Datenbeschreibung

RWI – Leibniz-Institut für Wirtschaftsforschung

FDZ Data Description: Real-Estate Data for Germany (RWI-GEO-RED v15) - Advertisements on the Internet Platform ImmobilienScout24 2007-06/2026

Felix Heuer

August 2026

Impressum

Herausgeber

RWI – Leibniz-Institut für Wirtschaftsforschung e.V.

Hohenzollernstraße 1–3 | 45128 Essen, Germany

Postfach 10 30 54 | 45030 Essen, Germany

Fon: +49 201–81 49-0 | E-Mail: rwi@rwi-essen.de

www.rwi-essen.de

Vorstand

Prof. Dr. Thomas K. Bauer (Vizepräsident)
Dr. Stefan Rumpf (Administrativer Vorstand)
Prof. Dr. Kerstin Schneider (Mitglied des erweiterten Vorstands)

© RWI 2026
Der Nachdruck, auch auszugsweise, ist nur mit Genehmigung des RWI gestattet.

RWI Datenbeschreibung

Schriftleitung: Prof. Dr. Thomas K. Bauer
Gestaltung: Magdalena Franke, Claudia Lohkamp

FDZ Data Description: Real-Estate Data for Germany
(RWI-GEO-RED v15) - Advertisements on the Internet Platform
ImmobilienScout24 2007-06/2026

August 2026

RWI: Felix Heuer

Leibniz-Institut für Wirtschaftsforschung

Leibniz Gemeinschaft

Das RWI wird vom Bund und vom Land Nordrhein-Westfalen gefördert.

Contents

List of Figures **4**

List of Tables **4**

Abstract **5**

1 Introduction and Short Description **6**
1.1 Introduction 6
1.2 Short Description 7
1.3 List of Variables 7
1.4 Data Access 10

2 Data Preparation **12**
2.1 Data Source 12
2.2 Georeferencing 12
2.3 Anonymization 13
2.4 Missings 13
2.5 Data problems 14

3 Variable Description **18**
3.1 Identifier 18
3.2 Time Period 19
3.3 Housing Unit Features 19
3.4 General Information of the Housing Unit 21
3.5 Area Information 23
3.6 Energy and Structure Information 23
3.7 Price Information 24
3.8 Regional Information 26
3.9 Meta-Information of the Advertisement 28
3.10 Generated Technical Variables 30

4 Changes Between Versions **31**

5 Further Material **32**
5.1 Value Labels 32
5.2 Summary Statistics 32

6 References **33**

List of Figures

List of Tables

1 Data overview . . . 7
2 List of Variables . . . 7
3 Missing Values . . . 13
4 Censoring thresholds . . . 14
5 Categorization of duplicate object identifiers . . . 16
6 Detailed variable descriptions of identifiers . . . 18
7 Detailed variable descriptions of variables describing time . . . 19
8 Detailed variable descriptions of variables describing housing unit features . . . 19
9 Detailed Variable Descriptions of general housing unit information . . . 21
10 Detailed Variable Descriptions of Area Information . . . 23
11 Detailed Variable Descriptions of Energy and Structure Information . . . 23
12 Detailed Variable Descriptions of Price Information . . . 25
13 Detailed Variable Descriptions of Regional Information . . . 26
14 Detailed Variable Descriptions of Variables Containing Meta-information of the Advertisement . . . 28
15 Detailed Variable Descriptions of Technical Variables . . . 30

Abstract

The FDZ Ruhr provides a dataset on private real estate advertisements in Germany from 2007 onwards. The data are based on information from the internet platform ImmobilienScout24 and covers both apartments and houses for sale and for rent. The data are available for scientific research only. The provided dataset covers detailed regional information and a rich set of housing characteristics. This data report gives a brief overview on the data as well as its limitations and specifics. The data report is addressed to (potential) users of the data as support for their data preparation.

This data report refers to RWI-GEO-RED V15 that covers data until June 2026.

1. Introduction and Short Description

1.1. Introduction

The Research Data Center (FDZ) Ruhr at the RWI provides a unique dataset on German real estate prices, obtained by ImmobilienScout24. The dataset entails information of real estate offerings on prices as well as on various observable characteristics that determine the value of a property. It is provided on a monthly basis. The present dataset (V15) covers January 2007 until June 2026. The data are organized into four separate datasets: houses for sale, houses for rent, flats for sale, and flats for rent.

ImmobilienScout24 is the largest internet platform on real estate in Germany. It gives real estate owners the opportunity to advertise their objects for a fee. The platform is open to private and commercial entities. Yet, the present dataset only includes residential real estate. It distinguishes between houses and apartments and features objects both for rent and for sale. ImmobilienScout24 has a self-reported market share of about 50% of all real estate objects offered for sale or rent in Germany (Georgi & Barkow, 2010).

To advertise an object, the owner has to fill out a questionnaire asking for different characteristics of the property. This means that the price at which the owner is willing to sell or rent out the object has to be understood as an offering price. The advertised price is not binding, i.e., the data do not comprise transaction prices. The price information is available for almost all advertisements. In addition to the price, advertisers are free to include further object-specific characteristics. This helps to present an object adequately, and ideally, it increases the chance of selling at a favorable price or renting out a flat respectively.

This unique dataset has many advantages compared to other house price data available for the German real estate market. First, due to the high market share of ImmobilienScout24, the dataset consists of a substantial number of observations. This allows for analyzing small-scale housing markets, such as the investigation of the effect of policy intervention on local house prices (e.g., Bauer et al., 2015) and house price responses to location-specific shocks (e.g., Bauer et al., 2017). Second, the FDZ Ruhr generates regional price indices for apartments and houses based on these data which are updated regularly (see RWI-GEO-REDX, DOI: https://doi.org/10.7807/immo:redx:suf:v17). Those price indices are provided for districts (Kreise) and municipalities (Gemeinden). See Heuer and Thiel (2026b) for more details. Both the raw data and the price indices are made available for scientific research.

