



Datenbeschreibung

RWI – Leibniz-Institut für Wirtschaftsforschung

**FDZ Data Description: Commercial Real-Estate Data
for Germany (RWI-GEO-REDC V4) - Commercial
advertisements on the Internet Platform
ImmoScout24 01/2010-06/2024**

Felix Heuer and Patrick Thiel

August 2025



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. Dr. h.c. Christoph M. Schmidt (Präsident)
Prof. Dr. Thomas K. Bauer (Vizepräsident)
Dr. Stefan Rumpf (Administrativer Vorstand)
Prof. Dr. Kerstin Schneider (Mitglied des erweiterten Vorstands)

© RWI 2025
Der Nachdruck, auch auszugsweise, ist nur mit Genehmigung des RWI gestattet.

RWI Datenbeschreibung

Schriftleitung: Prof. Dr. Dr. h. c. Christoph M. Schmidt
Gestaltung: Magdalena Franke, Claudia Lohkamp

**FDZ Data Description: Commercial Real-Estate Data for Germany
(RWI-GEO-REDC V4) – Commercial advertisements on the Internet
Platform ImmoScout24 01/2010-06/2024**

August 2025

Projektteam

RWI: Heuer, Felix and Thiel, Patrick



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	9
2 Data Preparation	11
2.1 Data Source	11
2.2 Georeferencing	11
2.3 Anonymization	12
2.4 Missings	12
2.5 Data problems	14
3 Variable Description	18
3.1 Identifier	18
3.2 Temporal Information	18
3.3 General Features	19
3.4 Features Related to Size	20
3.5 Features Related to Energy and Building Structure	20
3.6 Features Related to Price	22
3.7 Special Features	23
3.8 Regional Information	24
3.9 Generated Features	26
3.10 Features Related to Advertisement	27
4 Value Labels for Categorical Variables	29
5 Changes Between Versions	34
6 Further Material	35
7 References	36

List of Figures

List of Tables

1	Data Overview	7
2	List of Variables	7
3	Missing Values Classification	12
4	Censoring Thresholds for Implausible Values	13
5	Categorization of Duplicated Property Identifiers reflected by <i>dupID_gen</i>	15
6	Definition of Attribute Sets for Comparison	15
7	Definition of Thresholds for Code 9 of <i>dupID_gen</i>	16
8	Detailed Variable Descriptions for Identifiers	18
9	Detailed Variable Descriptions for Temporal Information	18
10	Detailed Variable Descriptions for General Features	19
11	Detailed Variable Descriptions for Features Related to Size	20
12	Detailed Variable Descriptions for Features Related to Energy and Building Structure	20
13	Detailed Variable Descriptions for Features Related to Price	22
14	Detailed Variable Descriptions for Special Features	23
15	Detailed Variable Descriptions for Regional Information	24
16	Detailed Variable Descriptions for Generated Features	26
17	Detailed Variable Descriptions for Features Related to Advertisement	27
18	List of Value for Categorical Variables	29

Abstract

The FDZ Ruhr provides a dataset on commercial real estate advertisements in Germany, drawing from information obtained from the internet platform ImmoScout24. The dataset encompasses a wide range of real estate types, including but not limited to offices, retail, and hotels. It is important to note that the data are available for scientific research purposes only. The provided dataset offers detailed regional information and a comprehensive set of characteristics. This data report offers a concise overview of the data, its limitations, and specific details. The data report is intended for (potential) users of the data as support for their data preparation.

This data report refers to RWI-GEO-REDC V4, which covers data from 01/2010 to 06/2024. Note that the years 2021 and 2022 are currently not present in the data.

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 ImmoScout24. The dataset encompasses information regarding real estate offerings on prices (listing prices and rents) as well as on various observable characteristics that determine the value of a property. It is provided on a monthly basis, with the present dataset (V4) covering 01/2010 to 06/2024. Note that the years 2021 and 2022 are currently not present in the data.

ImmoScout24 is the largest internet platform for real estate in Germany. It gives real estate owners, agents and companies the opportunity to advertise their properties for a fee. ImmoScout24 has a self-reported market share of about 50% of all real estate properties offered for sale or rent in Germany (Georgi & Barkow, 2010). The platform is open to both private and commercial entities. However, it should be noted that RWI-GEO-REDC offers data on commercial listings exclusively. For information pertaining to private listings, users are directed to the dataset RWI-GEO-RED.

In order to advertise a property, the owner is required to complete a questionnaire that requests specific information regarding the property's characteristics. Consequently, the price at which the owner is willing to sell or rent out the property is considered an offering price, or listing price. It is important to note that the advertised price is not binding; that is to say, the data does not include transaction prices. Price information is available for nearly all advertisements, and advertisers are at liberty to include additional property-specific characteristics alongside the price. This practice enables a comprehensive presentation of the property, thereby enhancing the probability of securing a favorable price or successful renting.

The dataset is distinctive in its comprehensive coverage of the commercial real estate market in Germany, a characteristic that, to the best of our knowledge, sets it apart from other available datasets. The distinctive features and advantages of the dataset can be outlined as follows.

Firstly, the dataset is sizable, encompassing approximately 5.1 mio. observations in the latest version (V4), which covers the entirety of Germany. Additionally, the dataset undergoes regular updates, thereby expanding the database continuously. This facilitates comprehensive analyses of numerous questions and enables researchers to conduct (sub-)analyses at the desired level and region.

Secondly, the dataset contains detailed information on each real estate advertisement. With 93 variables available in the on-site version of the current version (V4), which cover crucial information such as price, rent, size, and endowment, RWI-GEO-REDC enables researchers to tailor the data to various settings and extract the necessary features to answer the research question at hand.

Thirdly, RWI-GEO-REDC offers high-resolution data. In addition to being available at the monthly level to study temporal patterns, the data includes a listing of each advertisement with its address. This allows for analysis at the coordinate level. It should be noted that the FDZ Ruhr at RWI offers coordinate-level data on-site, but this data must be accessed in the data secure room. Researchers can obtain a Scientific Use File (SUF) that can be used at their workplace. This file contains a grid identifier (1 x 1 km) as the lowest regional unit.

