnostic technique. *Psychol Methods* 11(2):178–192
- Yanchi L, Zhongmou L, Hui X, Xuedong G, Junjie W (2010) Understanding of internal clustering validation measures. In: *IEEE international conference on data mining*. <http://datamining.it.uts.edu.au/icdm10/index.php>. Accessed on 18 Sept 2015