Besides the described scientific use file and the on-site use, there exist two versions of campus files that are available for teaching purposes. For more details see (Heuer & Thiel, 2026a).

This data description is an updated version of Heuer and Thiel (2025). Therefore, there are several text parts that are the same.

1.2. Short Description

Table 1
Data overview

Data unit	Residential housing unit offered for rent or sale in Germany on Immobilienscout24.de
Period covered	January 2007 until June 2026
Time reference	Monthly
Regional structure	Geocoded addresses (on-site), as well as 1 sqm grid identifier and larger administrative areas (scientific use file)
Date of territorial status	End of 2015 and 2019 (for administrative boundaries)
Sample	Full population of all housing units offered on the platform
Update frequency	Continuous
Data access	Available for scientific research only
Anonymization	1 sqm grid identifier anonymized for scarcely populated cells in scientific use file
Current version	01/2007–06/2026 V15.0

1.3. List of Variables

The dataset contains the following variables where some variables are only relevant for certain types of real estate. Note that we print the availability of a given variable for the most recent data delivery. Please refer to Section 4 for changes between versions.

Table 2
List of Variables

Category	Variable name	Description	House rent	House sale	Flat rent	Flat sale
Identifier	obid	Housing unit identifier	1	1	1	1
	uniqueID_gen	Unique housing unit identifier (generated)	1	1	1	1
Time period	ajahr	Beginning of ad, year	1	1	1	1
	amonat	Beginning of ad, month	1	1	1	1
	ejahr	Ending of ad, year	1	1	1	1
	emonat	Ending of ad, month	1	1	1	1
Housing unit features	aufzug	Elevator in housing unit	0	0	1	1
	ausstattung	Facilities of housing unit	1	1	1	1
	badezimmer	Number of bathrooms	1	1	1	1
	balkon	Balcony at housing unit	0	0	1	1
	denkmalob-jekt	Protected historic building	0	1	0	1
	einbaukueche	Kitchenette in housing unit	1	0	1	1
	etage	Floor on which housing unit is lo-cated	0	0	1	1
	ferienhaus	Usable as holiday home	0	1	0	1
	freiab	Available from	1	1	1	1

	gaestewc	Guest toilet in housing unit	1	1	1	1
	garten	(Shared) garden available	0	0	1	1
	haustier_erlaubt	Pets allowed	1	0	1	0
	kategorie_Haus	House type	1	1	0	0
	kategorie_Wohnung	Flat type	0	0	1	1
	keller	Cellar in housing unit	1	1	1	1
	parkplatz	Garage/ parking space available	1	1	1	1
	zimmeranzahl	Number of rooms	1	1	1	1
	anzahletagen	Number of floors	1	1	1	1
	bauphase	Construction phase	0	1	0	0
	betreut	Assisted living for the elderly	0	0	0	0
	einliegerwohnung	Granny flat in housing unit	0	1	0	0
	foerderung	Public housing – certificate of eligibility is needed	0	0	1	0
	immobilientyp	Type of real estate	1	1	1	1
	kaufvermietet	Rented when sold	0	1	0	1
	mieteinnahmenpromonat	Rental income per month in EUR	0	1	0	1
	nebenraeume	Number of ancillary rooms	0	0	0	0
	rollstuhlgerecht	Accessible, no steps	1	1	1	1
	schlafzimmer	Number of bedrooms	1	1	1	1
	wohngeld	Common charge for community association in EUR/month	0	0	0	1
	grundstueckflaeche	Plot area	1	1	0	0
	nutzflaeche	Usable floor space	1	1	1	1
	wohnflaeche	Living area	1	1	1	1
Energy and structure information	baujahr	Year that housing unit was built	1	1	1	1
	energieausweistyp	Type of Energy Performance Certificates (EPCs)	1	1	1	1
	energieeffizienzklasse	Energy Efficiency Rating	1	1	1	1
	ev_kennwert	Energy consumption per year and square meter	1	1	1	1
	ev_wwenthalten	Warm water consumption included in energy consumption	1	1	1	1
	heizkosten	Heating costs	1	0	1	0
	heizungsart	Type of heating	1	1	1	1
	bef1-bef10	firing type	1	1	1	1

	letzte_mod-ernisierung	Year of last modernisation of housing unit	1	1	1	1
	objektzustand	Condition of housing unit	1	1	1	1
Price information	courtage	Brokerage at contract conclusion	1	1	1	1
	heizkosten_in_wm_ enthal-ten	Heating costs covered by inclusive rent	1	0	1	0
	kaufpreis	Purchasing price in EUR	0	1	0	1
	mietekalt	Exclusive rent in EUR	1	0	1	0
	mietekaution	Security deposit	1	1	1	1
	mietewarm	Inclusive rent in EUR	1	0	1	0
	nebenkosten	Utilities in EUR	1	0	1	0
	parkplatzpreis	Price of parking space in EUR	1	1	1	1
Regional information	blid	German state	1	1	1	1
	erg_amd	Local labour market (Kosfeld and Werner, 2012)	1	1	1	1
	ergg_1km	1-skm raster cell following INSPIRE	1	1	1	1
	gid2015	Municipality identifier (AGS, 2015)	1	1	1	1
	gid2019	Municipality identifier (AGS, 2019)	1	1	1	1
	kid2015	District identifier (AGS, 2015)	1	1	1	1
	kid2019	District identifier (AGS, 2019)	1	1	1	1
	plz	Address: postcode	1	1	1	1
	gid_updated	Municipality identifier updated by IS24 information	1	1	1	1
	kid_updated	District identifier updated by IS24 information	1	1	1	1