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

1.2. Short Description

Table 1

Data Overview

Data unit	Commercial properties offered for rent or sale in Germany on ImmoScout24
Period covered	01/2010 until 06/2024 (2021 and 2022 not included)
Time reference	Monthly
Regional structure	Geocoded addresses (on-site only), as well as 1 sqm grid identifier and larger administrative areas (scientific use file)
Date of territorial status	End of 2019 (for administrative boundaries)
Sample	Full population of all properties offered on the platform
Update frequency	Continuously
Data access	Available for scientific research only
Anonymization	1 sqm grid identifier anonymized for scarcely populated cells in Scientific Use File (SUF)
Current version	01/2010–06/2024 V4.0

1.3. List of Variables

The dataset contains the following variables. Please refer to Section 5 for changes between versions.

Table 2

List of Variables

Category	Variable name	Variable label	Variable type
Identifier	obid	Property unit identifier	Character
	uniqueID_gen	Unique ID (RWI)	Numeric
Temporal information	ajahr	Beginning of offer (year)	Numeric
	amonat	Start of offer (month)	Numeric
	ejahr	End of offer (year)	Numeric
	emonat	End of offer (month)	Numeric
	available_from	Available from (on-site only)	Character
	duration_days	Duration in days	Numeric
General features	category_business	Category business	Categorical
	property_type	Property type	Categorical
	provider	Provider of ad	Categorical
Features related to size	divisible_from	Divisible from	Numeric
	plot_area	Plot area (sq. meter)	Numeric
	usable_area	Usable area (sq. meter)	Numeric
Features related to energy and building structure	bef1-bef*	Indicator for heating	Numeric

	construction_year	Construction year	Numeric
	energy_certificate_type	Energy certificate type	Categorical
	energy_consumption_index	Energy consumption index	Numeric
	energy_efficiency_class	Energy efficiency class	Categorical
	heating_type	Heating type	Categorical
	last_modernization	Last modernization	Numeric
	property_condition	Property condition	Categorical
	heating_costs	Heating costs (EUR)	Numeric
	heating_costs_included_rent	Heating costs included inclusive rent	Categorical
	warm_water_cons_included_energy_cons	Warm water consumption included in energy consumption	Categorical
Features related to price	ancillary_costs	Ancillary costs (EUR)	Numeric
	ancillary_costs_per_sqm	Ancillary costs per square meter (EUR)	Numeric
	cold_rent	Cold rent (excluding heating costs) (EUR)	Numeric
	commission	Commission (Brokerage fee) (EUR) (on-site only)	Character
	listing_price	Listing price (EUR)	Numeric
	rent_per_sqm	Rent per square meter (EUR)	Numeric
	security_deposit_months	Security deposit in months	Numeric
	security_deposit_price	Security deposit price (EUR)	Numeric
	security_deposit_type	Security deposit type	Categorical
Specific features	basement	Indicator for basement	Categorical
	elevator	Indicator for elevator	Categorical
	endowment	Endowment	Categorical
	floor	Floor	Character
	parking	Indicator for parking	Categorical
	protected_building	Indicator for protected building	Categorical
	wheelchair_accessible	Indicator for wheelchair accessibility	Categorical
Regional information	blid	Federal state ID	Categorical
	ergg_1km	1 sq. km raster cell following INSPIRE	Character
	geox	Geographical coordinate longitude (ImmoScout) (on-site only)	Numeric
	geoy	Geographical coordinate latitude (ImmoScout) (on-site only)	Numeric
	gid2019	Municipality ID 2019 (AGS)	Character

	gname2019	Municipality name 2019	Character
	house_number	House number (on-site only)	Character
	kid2019	District ID 2019	Character
	kname2019	District name 2019	Character
	lat_gps	Geographical coordinate latitude GPS (EPSG: 4326) (on-site only)	Numeric
	lat_utm	Geographical coordinate latitude GPS (EPSG: 25832) (on-site only)	Numeric
	lon_gps	Geographical coordinate longitude GPS (EPSG: 4326) (on-site only)	Numeric
	lon_utm	Geographical coordinate longitude UTM (EPSG: 25832) (on-site only)	Numeric
	street	Street (on-site only)	Character
	zipcode	Zip code	Character
	lmr2018	Labor market region ID 2018 (RWI)	Character
Generated features	dupID_gen	Classification of duplicates (RWI)	Categorical
	duplicateid	Relating obid (ImmoScout)	Character
	redc_delivery	Indicator for RWI-GEO-REDC delivery	Categorical
	redc_version	Indicator for RWI-GEO-REDC version	Character
	spell	Indicator for spell	Numeric
Features related to advertisement	num_clicks_contact_button	Number of clicks on contact button	Numeric
	num_clicks_share_button	Number of clicks on share button	Numeric
	num_clicks_url	Number of clicks on customer URL	Numeric
	num_hits_ad	Number of hits on ad	Numeric
	num_hits_result_list	Number of hits through result list	Numeric
	num_hits_search_request	Number of hits through search request	Numeric

1.4. Data Access

The data are available to researchers for non-commercial use. There are two versions of the RWI-GEO-REDC data. Firstly, the Scientific Use File (SUF) covers all information with the exception of exact geo-coordinates and other variables related to the address. Furthermore, some grid cells are anonymized due to the anonymization rules outlined in Section 2.3. Secondly, the complete dataset is available in the Data Secure Room of the FDZ Ruhr in Essen (with on-site access). The data can be obtained in both .csv and .parquet file formats. It should be noted that data access to both versions requires a signed data use agreement. Both versions are restricted to non-commercial research; only researchers from 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 (V4) of the datasets are:

Scientific Use File (SUF)

10.7807/immo:redc:suf:v4

On-site dataset

10.7807/immo:redc:v4

2. Data Preparation

2.1. Data Source

The FDZ Ruhr at the RWI acquires the raw data directly from ImmoScout24 in the form of spell data, wherein a spell denotes an advertisement. The FDZ Ruhr at the RWI obtains all active advertisements of commercial real estate between two deliveries.

