

Schaffner, Sandra; Thiel, Patrick

Research Report

FDZ data description: Real-estate data for Germany campus files (RWI-GEO-RED Panel and RWI-GEO-RED Cross v5) - Advertisements on the internet platform ImmobilienScout24 for teaching purposes

RWI Datenbeschreibung

Provided in Cooperation with:

RWI – Leibniz-Institut für Wirtschaftsforschung, Essen

Suggested Citation: Schaffner, Sandra; Thiel, Patrick (2024) : FDZ data description: Real-estate data for Germany campus files (RWI-GEO-RED Panel and RWI-GEO-RED Cross v5) - Advertisements on the internet platform ImmobilienScout24 for teaching purposes, RWI Datenbeschreibung, RWI - Leibniz-Institut für Wirtschaftsforschung, Essen

This Version is available at:

<https://hdl.handle.net/10419/296881>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



Datenbeschreibung

RWI – Leibniz-Institut für Wirtschaftsforschung

**FDZ Data description: Real-Estate Data for
Germany Campus Files (RWI-GEO-RED Panel
and RWI-GEO-RED Cross v5) - Advertisements
on the Internet Platform ImmobilienScout24
for teaching purposes**

April 2024

Sandra Schaffner

Patrick Thiel



Impressum

Herausgeber:

RWI – Leibniz-Institut für Wirtschaftsforschung
Hohenzollernstraße 1-3 | 45128 Essen, Germany

Postanschrift:

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)

© RWI 2024

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: Real-Estate Data for Germany Campus Files
(RWI-GEO-RED Panel and RWI-GEO-RED Cross v5) - Advertisements on the
Internet Platform ImmobilienScout24 for teaching purposes

April 2024

Sandra Schaffner

Patrick Thiel

Datenbeschreibung

RWI – Leibniz-Institut für Wirtschaftsforschung

FDZ Data description: Real-Estate Data for Germany Campus Files (RWI-GEO-RED Panel and RWI-GEO-RED Cross v5) - Advertisements on the Internet Platform ImmobilienScout24 for teaching purposes

April 2024

**Sandra Schaffner
Patrick Thiel**

Contents

List of Figures	4
List of Tables	4
Abstract	5
1 Introduction and Short Description	6
1.1 Introduction	6
1.2 Original Database	6
2 The Campus Files	7
2.1 Panel Campus File	7
2.2 Cross-Sectional Campus File	8
2.3 Data Preparation	10
3 Variable Description	11
3.1 Identifier	11
3.2 Time Period	11
3.3 Housing Unit Features	11
3.4 General Information of the Housing Unit	13
3.5 Area Information	15
3.6 Energy and Structure Information	15
3.7 Price Information	17
3.8 Regional Information	18
3.9 Meta-Information of the Advertisement	18
3.10 Generated Technical Variables	21
4 Changes Between Versions	22
5 Data Access	23
6 References	24

List of Figures

List of Tables

1	(Rough) Number of observations per city per year in the Panel Campus File . . .	7
2	Variables covered in the different campus files	8
3	Number of municipalities in the Cross-Sectional Campus file 2023 that can be identified	10
4	Detailed Variable Descriptions of Identifiers	11
5	Detailed Variable Descriptions of Variables Describing Time	11
6	Detailed Variable Descriptions of Variables Describing Housing Unit Features . .	12
7	Detailed Variable Descriptions of General housing unit Information	13
8	Detailed Variable Descriptions of Area Information	15
9	Detailed Variable Descriptions of Energy and Structure Information	15
10	Detailed Variable Descriptions of Price Information	17
11	Detailed Variable Descriptions of Regional Information	18
12	Detailed Variable Descriptions of Variables Containing Meta-information of the Advertisement	18
13	Detailed Variable Descriptions of Technical Variables	21