Meta-information of ad	click_customer	Number of clicks on customer profile	1	1	1	1
	click_schnellkontakte	Number of clicks on contact button	1	1	1	1
	click_url	Number of clicks on customer URL	1	1	1	1
	click_weitersagen	Number of clicks on share button	1	1	1	1
	hits	Number of hits of ad	1	1	1	1
	laufzeittage	Days of availability of ad	1	1	1	1
	liste_show	Number of hits through result list	1	1	1	1
	liste_match	Number of hits through search request	1	1	1	1
	click_schnellkontakte_gen	Number of clicks on contract button, adjusted by ad availability	1	1	1	1
	click_url_gen	Number of clicks on customer URL, adjusted by ad availability	1	1	1	1
	click_weitersagen_gen	Number of clicks on share button, adjusted by ad availability	1	1	1	1
	hits_gen	Number of hits of ad, adjusted by ad availability	1	1	1	1
	liste_show_gen	Number of hits through result list, adjusted by ad availability	1	1	1	1
	liste_match_gen	Number of hits through search request, adjusted by ad availability	1	1	1	1
	lieferung	Date of data retrieval	1	1	1	1
	anbieter	Type of provider (seller/landlord)	1	1	1	1
Generated technical variables	dupID_gen	Classification of housing unit identifiers used more than once	1	1	1	1
	duplicateid	obid of possible duplicate	1	1	1	1
	spell	Spell counter within housing unit identifier	1	1	1	1
	dataset_version	Version of dataset	1	1	1	1

1.4. Data Access

The data are available to researchers for non-commercial use. There are two versions of the RWI-GEO-RED data. First, the Scientific Use Files (SUF) cover all information except the exact geo-coordinates. Further, some grid cells are anonymized due to the anonymization rules in Section 2.3. Second, the full datasets are available in the Data Secure Room of the FDZ Ruhr in Essen (on-site access). The data can be obtained as a Stata® datasets (.dta), .csv files, and .parquet files. Data access to both versions requires a signed data use agreement. Both versions are restricted to non-commercial research and only researchers of scientific institutions are eligible to apply for data access. The SUF may be used at the workplace of the users.

Data access is provided by the Research Data Centre Ruhr at the RWI – Leibniz-Institute for Economic Research (FDZ Ruhr). The data can be accessed at https://www.rwi-essen.de/en/research-advice/further/research-data-center-ruhr-fdz/data-access. The application form includes a brief

description and title of the project, potential cooperation, information on the applying department, expected duration of data usage as well as further participants in the project.

Data users shall cite the datasets properly with the respective DOIs. The DOIs of the current version (V15) of the Scientific Use Files (SUF) are:

10.7807/immo:red:hk:suf:V15 for houses for sales
10.7807/immo:red:hm:suf:V15 for houses for rent
10.7807/immo:red:wk:suf:V15 for flats for sales
10.7807/immo:red:wm:suf:V15 for flats for rent

2. Data Preparation

2.1. Data Source

The FDZ Ruhr at the RWI obtains the raw data directly from ImmobilienScout24 in form of spell data, where a spell indicates an advertisement. We obtain all active advertisements of residential real estate between the last delivery and the time of data delivery.

Users provide information on the real estate they offer themselves, and ImmobilienScout24 does not verify this information. Both, private and commercial users can use the platform to advertise offers. They are guided through an online questionnaire to create the advertisement. The questionnaire restricts the type of information that can be added depending on the real estate type, for example houses for sale cannot have an indicator whether the monthly rent includes bills. Most fields on the questionnaire are not mandatory apart from the basic information, such as address, price, and space of the real estate. At the end of the questionnaire users can further describe the real estate in a text field and give a title to their advertisement. We obtain all data input by the user apart from the heading and the free description. We do not obtain the name and contact details of the offering user either.

The use of ImmobilienScout24 is not always free for users offering real estates. The price of an advertisement varies by type of the real estate, its regions, its value, and the duration of the advertisement. For example, the price for an advertisement looking for a new tenant as the current tenant currently starts for free per fortnight. In addition, users can pay for premium features, such as special highlighting in the search results or a longer period. We do not obtain the information whether the advertisement was such a premium advertisement.

2.2. Georeferencing

Users provide the address of the real estate they offer themselves. In the most recent years, it is mandatory to put the address into the platform. However, it is possible to show only urban district or municipality for public use. ImmobilienScout24 geocodes these addresses according to their own Mercator projection. Due to inaccuracies in the first years (the non-mandatory address), several offers are not geocoded while from 2016 onwards almost all offers are geocoded.

The on-site data sets (usable in the FDZ's data security room) make the geographical location of the housing objects accessible to the user. They contain the variables: *lon_gps*, *lat_gps*, *lon_utm* and *lat_utm* which represent the longitude and latitude of the location. The used coordinate reference systems (CRS) are EPSG:4326 (WGS84) for the "_gps" variables and EPSG:25832 (ETRS89/UTM zone 32N) for the "_utm" variables.

The research data center (FDZ) Ruhr at the RWI has access to the projection data geocoded by ImmobilienScout24 and converts the coordinates into the European standard ETRS89-LAEA according to INSPIRE (INSPIRE Thematic Working Group Coordinate Reference Systems and Geographical Grid Systems, 2014). This is a grid of 1-square-meter raster cells covering all of Germany. Grid cells are each matched to administrative regions, in particular municipalities (*Gemeinden*) and districts (*Kreise*) in the territorial status as of the end of 2015 and 2019, respectively. This is done based on the shapefiles provided by the Federal Agency for Cartography and Geodesy (BKG). Additionally, cells are matched to commuting zones (Arbeitsmarktregionen). Commuting zones are based on commuting behavior between districts, and districts with strong commuter flows are subsumed to a single labor market. The definitions of commuting zones are based on the commuter flows of 2009 (Kosfeld & Werner, 2012).

2.3. Anonymization

The exact coordinates of the property are replaced by the 1x1 km grid information. Subsequently, the 1x1 km grid information is also anonymized to cover the General Data Protection Regulation of the European Union 2018 for scarcely populated grids. The information is censored for house purchase and house rent offers located in grids containing (i) less than 5 houses in total or (ii) less than 10 houses and at least 50 percent of these houses offered on the website. Exceptions to (ii) are houses that are still under construction during the time of offer. The strategy for the anonymization of apartment data is analogous, the relevant benchmark given by the number of households in a grid cell (in contrast to houses). The municipality and district information are kept in the data set after anonymization on the grid level.