Users submit information regarding the real estate they offer, and ImmoScout24 does not verify this information. The platform is available to both private and commercial users, who can utilize it for the purpose of advertising their offers. These users are guided through an online questionnaire to create their advertisements. The majority of the questionnaire's fields are not mandatory, with the exception of basic information such as the address, price, and size of the property. At the conclusion of the questionnaire, users have the option to provide additional details about the property in a text field and to assign a title to their advertisement. It is important to note that we do not obtain the name and contact information of the user who has placed the advertisement.

The use of ImmoScout24 is not always free for users offering real estates. The cost of posting an advertisement is contingent on the type of real estate, its geographical location, its market value, and the duration of the advertisement's placement. Additionally, users have the option to purchase premium features, such as enhanced visibility in search results or extended duration. However, the obtained raw data does not provide insight into whether an advertisement has been designated as a premium feature.

2.2. Georeferencing

Users are required to provide the address of the real estate they are offering. In recent years, it has become obligatory to enter the address into the platform. However, it is possible to display only the urban district or municipality for public use. ImmoScout24 geocodes these addresses according to its own Mercator projection.

The on-site dataset, which is accessible in the FDZ's data secure room, contains the geographical location of the properties. The dataset includes the variables: *lon_gps*, *lat_gps*, *lon_utm* and *lat_utm* which represent the longitude and latitude of the location. The coordinate reference systems (CRS) employed include 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 ImmoScout24 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). The grid network under consideration is comprised of 1-square-meter raster cells, which cover the entirety of Germany. Each grid cell is matched to an administrative region, specifically a municipality (*Gemeinde*) or a district (*Kreis*), according to the territorial status as of the end of 2019. This matching is conducted in accordance with the shapefiles provided by the Federal Agency for Cartography and Geodesy (BKG).

2.3. Anonymization

The precise geographical coordinates of the property are substituted by the 1x1 km grid reference information within the SUF file, representing the lowest regional information.¹ Subsequently, the 1x1 km grid information is subjected to anonymization in order to comply with the General Data Protection Regulation of the European Union (2018) for grids with sparse population density. The aforementioned grid information is censored for advertisements in grids with fewer than five business entities. The data regarding the number of businesses per grid is derived from RWI - Leibniz Institute for Economic Research (2024). The municipality and district information remain constant in the dataset subsequent to anonymization at the grid level.

2.4. Missings

Missing values are consistently encoded in the following manner:²

Table 3
Missing Values Classification

Value	Description
-5	Implausible value (set by FDZ Ruhr RWI)
-6	Not available for this delivery
-7	Not specified
-9	Other missing
-11	Value has been anonymized (RWI)

We employ a data-driven approach to censor implausible values, which we encode as -5, as they appear to be the result of typing errors (see Table 4 for details).³ However, it is plausible that there are still typing errors and implausible values remaining. Unlike in surveys, there is no interviewer to assist in the measurement process, which suggests that the prevalence of mismeasured values may be higher than in standard surveys. However, it is not possible to identify these values with certainty.

¹According to the most recent data, there are approximately 70,725 unique grids in the current version of the dataset (V4). The total sample size is approximately 5.1 mio. observations, of which 190,998 (3.8%) observations could not have been assigned corresponding grid information due to an absence of address information. This number increases to 228339 (4.5%) grids after the anonymization procedure.

²It has to be noted that users of the related RWI-GEO-RED dataset for private real estate listings may be aware of additional categories for missing values (including -10). These, at the present moment, are not required for RWI-GEO-REDC, but there is a possibility of their introduction in future versions. The order of categories remains consistent to facilitate the transition between datasets.

³Note that the censoring is applied only to incoming data, i.e. the values in Table 4 refer to threshold with respect to the latest delivery. The censoring of previous waves happened in the past and is only appended to the incoming wave.

Table 4
Censoring Thresholds for Implausible Values

Variable	Threshold	Threshold type	Value
ancillary_costs_per_sqm	0.010	percentile	1.00
ancillary_costs_per_sqm	0.990	percentile	350.00
rent_per_sqm	0.010	percentile	3.50
rent_per_sqm	0.990	percentile	38.50
listing_price	0.001	percentile	1.00
listing_price	0.999	percentile	19,456,000.00
plot_area	0.001	percentile	1.00
plot_area	0.999	percentile	721,935.06
usable_area	0.001	percentile	1.00
usable_area	0.999	percentile	120,000.00
energy_consumption_index	0.001	percentile	2.32
energy_consumption_index	0.999	percentile	871.82
cold_rent	0.001	percentile	1.00
cold_rent	0.999	percentile	376,049.90
ancillary_costs	0.001	percentile	0.50
ancillary_costs	0.999	percentile	11,012.65
divisible_from	0.001	percentile	1.00
divisible_from	0.999	percentile	22,860.90
heating_costs	0.001	percentile	NA
heating_costs	0.999	percentile	NA
last_modernization	1800.00	minimum year	1800.00
last_modernization	2024.00	maximum year	2024.00
construction_year	1000.00	minimum year	1000.00
construction_year	2024.00	maximum year	2024.00
zipcode	4.00	character length	4.00

Notes: Values outside these thresholds have been recoded as implausible (-5).

Missing data encoded as -6 indicates the presence of values that are not available for the particular delivery. -7 represents entries where the user chose not to provide additional information regarding the property.⁴ -9 summarizes all other types of missings like leaving a question on ImmoScout24 unanswered.

⁴The following are some common value entries: "keine Angabe" (no information provided), "auf Nachfrage" (upon request), and abbreviations such as "n.A." or "tbd".

2.5. Data problems

This section discusses issues the user should be aware of when working with the data.

Lack of observations

Firstly, since most information is not mandatory, some variables are only filled for very few observations.