Abstract

The FDZ Ruhr provides two campus files on real estate advertisements in Germany: the Panel Campus File (RWI-GEO-RED Panel) and a Cross-Sectional Campus File (RWI-GEO-RED Cross-Section). The datasets are extractions of the Scientific Use Files of RWI-GEO-RED. The Panel File covers the 15 largest cities in Germany over the whole time period, whereas the Cross-Section File covers all of Germany within one year (2023). The RWI-GEO-RED data are based on information from the internet platform ImmobilienScout24 and cover residential advertisements only. The campus files include apartments for sale and for rent and houses for sale. The data are available for lectures, tutorials, seminar thesis, bachelor and master thesis and scientific research. The provided dataset covers detailed regional information and a rich set of housing characteristics. Both datasets are samples drawn and not comprehensive databases like the Scientific Use File. The already implemented data cleaning eases the data work for students. This data report gives a brief overview of 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 is an updated version of previous reports and refers to the Campus Files v5, which covers data until 2023.

1. Introduction and Short Description

1.1. Introduction

The campus files of the RWI-GEO-RED database on the respective Scientific Use Files. Scientific Use Files (SUFs) are available for scientific research on a contractual basis. However, due to increased demand for a campus file usable within, for instance, teaching the FDZ Ruhr and ImmobilienScout24 decided to release campus files. These files do not cover the whole database regarding variables and observations. Further, the regional information is restricted to municipalities and districts, whereby grids provide the lowest level of regional information within a municipality. Due to the process of sampling, it is not possible to analyze the number of advertisements over time. However, the files are representative of all other characteristics, especially prices. Further, the size of the data files is significantly smaller than the SUFs, which makes it much easier to work with the data. There exist two types of data: a longitudinal version for the biggest cities in Germany as well as a cross-sectional version for all of Germany.

This data description gives a brief overview of the data. Since the data is based on the SUFs we also recommend the data description of the SUFs for more information. This description covers the version v5.

1.2. Original Database

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 on real estate offerings on prices as well as on various observable characteristics that determine the value of a property. The present dataset covers the time span from January 2007 to December 2023. The campus files only contain information for completed years, i.e., the last observed month is December 2023. The data of the campus files are organized into three separate datasets: houses for sale, apartments for sale, and apartments for rent.

ImmobilienScout24 is the largest internet platform for real estate in Germany. It gives real estate owners the opportunity to advertise their homes 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 homes 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). Due to the high market share of ImmobilienScout24, the dataset consists of a substantial number of observations.

To advertise a home, the owner must fill out a questionnaire asking for different characteristics of the property, including the price. This means that the price at which the owner is willing to sell or rent out the property is 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 property-specific characteristics. This helps to present a property adequately, and ideally, it increases the chance of selling at a favorable price or renting out an apartment/ house respectively.

2. The Campus Files

There exist two different types of campus files: the Panel Campus File and the Cross-sectional Campus File.

2.1. Panel Campus File

The Panel Campus File covers the 15 biggest cities in Germany. These cities are Berlin, Hamburg, Munich, Cologne, Frankfurt, Stuttgart, Düsseldorf, Leipzig, Dortmund, Essen, Bremen, Dresden, Hannover, Nuremberg and Duisburg. All of these cities have at least almost 500 thousand residents (Duisburg had about 498 thousand residents in 2019). For each of the cities a large sample is drawn in a way that there is almost the same number of advertisements each year (ending date of the advertisement). The number of observations varies between 1,000 per city per year in the houses for sale sample up to 50,000 per year in Berlin and Leipzig and apartments for rent.

Table 1 gives an overview of the number of observations per ending year and city. Most times, due to the randomization process, the exact number of observations is not specified. The sample drawn includes about 27% apartments for sale and about 37% houses for sale and apartments for rent. However, the samples only cover observations that fulfill some quality standards like fewer missing values, no outliers, etc. A more detailed description of the sample drawn is described in Section 2.3.

The Panel Campus File is still in the original spell form. Each advertisement is surveyed as a spell with a beginning and ending date. It is possible that an advertisement is deactivated and reactivated again. Then a new spell starts. Also, changes made by the offeror (more detailed information, different prices, etc.) lead to a new spell generated by the system. Therefore, the data set allows the identification of price adjustments. Table 2 gives an overview of the variables in each data set.