2.4. Missings

Missings are consistently encoded in the following manner:

Table 3
Missing Values

Value	Description
-5	Implausible value (set by FDZ Ruhr)
-6	Old variable (no longer part of the platform)
-7	Not specified
-8	Variable for other types only
-9	Other missing
-10	Future variable
-11	Anonymized

We censor implausibly high values, encoded as -5, since they seem to result from typing errors. The respective thresholds can be found in Table 4. Additionally, some obvious typing errors like year of construction 2116 instead of 2016 are corrected. However, it is likely that there are still typing errors and possibly implausible values left. Since there is no interviewer as in surveys, we assume that the share of such mismeasured values is higher than in standard surveys. Unfortunately, it is not possible to identify them for sure.

Table 4
Censoring thresholds

Variable name	Censoring threshold
badezimmer	>15 if house, >5 if flat
baujahr	>2,120 and <1,000
etage	>45
ev_kennwert	>600
grundstuecksflaeche	>5,000
heizkosten	>1,500
kaufpreis	>50,000,000
letzte_modernisierung	>2018 and <1800
mieteinnahmen-promonat	>100,000
mietekalt, mietewarm	>15,000
nebenkosten	>1,000
nebenraeume	>10
nutzflaeche	>1,000
parkplatzpreis	>50,000 when buying, >1,000 when renting
schlafzimmer	>20 and when this is larger than total number of rooms (zimmer-anzahl)
wohnflaeche	>10,000 if house, >1,000 if flat
wohngeld	>50,000
zimmeranzahl	>25 if house, >10 if flat

Missings encoded as -6 indicate old variables. -7 represents entries where the user chose to provide no further information on the housing object. If variables are not part of a specific data type, they are encoded as -8. It represents an additional option in the questionnaire on ImmobilienScout24 when listing the object. Table 2 shows the included variables for each type. -9 summarizes all other types of missings like leaving a question on ImmobilienScout24 unanswered. -10 are variables that might arise in the future and -11 indicates entries which have been deleted due to data security concerns.

2.5. Data problems

Lack of observations

First, since most information is not mandatory, some variables are only filled for very few observations. In particular, these are heating costs (Heizkosten) for house purchases, the number of ancillary rooms (Nebenraueme) for all types, and the elevator indicator (Aufzug) for house offers, both for sale and rent.

Inconsistencies in provided information

Secondly, the way certain variables were entered online might not be consistent over time. We have detected the following inconsistencies:

1. Many of the binary variables might have been entered in different ways at different points in time, namely either as a yes/no/missing scale or as a tag, which by default can only take the value 1. We suspect that for some binary variables the way of entering them changed over time. Since old advertisements are frequently re-used, there is no clear cut in the data, but the missing patterns of many of the binary variables hint at this. One way to deal with this, is to recode all missing values as 0. Since almost all of these variables are desirable features which users are likely to advertise, the measurement error induced by this might not be too grave. In particular, this refers to the variables denkmalobjekt, enliegerwohnung, ferienhaus, gaestewc, kaufvermietet, keller, haustier_erlaubt, aufzug, balkon, garten, einbaukueche, foerderung.

2. It might not have been possible to use all variables at all times in the online questionnaire. We have adjusted missing values to indicate this when possible. However, some variables show a missing pattern which indicates that the variable was only disabled temporarily.

   (a) betreut: This variable indicating assisted living for the elderly seems to have been deprecated in later years.

   (b) ev_wwenthalten: This variable indicates whether the warm water consumption is included in the energy consumption. It shows an increasingly large share of missings from 2015 onwards, whereas before it was constantly increasing.

   (c) Variable parkplatz, indicating the availability of a parking space, seems to have been deactivated for the years 2009 to 2015 in which we observe hardly any values. Yet, the price for a parking lot (parkplatzpreis) does not mirror this pattern. This could be used by data users to infer the availability of the parking lot.

   (d) The variable rollstuhl, which indicates the accessibility of the object, has a large share missing values for the years 2009 to 2015. Potentially, it was not always possible to use this feature when creating a new advertisement, but old advertisements could still use it.

3. Since 2007, an Energy Performance Certificate (EPC) is mandatory for everybody who wants to rent or sell a property in Germany. Since May 2014, it is further mandatory to report the following information in a real estate advertisement: type of Energy Performance Certificate (energieausweistyp), the Energy Efficiency Rating for new EPCs (energieeffizienzklasse), the energy consumption (ev_kennwert). For these variables, we see a substantial increase in the share filled from 2014 onwards. It is further required to report the year the building was built (baujahr) and the fuel source for the heating system. However, many advertisements already reported these before the new law, and hence we do not observe a change in the share of missings for these variables.

4. The variable kategorie_haus, indicating the type of house, has changed categories over time. In 2008, the category “single family home” was introduced in addition to the category “single family home (detached)”, the latter of which was discontinued from 2009 onwards. Given the shares of the two categories in 2008, we recommend combining both categories to get a consistent variable over time. Likewise, the category “block of flats” was introduced in 2008 in addition to the category “two-family home” which was not used after 2008. It seems reasonable to combine both categories. The same applies for the categories “other property for living” and “other property”. Additionally, in 2008, the category “terraced house” was split up in middle unit and end unit. However, the overall category continued to be used for

some time. If the distinction between middle and end unit is not important for the analysis, one might consider merging all three categories for consistency purposes. The categories "bungalow" and "castle" were only introduced in 2008.