Duplicated observations

A second issue is the presence of non-uniqueness in certain advertisement identifiers. This phenomenon may arise for a number of reasons. Firstly, as we obtain spells that have not been concluded at the time of data delivery, these will also occur in the subsequent delivery, which continues from the time of the previous delivery. Furthermore, users have the capacity to make minor alterations to the advertisement to attract a greater number of people. In the data, we only observe the status of the advertisement at the time of data delivery. Consequently, the same advertisement might appear twice but with slightly different features in the data when a change was made after the delivery date. Fourthly, users have the option to designate a property listing as inactive for a limited period. This may be a rational course of action when a prospective buyer has expressed interest in a property but the transaction has not yet been concluded. During the inactive period, properties will be excluded from queries by potential buyers and will not be incorporated into the dataset. However, if the prospective buyer retracts their offer, the user may elect to reactivate the advertisement. Finally, users may elect to employ an existing advertisement as a template for a new one.

In order to establish a framework for addressing duplicate spells, a systematic approach has been devised to categorize these instances, as illustrated in Table 5 (given in the variable *dupID_gen*). The present approach is predicated on the similarity between attributes and the time difference between spells. Two sets of attributes are distinguished: the first set, which includes variables like construction year or zip code, is more crucial, as these should not differ in order to classify two advertisements as the same property. The second set of variables, including price or condition, may vary between spells due to the addition of new features to the property or price adjustments to attract new demand (see Table 6). Given these two sets of variables, we also examine the time difference between the listing of each advertisement. We establish a cutoff of six months to differentiate between advertisements. The resulting combinations are displayed in Table 5. However, if the deviations are too significant, we cannot assume that the advertisement refers to the same property as the previous spell and would assign a code of 9 (see Table 7).

Finally, it is conceivable that some offerors may have entered a duplicate into the database. In such an instance, ImmoScout24 endeavors to identify these duplicates. The specific information is provided by the variable *duplicateid*.

Table 5
Categorization of Duplicated Property Identifiers reflected by *dupID_gen*

Code	Criterion				
	Spell = 1?	Set 1 variables identical?	Set 2 variables identical?	Deviations below thresholds?	Time difference ≤ 6 months
0	✓	-	-	-	-
1	×	✓	✓	-	✓
2	×	✓	✓	-	×
3	×	✓	×	✓	✓
4	×	✓	×	✓	×
5	×	×	✓	✓	✓
6	×	×	✓	✓	×
7	×	×	×	✓	✓
8	×	×	×	✓	×
9	×	×	×	×	-

Notes: Code refers to the encoding of the variable *dupID_gen*. The following depiction is utilized in the table: The symbol ✓ is used to indicate that the specified condition has been met, whereas × is employed to signify a deviation from the established condition. - is used to indicate that the specified condition is not applicable to the specified condition. Please refer to Table 6 and Table 7 for information on the different variable sets and the considered thresholds.

Table 6
Definition of Attribute Sets for Comparison

Set 1 variables	Set 2 variables
zipcode	listing_price
plot_area	cold_rent
construction_year	endowment
floor	category_business
	energy_efficiency_class
	heating_type
	condition

Notes: All variables are considered collectively within a designated set to ascertain the conditions delineated in Table 5. In the event that a variable deviates from the established parameters, the set cannot be regarded as identical anymore, and the two spells are deemed to be at least to some degree different.

Table 7

Definition of Thresholds for Code 9 of *dupID_gen*

Variables	Threshold
listing_price	20%
cold_rent	20%
plot_area	10%

Notes: The *threshold* refers to the difference between spells which needs to be exceeded in order to classify the difference as *too significant*. In the event that the threshold is exceeded in at least one of the specified variables, *code 9* will be assigned in *dupID_gen*.

Partial cleaning and other notes

Security deposit. The variables related to the security deposit (*security_deposit_months*, *security_deposit_price*, *security_deposit_type*) are derived from the provided security deposit variable, which is designated as a text variable. We then extract the relevant price/rent information from this text variable, specifically the security deposit price and the month, and subsequently specify the corresponding type. The initial security deposit variable has been eliminated following the extraction process.

Floor. The floor variable was originally a text variable on ImmoScout. This structure was maintained, but the variable was partially cleaned by replacing obvious cases with their corresponding numbers. For example, values such as "Keller" (basement) or "Erdgeschoss" (ground floor) are recorded as -1 and 0, respectively. This cleaning solves most cases in the data, but some more ambiguous types are not cleaned. How to deal with these is left to the researcher. However, the partial cleaning should ease the use of the variable.

3. Variable Description

Each variable is described below. The variables are arranged by category.

3.1. Identifier

Table 8

Detailed Variable Descriptions for Identifiers

Feature	Description
Label	Property unit identifier
Name	obid
Data type	Character
Description	Each property is distinguished by a unique artificial ID number. These IDs are specific to each property and remain constant over time, even in the event of a property unit being temporarily withdrawn from the pool of advertised properties and subsequently offered for sale/ rent at a later date. It should be noted that some IDs may be reused over time when agents reuse previous advertisements. In the event of duplicate spells, a classification is provided.
Label	Unique ID (RWI)
Name	uniqueID_gen
Data type	Numeric
Description	<i>uniqueID_gen</i> is a unique identifier for each advertisement. It has been artificially generated and is defined as a running ID.

3.2. Temporal Information

Table 9

Detailed Variable Descriptions for Temporal Information

Feature	Description
Label	Start of offer (year)
Name	ajahr
Data type	Numeric
Description	<i>ajahr</i> is a numerical variable, denoting the year in which the property is initially advertised.
Label	Start of offer (month)
Name	amonat
Data type	Numeric
Description	<i>amonat</i> is a numerical variable and refers to the month during which a property is first advertised. If an advertisement for a property is published at some point during a specific month, then that advertisement is included in the respective wave. However, if an advertisement is updated during a certain month, only the most recent update is recorded and entered into the dataset.

Label	End of offer (year)
Name	<i>ejahr</i>
Data type	Numeric
Description	<i>ejahr</i> is a numerical variable, referring to the year of the end of the advertisement.
Label	End of offer (month)
Name	<i>emonat</i>
Data type	Numeric
Description	<i>emonat</i> is a numerical variable, referring to the month of the end of the advertisement. It should be noted that this variable may prove to be misleading in the event that the advertisement is divided into two segments due to the timing of the data extraction process from the database.
Label	Available from (on-site only)
Name	<i>available_from</i>
Data type	Character
Description	<i>available_from</i> is a text variable used to specify the point in time at which the property in question would become available for use by the prospective buyer or tenant. It should be noted that this variable is only available for the on-site dataset.
Label	Duration in days
Name	<i>duration_days</i>
Data type	Numeric
Description	<i>duration_days</i> is a numerical variable, denoting the number of days for which the respective advertisement has been available to view online. It is important to note that the specified days are only valid for the specific spell in question.

