Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

idealista18: A data package with real estate information in three major Spanish markets from the Idealista database*

David Rey Blanco[a], Pelayo González Arbues[a,1], Fernando López Hernández[b], Antonio Páez[c,*]

[a] *idealista, Plaza de las Cortes 5, 28014 Madrid, Spain*
[b] *Facultad de CC de la Empresa, C/ Real, 3. 30201 Cartagena, Murcia (Spain)*
[c] *School of Earth, Environment and Society, McMaster University, 1280 Main St W, Hamilton, Ontario L8S 4K1 Canada*

ARTICLE INFO

Keywords:

Property values
Spain
Spatial analysis
Machine learning
Hedonic price analysis

ABSTRACT

This dataset contains three items for each of the three major cities in Spain: Madrid, Barcelona and Valencia. First real estate listings published on idealista portal in 2018. All listings have been enriched with cadastral information (i.e. building year of construction, built quality materials grade) plus some geographical features such as distance to relevant city areas and the coordinates themselves. To comply with european personal protection laws, we have processeed some sensible variables yet preserving their spatial properties. The second and third items are a list of points of interest and the administrative boundaries for each municipality. This dataset is suitable to house market analysis, hedonic house price models and other spatial research related with real estate markets.

¡falta cita/referencia al artículo¿

Specifications Table

Every section of this table is mandatory. Please enter information in the right-hand column and remove all the instructions

*This is an example for title footnote coding.

*Corresponding author: Tel.: +1-905-525-9140 ext 26099
e-mail: paezha@mcmaster.ca (Antonio Páez)

[1] This is author footnote for second author.

http://dx.doi.org/10.1016/j.dib.xxxx.xx.xxx

Subject	Geography, Economics
Specific subject area	Spatial analysis, machine learning, hedonic price analysis
Type of data How data were acquired	Table [State how the data were acquired: E.g. Microscope, SEM, NMR, mass spectrometry, survey* etc. Instruments: E.g. hardware, software, program Make and model and of the instruments used: * if you conducted a survey you must submit a copy of the survey(s) used (either provide these as supplementary material file or provide a URL link to the survey in this section of the table). If the survey is not written in English, please provide an English-language translation.]
Data format	Spatially masked
Parameters for data collection	[Provide a brief description of which conditions were considered for data collection. Max 400 characters]
Description of data collection	[Provide a brief description of how these data were collected. Max 600 characters]
Data source location	Institution: Idealista City/Town/Region: Madrid Country: Spain Latitude and longitude (and GPS coordinates, if possible) for collected samples/data: If you are describing secondary data, you are required to provide a list of the primary data sources used in the section. Primary data sources:]
Data accessibility	Repository name: GitHub Direct URL to data: https://github.com/paezha/idealista18
Related research article	D. Rey Blanco, P. González Arbues, F. López Hernández, A. Páez, Using machine learning to identify spatial market segments: A reproducible study of major Spanish markets, Comput Environ Urban Syst. In Press.

Value of the Data

- A cleaned and enriched dataset consisting of real estate listings for three major cities in Spain, constructed to analyze the impacto of using machine learning models to identify spatial market segments to build house price hedonic models.
- The dataset can be used to extend the topic of market segments automatic or semi-automatic identification.
- Official boundaries combined with spatial patterns can be used to analyse the suitability of these boundaries for real estate value purposes.
- The dataset can be enlarged with complementary information to develop hedonic models.
- The data can be processed by quantitative analysis and statistical modeling to study the different factors that affect house prices in the three areas.
- Identification of spatial patterns in the real estate scope using the geo-referenced data points. For either value or urban patterns discovery.

Data Description

[Individually describe each data file (i.e. figure 1, figure 2, table 1, dataset, raw data, supplementary data, etc.) that are included in this article. Please make sure you refer to every data

file and provide a clear description for each - do not simply list them. No insight, interpretation, background or conclusions should be included in this section. Please include legends with any tables, figures or graphs.
Tip: do not forget to describe any supplementary data files.]

Experimental Design, Materials and Methods

[Offer a complete description of the experimental design and methods used to acquire these data. Please provide any programs or code files used for filtering and analyzing these data. It is very important that this section is as comprehensive as possible. If you are submitting via another Elsevier journal (a co-submission) you are encouraged to provide more detail than in your accompanying research article. There is no character limit for this section; however, no insight, interpretation, or background should be included in this section.
Tip: do not describe your data (figures, tables, etc.) in this section, do this in the Data Description section above.]

Ethics Statement

[Please refer to the journal's Guide for Authors for more information on the ethical requirements for publication in Data in Brief. In addition to these requirements:
If the work involved the use of human subjects: please include a statement here confirming that informed consent was obtained for experimentation with human subjects;
If the work involved animal experiments: please include a statement confirming that all experiments comply with the ARRIVE guidelines and were be carried out in accordance with the U.K. Animals (Scientific Procedures) Act, 1986 and associated guidelines, EU Directive 2010/63/EU for animal experiments, or the National Institutes of Health guide for the care and use of Laboratory animals (NIH Publications No. 8023, revised 1978)]

Acknowledgments

Acknowledgments should be inserted at the end of the paper, before the references, not as a footnote to the title. Use the unnumbered Acknowledgements Head style for the Acknowledgments heading.

Declaration of Competing Interest

[All authors are required to report the following information:

(1) All third-party financial support for the work this article;
(2) All financial relationships with any entity that could be viewed as relevant to data described in this manuscript;
(3) All sources of revenue with relevance to this work where payments have been made to authors, or their institutions on their behalf, within the 36 months prior to submission;
(4) Any other interactions with the sponsor, outside of the submitted work;
(5) Any relevant patents or copyrights (planned, pending or issued);
(6) Any other relationships or affiliations that may be perceived by readers to have influenced, or give the appearance of potentially influencing, what has been written in this article.

As a general guideline, it is usually better to disclose a relationship than not. This information will be acknowledged at publication in the manuscript. If there is no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper, please include this statement.]

The authors declare that they have no known competing financial interests or personal relationships which have, or could be perceived to have, influenced the work reported in this article.

[If there are financial interests/personal relationships which may be considered as potential competing interests, please declare them here.]

Note

Any instructions relevant to the `elsarticle.cls` are applicable here as well. See the online instruction available on:

`http://support.river-valley.com/wiki/index.php?title=Elsarticle.cls`

References

[References are limited (approx. 15) and excessive self-citation is not allowed. **If your data article is co-submitted via another Elsevier journal, please cite your associated research article here.**
Reference style: Text:?Indicate references by number(s) in square brackets in line with the text. The actual authors can be referred to, but the reference number(s) must always be given.?
Example: '..... as demonstrated [3,6]. Barnaby and Jones [8] obtained a different result ....'?
[Use `\cite` command to cite a reference list item in text.
These are examples for reference citations [1]. [2]. [3].]

References

[1] M. E. J. Newman, M. Girvan, Finding and evaluating community structure in networks, Phys. Rev. E. 69 (2004) 026113.
[2] C. Vehlow, T. Reinhardt, D. Weiskopf, Visualizing fuzzy overlapping communities in networks, IEEE Trans. Vis. Comput. Graph. 19 (2013) 2486–2495.
[3] E. Hullermeier, M. Rifqi, A fuzzy variant of the rand index for comparing clustering structures, in: in Proc. IFSA/EUSFLAT Conf., 2009, pp. 1294–1298.