Table 5
Categorization of duplicate object identifiers

Code	Description
0	Identified only occurs once, or it is the first occurrence.
1	Spell most likely is part of the previous spells, e.g., split up by data delivery. Housing units with the same identifier are classified as similar if for at least one of the two groups of variables all of the variables are identical: 1. Similar features in general: rent, rent inclusive bills, heating costs in inclusive rent, price for sale, plot area, floor, building year, year of last renovation, living area, number of floors, number of rooms, number of bedrooms, number of bathrooms, energy efficiency, facility index; 2. Allowing for the manipulation of key variables (except for post code): post code, plot area, floor, building year, living area, number of floors, number of rooms, number of bedrooms, number of bathrooms.
2	Like (1), but the gap between the advertisements is larger than 6 months, i.e., this could also be a new rental offer for the same housing unit.
3	The spells are close in time (up to 6 months' gap), but there are differences in important housing unit features.
4	Like (3), but the gap between the advertisements is larger than 6 months.
5	Considerable differences in important features, hence it seems unlikely that the identifier refers to the same advertisement. Defaults for type 5 – at least one of the following is true: Price (sale, inclusive rent, exclusive rent) differs by more than 20%, Living area differs by more than 20%, Plot area differs by more than 10% (for houses for sale), Number of rooms differs by more than 1.

Duplicated observations

A third issue is that some advertisement identifiers are not unique. This can occur for several reasons: First, since we obtain spells that have not been concluded at the time of data delivery, these will also occur in the next delivery which continues from the time of the previous delivery. Moreover, users can make small changes to the advertisement to attract more people. In the data, we only observe the status of the advertisement at the time of data delivery. Hence, the same

advertisement might appear twice but with slightly different features in the data when a change was made after the delivery date. Fourth, users can temporarily set an object as inactive. This may be reasonable when a prospective buyer has committed to buy an object, but the deal has not yet been finalized. While inactive, objects will not be included in queries of potential buyers and will thus not be included in the dataset. However, if the potential buyer withdraws their offer to buy, the user might decide to reactivate the advertisement. Lastly, users might decide to use an old advertisement as a template for a new advertisement, e.g., when renting out two similar flats in the same house with only a short period in between.

In order to give a guideline how to deal with duplicate spells, we have developed a routine to classify duplicate spells as displayed in Table 5 (found in the variable dupID_gen).

Finally, it is possible, that some offerors put a duplicate into the database. In this case ImmobilienScout24 tries to identify these duplicates. In this version it is the first time we can also provide this information by the variable duplicateid.

Time-inconsistence in performance variables

The variables *hits*, *click_schnellkontake*, *click_weitersagen*, *click_url* (so-called performance variables) experience a break in values for months starting in June 2019 (delivery wave 7) and afterwards such that much lower values are registered than previously observed. This problem is related to a coding error while collecting the raw data. The issue concerns the previous versions (up to V6). The anomaly was adjusted with version 6.1. such that the performance variables report the aggregated value for periods after June 2019 (wave 7). Due to the different meaning of the variables numbers are not comparable-

To offer a better representation of performance, we also constructed performance variables which were adjusted by the ad's availability (laufzeittage) and are valid for the whole period of availability of the ad not the spell only. The variables are labelled by *_gen* (e.g., *hits_gen*). This feature is implemented since version 6.1.

3. Variable Description

In the following, we describe each variable. Variables are ordered by category.

3.1. Identifier

Table 6
Detailed variable descriptions of identifiers

Feature	Description
Label	Housing unit identifier
Name	obid
Data type	Numeric
Description	Each property is uniquely identified by an artificial ID number. IDs are property-specific and do not change over time even if the housing unit is temporarily withdrawn from the pool of advised real estates and offered again at a later time. Some IDs may be re-used over time when agents re-use previous advertisements. For duplicate spells, we provide a classification of these (see dupID_gen).
Label	Unique housing unit identifier (generated)
Name	uniqueID_gen
Data type	Numeric
Description	This is a unique identifier for each advertisement. It has been artificially generated.

3.2. Time Period

Table 7
Detailed variable descriptions of variables describing time

Feature	Description
Label	Beginning of ad, year
Name	ajahr
Data type	Numeric
Description	This is a numerical variable, which refers to the year during which a housing unit is first advertised.
Label	Beginning of ad, month
Name	amonat
Data type	Numeric
Description	This is a numerical variable, which refers to the month during which an housing unit is first advertised. If a housing unit is advertised at least at some point in time during a certain month, this advertisement is included in the respective wave. If an advertisement is updated during a specific month, only the last update is recorded and enters the dataset.
Label	Ending of ad, year
Name	ejahr
Data type	Numeric
Description	This numeric variable refers to the year of the end of the advertisement.
Label	Ending of ad, month
Name	emonat
Data type	Numeric
Description	This numeric variable refers to the month of the end of the advertisement. This can be misleading of the advertisement is split into two spells due to the timing of the data extraction from the database (see Data preparation/Data problems for more information).

3.3. Housing Unit Features

Table 8
Detailed variable descriptions of variables describing housing unit features

Feature	Description
Label	Elevator in housing unit
Name	aufzug
Data type	Numeric
Description	This variable indicates if a housing unit has an elevator.
Label	Facilities of housing unit

Name	ausstattung
Data type	Numeric
Description	This is an artificial category number indicating the facilities of the housing unit.
Label	Number of bathrooms
Name	badezimmer
Data type	Numeric
Description	This indicates the number of bathrooms in the housing unit.
Label	Balcony at housing unit
Name	balkon
Data type	Numeric
Description	This variable indicates the presence of a balcony.
Label	Protected historic building
Name	denkmalobjekt
Data type	Numeric
Description	This is an indicator of whether or not the property is protected.
Label	Kitchenette in housing unit
Name	einbaukueche
Data type	Numeric
Description	This variable indicates the presence of a fitted kitchen.
Label	Floor on which housing unit is located
Name	etage
Data type	Numeric
Description	Apartment-specific variable indicates the floor the apartment is located in.
Label	Usable as holiday home
Name	ferienhaus
Data type	Numeric
Description	This is a binary indicator for whether the housing unit can be used as a holiday home. It is only filled for flats.
Label	Available from
Name	freiab
Data type	String
Description	This variable indicates the date from which the housing unit is available.
Label	Guest toilet in housing unit
Name	gaestewc
Data type	Numeric
Description	This binary variable indicates the presence of a guest toilet.
Label	(Shared) garden available
Name	garten
Data type	Numeric
Description	This variable indicates the presence of a garden.
Label	Pets allowed
Name	haustier_erlaubt
Data type	Numeric