3.3. General Features

Table 10
Detailed Variable Descriptions for General Features

Feature	Description
Label	Category business
Name	<i>category_business</i>
Data type	Categorical
Description	<i>category_business</i> is a categorical variable, denoting the category to which the property belongs. It is assigned to each property, with each property being assigned to exactly one category. The individual categories are listed in Section 4.
Label	Property type
Name	<i>property_type</i>
Data type	Categorical
Description	<i>property_type</i> is a categorical variable, indicating the type of the property. The individual categories are listed in Section 4.
Label	Provider of ad

Name	provider
Data type	Categorical
Description	<i>provider</i> is a categorical variable, indicating the provider of the advertisement. The individual categories are listed in Section 4.

3.4. Features Related to Size

Table 11
Detailed Variable Descriptions for Features Related to Size

Feature	Description
Label Name Data type Description	Divisible from <i>divisible_from</i> Numeric <i>divisible_from</i> is a numerical variable, specifying the square meters in which the property is capable of being divided.
Label Name Data type Description	Plot area (sq. meter) <i>plot_area</i> Numeric <i>plot_area</i> is a numerical variable, indicating the plot area of the property in square meters.
Label Name Data type Description	Usable area (sq. meter) <i>usable_area</i> Numeric <i>usable_area</i> is a numerical variable, indicating the usable floor space of the property in square meters.

3.5. Features Related to Energy and Building Structure

Table 12
Detailed Variable Descriptions for Features Related to Energy and Building Structure

Feature	Description
Label Name Data type Description	Indicator for firing type <i>bef1-bef*</i> Numeric <i>bef1-bef*</i> are numerical variables, indicating the type of the firing type of the property. The individual categories are listed in Section 4.
Label Name Data type Description	Construction year <i>construction_year</i> Numeric <i>construction_year</i> is a numerical variable, listing the year the property was built.
Label Name Data type	Energy certificate type <i>energy_certificate_type</i> Categorical

Description	<i>energy_certificate_type</i> is a categorical variable, indicating the energy certificate type for the property. The individual categories are listed in Section 4.
Label	Energy consumption index
Name	<i>energy_consumption_index</i>
Data type	Numeric
Description	<i>energy_consumption_index</i> is a numerical variable, describing the energy consumption per year and square meter in kWh.
Label	Energy efficiency class
Name	<i>energy_efficiency_class</i>
Data type	Categorical
Description	<i>energy_efficiency_class</i> is a categorical variable, indicating the energy efficiency class for the property. The individual categories are listed in Section 4.
Label	Heating type
Name	<i>heating_type</i>
Data type	Categorical
Description	<i>heating_type</i> is a categorical variable, indicating the type of heating. The individual categories are listed in Section 4.
Label	Last modernization
Name	<i>last_modernization</i>
Data type	Numeric
Description	<i>last_modernization</i> is a numerical variable, indicating the last year of modernization.
Label	Property condition
Name	<i>property_condition</i>
Data type	Categorical
Description	<i>property_condition</i> is a categorical variable, indicating the condition of the property. Each property is assigned exactly one property condition. The individual categories are listed in Section 4.
Label	Heating costs (EUR)
Name	<i>heating_costs</i>
Data type	Numeric
Description	<i>heating_costs</i> is a numerical variable, listing the amount paid for the heating system.
Label	Heating costs included inclusive rent
Name	<i>heating_costs_included_rent</i>
Data type	Categorical
Description	<i>heating_costs_included_rent</i> is a categorical variable, indicating whether the heating costs are already covered by the inclusive rent (warm rent).
Label	Warm water consumption included in energy consumption
Name	<i>warm_water_cons_included_energy_cons</i>
Data type	Categorical
Description	<i>warm_water_cons_included_energy_cons</i> is a categorical variable, indicating whether warm water costs are covered by the energy consumption statement.

3.6. Features Related to Price

Table 13

Detailed Variable Descriptions for Features Related to Price

Feature	Description
Label	Ancillary costs (EUR)
Name	ancillary_costs
Data type	Numeric
Description	<i>ancillary_costs</i> is a numerical variable, indicating the additional monthly costs (in Euros) that are required to be paid on top of the base rent.
Label	Ancillary costs per square meter (EUR)
Name	ancillary_costs_per_sqm
Data type	Numeric
Description	<i>ancillary_costs_per_sqm</i> is a numerical variable indicating, the additional monthly costs (in Euros) per square meter that are required to be paid on top of the base rent.
Label	Cold rent (excluding heating costs) (EUR)
Name	cold_rent
Data type	Numeric
Description	<i>cold_rent</i> is a numerical variable, denoting the rent that covers expenses for the living space only (in Euros). It should be noted that amenities, as well as expenses for heating and additional fees, are not included in this definition.
Label	Commission (Brokerage fee) (EUR) (on-site only)
Name	commission
Data type	Character
Description	<i>property_condition</i> is a text variable, describing the brokerage fee to be paid to the agent. It should be noted that this variable is only available for the on-site dataset.
Label	Listing price (EUR)
Name	listing_price
Data type	Numeric
Description	<i>listing_price</i> is a numerical variable, denoting the price a potential buyer has to pay for the property (in Euros).
Label	Rent per square meter (EUR)
Name	rent_per_sqm
Data type	Numeric
Description	<i>rent_per_sqm</i> is a numerical variable, denoting the rent per square meter a tenant has to pay for renting the property (in Euros per square meter).
Label	Security deposit in months
Name	security_deposit_months
Data type	Numeric

Description	<i>security_deposit_months</i> is a numerical variable, describing the number of months (in rent) a potential buyer or tenant have to pay as deposit. The type of rent is specified in <i>security_deposit_type</i> . See also Section 2.5 for an additional note on cleaning.
Label	Security deposit price (EUR)
Name	<i>security_deposit_price</i>
Data type	Numeric
Description	<i>security_deposit_price</i> is a numerical variable, denoting price a potential buyer or tenant have to pay as deposit (in Euros). See also Section 2.5 for an additional note on cleaning.
Label	Security deposit type
Name	<i>security_deposit_type</i>
Data type	Categorical
Description	<i>security_deposit_months</i> is a categorical variable, describing the type of deposit due for buying or renting the property. The individual categories are listed in Section 4. See also Section 2.5 for an additional note on cleaning.