Description	This binary indicator shows whether pets are allowed in the housing unit.
Label	House type
Name	kategorie_Haus
Data type	Numeric
Description	The artificial category number indicates which housing unit category a property belongs to. Each property is assigned exactly one category number. This variable is filled for houses only.
Label	Flat type
Name	kategorie_Wohnung
Data type	Numeric
Description	The artificialcategory number indicates which housing unit category a property belongs to. Each property is assigned exactly one category number. This variable is filled for flats only.
Label	Cellar in housing unit
Name	keller
Data type	Numeric
Description	This variable indicates if a housing unit has a cellar or a cellar room.
Label	Garage/parking space available
Name	parkplatz
Data type	Numeric
Description	This is a binary variable indicating whether a parking space is available.
Label	Number of rooms
Name	zimmeranzahl
Data type	Numeric
Description	Number of rooms, excluding kitchen, bath or corridors. In several cases, “zimmeranzahl” is not a natural number, which is not necessarily due to a faulty entry. In Germany there is the concept of half rooms. Following the DIN 283 norm, a half room is defined as a room with a size between 6 and 10 square meters. While this definition is outdated, it is still frequently in use.

3.4. General Information of the Housing Unit

Table 9
Detailed Variable Descriptions of general housing unit information

Feature	Description
Label	Number of floors
Name	anzahletagen
Data type	Numeric
Description	This indicates the number of floors in the building.
Label	Construction phase
Name	bauphase
Data type	Numeric

Description	This variable indicates whether the housing unit is still under construction. Missings likely indicate that the housing unit is not under construction.
Label	Assisted living for the elderly
Name	betreut
Data type	Numeric
Description	This indicates whether the property is part of an assisted living for the elderly-programme.
Label	Granny flat in housing unit
Name	einliegerwohnung
Data type	Numeric
Description	This variable indicates whether a granny flat is present in the given housing unit.
Label	Public housing
Name	foerderung
Data type	Numeric
Description	This is a binary variable indicating that a certificate of eligibility to public housing is needed to rent the apartment
Label	Type of real estate
Name	immobilientyp
Data type	Numeric
Description	This artificial number indicates the type of a property.
Label	Rented when sold
Name	kaufvermietet
Data type	Numeric
Description	This variable indicates if a housing unit for sale is already rented out.
Label	Rental income per month in EUR
Name	mieteinnahmenpromonat
Data type	Numeric
Description	For housing units offered for sale, this indicates the rent income if the housing unit is rented out. This is on a monthly basis and in EUR rounded to two decimal digits.
Label	Number of ancillary rooms
Name	nebenraeume
Data type	Numeric
Description	This is the number of ancillary rooms.
Label	Accessible, no steps
Name	rollstuhlgerecht
Data type	Numeric
Description	This is a binary indicator for step-free access of the housing unit.
Label	Number of bedrooms
Name	schlafzimmer
Data type	Numeric
Description	This is the number of bedrooms of the housing unit.
Label	Common charge for community association in EUR/month
Name	wohngeld

Data type	Numeric
Description	This variable refers to the amount of the common charge for community association in EUR per month.

3.5. Area Information

Table 10
Detailed Variable Descriptions of Area Information

Feature	Description
Label	Plot area
Name	grundstuecksflaeche
Data type	Numeric
Description	This variable indicates the plot area of the housing unit in square meters. Numbers are rounded to two decimal digits.
Label	Usable floor space
Name	nutzflaeche
Data type	Numeric
Description	This indicates the usable floor space in square meters. Numbers are rounded to two decimal digits.
Label	Living area
Name	wohnflaeche
Data type	Numeric
Description	Living space in square meters. The precision of entries varies between natural numbers and numbers with two decimal places. If users enter more than two decimal places, numbers are rounded to two decimal places.

3.6. Energy and Structure Information

Table 11
Detailed Variable Descriptions of Energy and Structure Information

Feature	Description
Label	Year that housing unit was built
Name	baujahr
Data type	Numeric
Description	Year in which the housing unit was built. Observations that lie in the future are not necessarily faulty entries, potentially indicating that a housing unit is still under construction.
Label	Type of Energy Performance Certificates (EPCs)
Name	energieausweistyp
Data type	Numeric
Description	This variable indicates the type of Energy Performance Certificate that the customer has for the housing unit.

Label	Energy Efficiency Rating
Name	energieeffizienzklasse
Data type	Numeric
Description	The Energy Efficiency rating is represented here as an artificially created categorical variable.
Label	Energy consumption per year and square meter
Name	ev_kennwert
Data type	Numeric
Description	This indicated the energy consumption per year and square meter in kWh.
Label	Warm water consumption included in energy consumption
Name	ev_wwenthalten
Data type	Numeric
Description	This is a binary variable indicating whether the warm water consumption was included in the calculation of the energy consumption value.
Label	Heating costs
Name	heizkosten
Data type	Numeric
Description	This indicates the monthly heating costs in EUR and is rounded to two decimal digits.
Label	Type of heating
Name	heizungsart
Data type	Numeric
Description	This is an artificially created category number indicating the type of heating.
Label	Type of firing
Name	Bef1-bef10
Data type	Numeric
Description	This is an artificially created category number indicating the type of firing of the heating system. The information is in bef1. If there is a second type of firing it is in bef2 and so on.
Label	Year of last odernization of housing unit
Name	letzte_modernisierung
Data type	Numeric
Description	Indicator for the year of the last odernization.
Label	Condition of housing unit
Name	objektzustand
Data type	Numeric
Description	The artificial condition number indicates the condition of a property. Each property is assigned exactly one out of 11 possible numbers.

3.7. Price Information

Table 12
Detailed Variable Descriptions of Price Information

Feature	Description
Label	Brokerage at contract conclusion
Name	courtage
Data type	String
Description	Brokerage to be paid to the agent.
Label	Heating costs covered by inclusive rent
Name	heizkosten_in_wm_enthalten
Data type	Numeric
Description	This is a binary indicator showing whether heating costs are included in the inclusive rent.
Label	Purchasing price in EUR
Name	kaufpreis
Data type	Numeric
Description	Price at which the owner advertises to sell the housing unit. Prices are expressed in EUR and rounded to two decimal digits. Only housing units offered for sale.
Label	Exclusive rent in EUR
Name	mietekalt
Data type	Numeric
Description	Price at which the owner is willing to rent out the housing unit. The rent covers expenses for the living space only. Amenities as well as expenses for heating or fees such as for garbage disposal are not included. Prices are expressed in EUR and rounded to two decimal digits. This is only filled for housing units offered for rent.
Label	Security deposit
Name	mietekaution
Data type	String
Description	This specifies the security deposit to be paid before renting out the housing unit. This can either be specified in EUR, however it is also common to refer to this in terms of base rent (KM/Kaltmiete/Monatsmiete) or inclusive rent (WM/Warmmiete).
Label	Inclusive rent in EUR
Name	mietewarm
Data type	Numeric
Description	This refers to the rent inclusive of certain bills, most commonly water, heating, garbage -- but excluding electricity. This is monthly and in EUR and rounded to two decimal digits.
Label	Utilities in EUR
Name	nebenkosten
Data type	Numeric
Description	This refers to the extra monthly costs that need to be paid for bills on top of the base rent. This usually does not include electricity, but other bills. Prices are expressed in EUR and rounded to two decimal digits.

Label	Price of parking space in EUR
Name	parkplatzpreis
Data type	Numeric
Description	This refers to the price of the parking place, expressed in EUR.

3.8. Regional Information

Table 13
Detailed Variable Descriptions of Regional Information

Feature	Description
Label	German state
Name	blid
Data type	Numeric
Description	Each German federal state is attributed to a specific number. For each housing unit, this number identifies the state it is located in. The state variable is obtained from the AGS code retrieved by our georeferencing procedure, i.e., it is based on the variable kid2015. If we failed to geocode a housing unit, the variable reports the state retrieved by Immoscout's internal georeferencing. You can find these cases by filtering to those observations where there is no AGS identifier.
Label	Local labour market (Kosfeld and Werner, 2012)
Name	erg_amd
Data type	Numeric
Description	This variable indicates the local labour market following definitions of Kosfeld and Werner, 2012. Labour market regions are based on commuting behaviour between districts.
Label	1-skm raster cell following INSPIRE
Name	ergg_1km
Data type	String
Description	This variable indicates the grid cell of a 1-square-km raster of Germany according to the INSPIRE guideline. Addresses are matched to this raster based on their geocoded location.
Label	Municipality Identifier (AGS, 2015)
Name	gid2015
Data type	Numeric
Description	This is the municipality identifier according to the German Official Municipality Key (Amtlicher Gemeindeschluessel). It is based on the territoral definition of 2015 (end of year).
Label	Municipality Identifier (AGS, 2019)
Name	gid2019
Data type	Numeric
Description	This is the municipality identifier according to the German Official Municipality Key (Amtlicher Gemeindeschluessel). It is based on the territoral definition of 2019 (end of year).
Label	Municipality identifier updated by IS24 information

Name	gid_updated
Data type	Numeric
Description	This identifier is 1 for those observations with no information on the 1km² grid cell but some information on the municipality provided by ImmobilienScout24 (IS24). In this case the municipality identifier according to the German Official Municipality Key (Amtlicher Gemeindeschluessel) (see gid2019) is implemented based on IS24 information.
Label	District identifier (AGS, 2015)
Name	kid2015
Data type	Numeric
Description	This is the district identifier according to the German Official Municipality Key (Amtlicher Gemeindeschluessel). It is based on the territorial definition of 2015 (end of year).
Label	District identifier (AGS, 2019)
Name	kid2019
Data type	Numeric
Description	This is the district identifier according to the German Official Municipality Key (Amtlicher Gemeindeschluessel). It is based on the territorial definition of 2019 (end of year).
Label	District identifier updated by IS24 information
Name	kid_updated
Data type	Numeric
Description	This identifier is 1 for those observations with no information on the 1km² grid cell but some information on the district provided by ImmobilienScout24 (IS24). In this case the district identifier according to the German Official Municipality Key (Amtlicher Gemeindeschluessel) (see kid2019) is implemented based on IS24 information.
Label	Latitude (UTM) [on-site access only]
Name	lat_utm
Data type	Numeric
Description	This is the latitude of the housing units according to UTM projection (CRS: 25832). Access restricted to on-site use only.
Label	Longitude (UTM) [on-site access only]
Name	lon_utm
Data type	Numeric
Description	This is the longitude of the housing unit's address according to UTM projection (CRS: 25832). Access restricted to on-site use only.
Label	Latitude (WGS84) [on-site access only]
Name	lat_gps
Data type	Numeric
Description	This is the latitude of the housing units according to WGS84 projection (CRS: 4326). Access restricted to on-site use only.
Label	Longitude (WGS84) [on-site access only]
Name	lon_gps

Data type	Numeric
Description	This is the longitude of the housing units according to WGS84 projection (CRS: 4326). Access restricted to on-site use only.
Label	Address: postcode
Name	plz
Data type	Numeric
Description	It gives the postal code of the city the housing unit is located in. This variable is obtained by the original address and in some cases updated by our georeferencing procedure.