3.7. Special Features

Table 14
Detailed Variable Descriptions for Special Features

Feature	Description
Label	Indicator for basement
Name	<i>basement</i>
Data type	Categorical
Description	<i>basement</i> is a categorical variable, indicating whether the property is equipped with a basement. The individual categories are listed in Section 4.
Label	Indicator for elevator
Name	<i>elevator</i>
Data type	Categorical
Description	<i>elevator</i> is a categorical variable, indicating whether the property is equipped with an elevator. The individual categories are listed in Section 4.
Label	Endowment
Name	<i>endowment</i>
Data type	Categorical
Description	<i>endowment</i> is a categorical variable, describing the endowment of the property. The individual categories are listed in Section 4.
Label	Floor
Name	<i>floor</i>
Data type	Character
Description	<i>floor</i> is a text variable, indicating the floor at which the property is located. See also Section 2.5 for an additional note on cleaning.

Label	Indicator for parking
Name	parking
Data type	Categorical
Description	<i>parking</i> is a categorical variable, indicating whether the property is equipped with a parking place. The individual categories are listed in Section 4.
Label	Indicator for protected building
Name	protected_building
Data type	Categorical
Description	<i>protected_building</i> is a categorical variable, indicating whether the property is classified as a protected building. The individual categories are listed in Section 4.
Label	Indicator for wheelchair accessibility
Name	wheelchair_accessible
Data type	Categorical
Description	<i>wheelchair_accessible</i> is a categorical variable, indicating whether the property is accessible for wheelchairs. The individual categories are listed in Section 4.

3.8. Regional Information

Table 15
Detailed Variable Descriptions for Regional Information

Feature	Description
Label	Federal state ID
Name	blid
Data type	Categorical
Description	<i>blid</i> is a categorical variable, indicating the federal state in which the property is located. The information is extracted from the district ID information. The individual categories are listed in Section 4.
Label	1 sq. km raster cell following INSPIRE
Name	ergg_1km
Data type	Character
Description	<i>ergg_1km</i> is a text variable, indicating the grid cell identifier of a 1-square-kilometer raster of Germany according to the INSPIRE guidelines.
Label	Geographical coordinate longitude (ImmoScout) (on-site only)
Name	geox
Data type	Numeric
Description	<i>geox</i> is a numerical variable, indicating the longitude of the property as defined by ImmoScout24's coordinate reference system. It should be noted that this variable is only available for the on-site dataset.
Label	Geographical coordinate latitude (ImmoScout) (on-site only)
Name	geoy

Data type	Numeric
Description	<i>geoy</i> is a numerical variable, indicating the latitude of the property as defined by ImmoScout24's coordinate reference system. It should be noted that this variable is only available for the on-site dataset.
Label	Municipality ID 2019 (AGS)
Name	<i>gid2019</i>
Data type	Character
Description	<i>gid2019</i> is a text variable, describing the municipality ID (AGS) in which the property is located. The territorial definition is based on the end of 2019.
Label	Municipality name 2019
Name	<i>gname2019</i>
Data type	Character
Description	<i>gname2019</i> is a text variable, describing the municipality name in which the property is located. The territorial definition is based on the end of 2019.
Label	House number (on-site only)
Name	<i>house_number</i>
Data type	Character
Description	<i>house_number</i> is a text variable, listing the house number belonging to the property. It should be noted that this variable is only available for the on-site dataset.
Label	District ID 2019
Name	<i>kid2019</i>
Data type	Character
Description	<i>kid2019</i> is a text variable, describing the district ID in which the property is located. The territorial definition is based on the end of 2019.
Label	District name 2019
Name	<i>kname2019</i>
Data type	Character
Description	<i>kname2019</i> is a text variable, describing the district name in which the property is located. The territorial definition is based on the end of 2019.
Label	Geographical coordinate latitude GPS (EPSG: 4326) (on-site only)
Name	<i>lat_gps</i>
Data type	Numeric
Description	<i>lat_gps</i> is a numerical variable, indicating the latitude of the property according to the WGS84 projection (CRS: 4326). It should be noted that this variable is only available for the on-site dataset.
Label	Geographical coordinate latitude UTM (EPSG: 25832) (on-site only)
Name	<i>lat_utm</i>
Data type	Numeric

Description	<i>lat_utm</i> is a numerical variable, indicating the latitude of the property according to the UTM projection (CRS: 25832). It should be noted that this variable is only available for the on-site dataset.
Label	Geographical coordinate longitude GPS (EPSG: 4326) (on-site only)
Name	<i>lon_gps</i>
Data type	Numeric
Description	<i>lon_gps</i> is a numerical variable, indicating the longitude of the property according to the WGS84 projection (CRS: 4326). It should be noted that this variable is only available for the on-site dataset.
Label	Geographical coordinate longitude UTM (EPSG: 25832) (on-site only)
Name	<i>lon_utm</i>
Data type	Numeric
Description	<i>lon_utm</i> is a numerical variable, indicating the longitude of the property according to the UTM projection (CRS: 25832). It should be noted that this variable is only available for the on-site dataset.
Label	Street (on-site only)
Name	<i>street</i>
Data type	Character
Description	<i>street</i> is a text variable, defining the street in which the property is located. It should be noted that this variable is only available for the on-site dataset.
Label	Zip code
Name	<i>zipcode</i>
Data type	Character
Description	<i>zipcode</i> is a text variable, indicating the zip code in which the observation is located. The zip code has five digits.
Label	Labor market region ID 2018 (RWI)
Name	<i>lmr2018</i>
Data type	Character
Description	<i>lmr2018</i> is a text variable, describing labor market region as defined by Breidenbach et al. (2018).