3.9. Meta-Information of the Advertisement

Table 14
Detailed Variable Descriptions of Variables Containing Meta-information of the Advertisement

Feature	Description
Label	Number of clicks on customer profile
Name	click_customer
Data type	Numeric
Description	This variable indicates the number of clicks on the customer profile. Valid for spell only.
Label	Number of clicks on contact button
Name	click_schnellkontakte
Data type	Numeric
Description	This variable indicates the number of clicks on the contact button. Valid for spell only.
Label	Number of clicks on customer URL
Name	click_url
Data type	Numeric
Description	This variable indicates the number of clicks on the customer's URL. Valid for spell only.
Label	Number of clicks on share button
Name	click_weitersagen
Data type	Numeric
Description	This variable indicates the number of clicks on the share button. Valid for spell only.
Label	Number of hits of ad
Name	hits
Data type	Numeric
Description	This variable shows the number of hits that the advertisement got.
Label	Days of availability of ad
Name	laufzeittage
Data type	Numeric
Description	This variable indicates the number of days the respective advertisement has been online. Valid for spell only.
Label	Number of hits through result list

Name	liste_show
Data type	Numeric
Description	The variable describes how often the advertisement appeared in the result list while searching the website. Valid for spell only.
Label	Number of hits through search request
Name	liste_match
Data type	Numeric
Description	The variable describes how often the advertisement fit a pre-specified search request. Valid for spell only.
Label	Date of data retrieval
Name	lieferung
Data type	Numeric
Description	This variable shows the year and month in which the data was extracted from the database and delivered to RWI. Advertisements which overlap with this cut-off date will hence be split and duplicated. See more under dupID_gen.
Label	Type of offeror
Name	Anbieter
Data type	Numeric
Description	This variable shows the type of offeror.
Label	Number of hits of ad, adjusted by ad availability
Name	hits_gen
Data type	Numeric
Description	This variable shows the number of hits that the advertisement got but was adjusted by the ad availability (laufzeittage).
Label	Number of clicks on contact button, adjusted by ad availability
Name	click_schnellkontakte_gen
Data type	Numeric
Description	This variable indicates the number of clicks on the contact button but was adjusted by the ad availability (laufzeittage).
Label	Number of clicks on share button, adjusted by ad availability
Name	click_weitersagen_gen
Data type	Numeric
Description	This variable indicates the number of clicks on the share button but was adjusted by the ad availability (laufzeittage).
Label	Number of clicks on customer URL, adjusted by ad availability
Name	click_url_gen
Data type	Numeric
Description	This variable indicates the number of clicks on the customer's URL but was adjusted by the ad availability (laufzeittage).
Label	Number of hits through result list, adjusted by ad availability
Name	liste_show_gen
Data type	Numeric
Description	The variable describes how often the advertisement appeared in the result list while searching the website. The variable was adjusted by the ad availability (laufzeittage).

Label	Number of hits through search request, adjusted by ad availability
Name	liste_match_gen
Data type	Numeric
Description	The variable describes how often the advertisement fit a pre-specified search request. The variable was adjusted by the ad availability (laufzeittage).

3.10. Generated Technical Variables

Table 15
Detailed Variable Descriptions of Technical Variables

Feature	Description
Label	Classification of housing unit identifiers used more than once
Name	dupID_gen
Data type	Numeric
Description	This is a generated variable, indicating whether a duplicate spell is likely to be the same housing unit, or a new housing unit. New housing units can have the same ID when customers re-use a previous advertisement for another flat, or when a housing unit is re-rented/resold. There is also the possibility to strategically change certain variables. See more under Data preparation/Data problems.
Label	Spell counter within housing unit identifier
Name	spell
Data type	Numeric
Description	This is an artificially generated variable indicating the spell within each housing unit ID, should an ID occur more than once. Spells are ordered chronologically.
Label	Obid of duplicate
Name	duplicateid
Data type	Numeric
Description	This is a variable generated by ImmobilienScout24, indicating which is a duplicate obid that is very likely to be the same housing unit.
Label	Dateset version
Name	dataset_version
Data type	String
Description	The variable indicates the version of the used dataset (e.g. v11).

4. Changes Between Versions

This section describes briefly the changes made between the current version (V15) and the previous version (V14). Other changes between earlier versions can be found in previous data reports.

The observation period was adjusted. Version V14 lists housing objects up to December 2025. Version V15 extends the period to June 2026.

5. Further Material

5.1. Value Labels

This dataset is fully labeled in both English and German. To review value labels, please download the complementary labels description file from . https://www.rwi-essen.de/en/research-advice/further/research-data-center-ruhr-fdz/data-sets/rwi-geo-red/x-real-estate-data-and-price-indices.

5.2. Summary Statistics

To get an overview of the data, we provide summary statistics of the variables by real estate type and year. These statistics can be downloaded from https://www.rwi-essen.de/en/research-advice/further/research-data-center-ruhr-fdz/data-sets/rwi-geo-red/x-real-estate-data-and-price-indices.

6. References

Bauer, T., Braun, S., & Kvasnicka, M. (2017). Nuclear Power Plant Closures and Local Housing Values: Evidence From Fukushima and the German Housing Market. *Journal of Urban Economics*, *99*, 94–106.

Bauer, T., Budde, R., Micheli, M., & Neumann, U. (2015). Immobilienmarkteffekte des Emscherumbaus? *Raumforschung und Raumordnung*, *73*(4), 269–283.

Georgi, S., & Barkow, P. (2010). Wohnimmobilien-Indizes: Vergleich Deutschland – Großbritannien [Residential Real Estate Indices – A Comparison Between Germany and the UK]. ZIA Projektberichte.

Heuer, F., & Thiel, P. (2025). Fdz data description: Real-estate data for germany (rwi-geo-red v13)-advertisements on the internet platform immobilienscout24 2007-06/2025.

Heuer, F., & Thiel, P. (2026a). Fdz data description: Real-estate data for germany campus files (rwi-geo-red panel and rwi-geo-red cross v7)-advertisements on the internet platform immobilienscout24 for teaching purposes.

Heuer, F., & Thiel, P. (2026b). Fdz data description: Regional real estate price indices for germany (rwi-geo-redx)-version 17.

INSPIRE Thematic Working Group Coordinate Reference Systems and Geographical Grid Systems. (2014). D2.8.I.2 Data Specification on Geographical Grid Systems – Technical Guidelines. http://inspire.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_GG_v3.1.pdf.

Kosfeld, R., & Werner, A. (2012). Deutsche Arbeitsmarktregionen – Neuabgrenzung nach den Kreisgebietsreformen 2007–2011. *Raumforschung und Raumordnung*, *70*(1), 49–64.