3.9. Generated Features

Table 16
Detailed Variable Descriptions for Generated Features

Feature	Description
Label	Classification of duplicates (RWI)
Name	<i>dupID_gen</i>
Data type	Categorical

Description	<i>dupID_gen</i> is a categorical variable, indicating whether a duplicate spell is likely to be a replica of the original property or a new entity. New properties can possess the same ID when customers reuse a previous advertisement for another property or when a property is resold or rented. Furthermore, there is the possibility to strategically modify certain variables. The individual categories are listed in Section 4. See also Section 2.5 for more information.
Label	Related obid (ImmoScout)
Name	duplicateid
Data type	Character
Description	<i>duplicateid</i> is a text variable generated by ImmoScout24, indicating which property IDs (obid) are likely referring to the same property.
Label	Indicator for RWI-GEO-REDC delivery
Name	redc_delivery
Data type	Categorical
Description	<i>redc_delivery</i> is a categorical variable, specifying the delivery wave the observation belongs to. The individual categories are listed in Section 4.
Label	Indicator for RWI-GEO-REDC version
Name	redc_version
Data type	Character
Description	<i>redc_version</i> is a text variable, specifying the version of the RWI-GEO-REDC dataset the observation belongs to.
Label	Indicator for spell
Name	spell
Data type	Numeric
Description	<i>spell</i> is a numerical variable, indicating the spell within a given property ID (obid). If an ID appears more than once, the spell is larger than one. Spells are ordered chronologically.

3.10. Features Related to Advertisement

Table 17

Detailed Variable Descriptions for Features Related to Advertisement

Feature	Description
Label	Number of clicks on contact button
Name	num_clicks_contact_button
Data type	Numeric
Description	<i>num_clicks_contact_button</i> is a numerical variable, identifying the number of clicks on the contact button on the advertisement on ImmoScout24.
Label	Number of clicks on share button
Name	num_clicks_share_button
Data type	Numeric

Description	<i>num_clicks_share_button</i> is a numerical variable, identifying the number of clicks on the share button on the advertisement on ImmoScout24.
Label	Number of clicks on customer URL
Name	<i>num_clicks_url</i>
Data type	Numeric
Description	<i>num_clicks_url</i> is a numerical variable, identifying the number of clicks on the customer URL button on the advertisement on ImmoScout24.
Label	Number of hits on ad
Name	<i>num_hits_ad</i>
Data type	Numeric
Description	<i>num_hits_ad</i> is a numerical variable, identifying the number of hits on the advertisement on ImmoScout24.
Label	Number of hits through result list
Name	<i>num_hits_result_list</i>
Data type	Numeric
Description	<i>num_hits_result_list</i> is a numerical variable, identifying the number of hits on the advertisement on ImmoScout24 through the result list.
Label	Number of hits through search request
Name	<i>num_hits_search_request</i>
Data type	Numeric
Description	<i>num_hits_search_request</i> is a numerical variable, identifying the number of hits on the advertisement on ImmoScout24 through search request.

4. Value Labels for Categorical Variables

Table 18
List of Value for Categorical Variables

Variable	Value	Label
basement	0	No
basement	1	Yes
bef	1	No information
bef	2	Geothermal
bef	3	Solar heating
bef	4	Pellet heating
bef	5	Gas heating
bef	6	Oil heating
bef	7	District heating
bef	8	Electricity
bef	9	Coal
bef	10	Natural gas light
bef	11	Natural gas heavy
bef	12	Liquid gas
bef	13	Steam district heating
bef	14	Wood
bef	15	Wood chips
bef	16	Coal coke
bef	17	Local heating
bef	18	No attribute assigned
bef	19	Heat supply
bef	20	Bio energy
bef	21	Wind energy
bef	22	Hydro energy
bef	23	Environmental thermal energy
bef	24	Combined heat and power fossil fuels
bef	25	Combined heat and power renewable energy
bef	26	Combined heat and power regenerative energy
bef	27	Combined heat and power bio energy
blid	1	Schleswig Holstein
blid	2	Hamburg
blid	3	Lower Saxony
blid	4	Free Hanseatic City of Bremen
blid	5	North Rhine-Westphalia
blid	6	Hesse
blid	7	Rhineland-Palatine
blid	8	Baden-Wuerttemberg
blid	9	Bavaria
blid	10	Saarland
blid	11	Berlin
blid	12	Brandenburg

blid	13	Mecklenburg-Western Pommerania
blid	14	The Free State of Saxony
blid	15	Saxony-Anhalt
blid	16	The Free State of Thuringia
category_business	1	Office and commercial buildings
category_business	2	Store
category_business	3	Living and commercial buildings
category_business	4	Sales area
category_business	5	Office building
category_business	6	Office floors
category_business	7	Special property
category_business	8	Office center
category_business	9	Exhibition space
category_business	10	Restaurant
category_business	11	Estate
category_business	12	Shopping center
category_business	13	Cafe
category_business	14	Guesthouse
category_business	15	Leisure facility
category_business	16	Commercial center
category_business	17	Hotel
category_business	18	Bar service and Lounge
category_business	19	Riding school
category_business	20	Warehouse with open space
category_business	21	Forwarding warehouse
category_business	22	High-bay warehouse
category_business	23	Club and discotheque
category_business	24	Cold storage
category_business	25	Department store
category_business	26	Factory outlet
category_business	27	Vacation bungalow
category_business	28	Practice
category_business	29	Office
category_business	30	Hall
category_business	31	Industrial hall
category_business	32	Office and warehouse buildings
category_business	33	Commercial space
category_business	34	Storage space
category_business	35	Warehouse
category_business	36	Workshop
category_business	37	Service area
category_business	38	Inn/ Tavern
category_business	39	Industrial hall with open space
category_business	40	Loft
category_business	41	Self-service market
category_business	42	Practice floor
category_business	43	Lodging house

category_business	44	Studio
category_business	45	Sales hall
category_business	46	Practice building
category_business	47	Kiosk
category_business	48	Hotel property
category_business	49	Business park
category_business	50	Farm
category_business	51	Hotel garni
category_business	52	Refrigerated warehouse
category_business	53	Winery
dupID_gen	0	First occurrence of the ID
dupID_gen	1	Identical to 1 and time difference is smaller/equal 6 months
dupID_gen	2	Identical to 1 and time difference is greater 6 months
dupID_gen	3	Differences in set 2 variables (below thresholds) and time difference is smaller/equal 6 months
dupID_gen	4	Differences in set 2 variables (below thresholds) and time difference is greater 6 months
dupID_gen	5	Differences in set 1 variables (below thresholds) and time difference is smaller/ equal 6 months
dupID_gen	6	Differences in set 1 variables (below thresholds) and time difference is greater 6 months
dupID_gen	7	Differences in set 1 and 2 variables (below thresholds) and time difference is smaller/equal 6 months
dupID_gen	8	Differences in set 1 and 2 variables (below thresholds) and time difference is greater 6 months
dupID_gen	9	Differences exceeding thresholds. Time difference is irrelevant.
elevator	0	No
elevator	1	Yes
endowment	1	Simple
endowment	2	Normal
endowment	3	Sophisticated
endowment	4	Deluxe
energy_certificate_type	1	Energy demand
energy_certificate_type	2	Energy usage
energy_efficiency_class	1	APLUS
energy_efficiency_class	2	A
energy_efficiency_class	3	B
energy_efficiency_class	4	C

energy_efficiency_class	5	D
energy_efficiency_class	6	E
energy_efficiency_class	7	F
energy_efficiency_class	8	G
energy_efficiency_class	9	H
heating_type	1	Cogeneration or combined heat and power plant
heating_type	2	Electric heating
heating_type	3	Self-contained central heating
heating_type	4	District heating
heating_type	5	Floor heating
heating_type	6	Gas heating
heating_type	7	Wood pellet heating
heating_type	8	Night storage heaters
heating_type	9	Heating by stove
heating_type	10	Oil heating
heating_type	11	Solar heating
heating_type	12	Thermal heat pump
heating_type	13	Central pump
parking	0	No
parking	1	Yes
property_condition	1	First occupancy
property_condition	2	First occupancy after refurbishment
property_condition	3	Like new
property_condition	4	Refurbished
property_condition	5	Modernized
property_condition	6	Completely renovated
property_condition	7	Well-kept
property_condition	8	Need of renovation
property_condition	9	By arrangement
property_condition	10	Dilapidated
property_type	1	Office and practice
property_type	2	Retail
property_type	3	Halls and production
property_type	4	Special trades
property_type	5	Gastronomy and hotel
protected_building	0	No
protected_building	1	Yes
provider	1	Private
provider	2	Broker
provider	3	Housing industry
provider	4	Property developer
provider	5	Financial institution
provider	6	Commercial provider
provider	7	House construction
provider	8	Relocation
security_deposit_type	1	Cold rent
security_deposit_type	2	Warm rent

security_deposit_type	3	Other rent (cold or warm)
security_deposit_type	4	Miscellaneous (insurance, credit, guarantee)
security_deposit_type	5	Price
security_deposit_type	6	No deposit
wheelchair_accessible	0	No
wheelchair_accessible	1	Yes
heating_costs_included_rent	0	No
heating_costs_included_rent	1	Yes
warm_water_cons_included_energy_cons	0	No
warm_water_cons_included_energy_cons	1	Yes
redc_delivery	1	Jun 2023
redc_delivery	2	Dec 2023
redc_delivery	3	Dec 2023-1
redc_delivery	4	Jun 2024

5. Changes Between Versions

ImmoScout24 periodically introduces modifications to the online questionnaire, which can result in the introduction of new variables and the subsequent removal of certain existing ones. Further, adjustments to the data generating process can lead to differences between versions. In the following section, an overview of these changes is provided. Previous changes are documented in the corresponding data reports.

In the updated version (V4), the data now ranges from January 2010 to June 2024, thereby encompassing a substantial historical period, thus facilitating enhanced analysis.

Furthermore, the subsequent variables are not available for the forth wave of RWI-GEO-REDC and hence, are missing for these observations.

- heating_costs
- heating_costs_included_rent
- warm_water_cons_included_energy_cons

6. Further Material

Users might find the coding for the data preparation useful. Please refer to our GitHub repository (<https://github.com/PThie/RWI-GEO-REDC>), which should be cited as: Felix Heuer and Patrick Thiel (2025). PThie/RWI-GEO-REDC: RWI-GEO-REDC V4 (V4.0). Zenodo. <https://doi.org/10.5281/zenodo.16910332>.

The variable and value labels are also delivered to the user as Excel files, in conjunction with the dataset files.

7. References

- Breidenbach, P., Budde, R., Eilers, L., Kaeding, M., & Schaffner, S. (2018). Überprüfung des Zuschnitts von Arbeitsmarktreionen für die Neuabgrenzung des GRW-Fördergebiets ab 2021. Endbericht: Studie im Auftrag des Bundesministeriums für Wirtschaft und Energie (BMWi). <https://www.econstor.eu/handle/10419/181939>
- Georgi, S., & Barkow, P. (2010). Wohnimmobilien-Indizes: Vergleich Deutschland – Großbritannien [Residential Real Estate Indices – A Comparison Between Germany and the UK]. ZIA Projektberichte.
- 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.
- RWI - Leibniz Institute for Economic Research. (2024). RWI-GEO-GRID: Socio-economic Data on Grid Level (Wave 14) [<https://doi.org/10.7807/microm:V14>. [Data set].].
- Thiel, P. (2025). FDZ Data Description: Commercial Real-Estate Data for Germany (RWI-GEO-REDC V3) - Commercial advertisements on the Internet Platform ImmoScout24 01/2010-12/2023. https://www.rwi-essen.de/fileadmin/user_upload/RWI/FDZ/FDZ_Datensatzbeschreibung_REDC_v3.pdf