Table 1

(Rough) Number of observations per city per year in the Panel Campus File

City	Houses for sale	Apartments for sale	Apartments for rent
Berlin	5,300	10,700	49,700
Bremen	2,100	2,100	7,400
Cologne	2,100	4,200	15,900
Dortmund	1,100	2,100	13,400
Dresden	1,100	4,200	31,600
Duisburg	2,100	2,100	10,400
Düsseldorf	1,100	4,200	19,100
Essen	1,100	2,100	14,900
Frankfurt	1,100	4,300	15,900
Hamburg	2,100	4,300	16,000
Hannover	1,100	2,100	11,400
Leipzig	1,100	2,100	48,800
Munich	1,000	10,600	22,800
Nuremberg	1,100	2,100	7,400
Stuttgart	1,100	4,200	12,400

Notes: The exact number of observations varies due to the randomization process to a small extent.

Source: Author's table.

2.2. Cross-Sectional Campus File

The Cross-Sectional Campus File is a sample drawn from one recent year (2023) for all of Germany. The sample is drawn from all advertisements that were in the database for at least one day within the respective year. For simplification, only the last spell of the advertisement is covered. For municipalities with less than 50 observations only the district information is available with at most 100 observations drawn for this district. Afterward, a sample of 50, 200, 1000, or 5000 observations is drawn depending on the number of observations in the full sample. Table 3 gives an overview of the number of municipalities identified in the data set. While most municipalities can be identified in the houses for sale sample, this is not the case in the apartments for sale since within the majority of municipalities less than 50 apartments (after data cleaning) are advertised. Only the district is available for these observations.

Table 2

Variables covered in the different campus files

	Panel Campus File			Cross-sectional Campus File		
	House sales (HS)	Apartment sales (AS)	Apartment rents (AR)	House sales (HS)	Apartment sales (AS)	Apartment rents (AR)
obid	x	x	x	x	x	x
mietekalt			x			x
nebenkosten			x			x
kaufpreis	x	x		x	x	
mieteinnahmenpromonat	x			x		
heizkosten	x			x		
baujahr	x	x	x	x	x	x
letzte_modernisierung	x	x	x	x	x	x
wohnflaeche	x	x	x	x	x	x
grundstuecksflaeche	x	x		x	x	
nutzflaeche	x	x	x	x	x	x
etage		x	x		x	x
anzahletagen		x	x		x	x
zimmeranzahl	x	x	x	x	x	x
nebenraeume	x	x	x	x	x	x
schlafzimmer	x	x	x	x	x	x
badezimmer	x	x	x	x	x	x
parkplatzplatzpreis	x	x		x	x	
wohngeld		x			x	
ev_kennwert	x	x	x	x	x	x
laufzeittage	x	x	x	x	x	x
hits	x	x	x	x	x	x
click_schnellkontakte	x	x	x	x	x	x
click_customer	x	x	x	x	x	x
click_weitersagen	x	x	x	x	x	x
click_url	x	x	x	x	x	x
hits_gen	x	x	x	x	x	x
click_schnellkontakte_gen	x	x	x	x	x	x

Continued on next page

Table 2 - Continued from previous page

	House sales (HS)	Apartment sales (AS)	Apartment rents (AR)	House sales (HS)	Apartment sales (AS)	Apartment rents (AR)
click_weitersagen_gen	X	X	X	X	X	X
liste_show	X	X	X	X	X	X
liste_match	X	X	X	X	X	X
immobilientyp	X	X	X	X	X	X
plz	X	X	X	X	X	X
denkmalobjekt	X	X		X	X	
aufzug		X	X		X	X
balkon		X	X		X	X
einbaukueche	X	X	X	X	X	X
einliegerwohnung	X	X		X	X	
ev_wwenthalten	X	X	X	X	X	X
ferienhaus	X			X		
foerderung	X		X			X
gaestewc	X	X	X	X	X	X
garten	X	X	X	X	X	X
kaufvermietet	X			X		
heizkosten			X			X
keller	X	X	X	X	X	X
parkplatz	X	X	X	X	X	X
rollstuhlgerecht	X	X	X	X	X	X
bauphase	X	X	X	X	X	X
ausstattung	X	X	X	X	X	X
energieeffizienzklasse	X	X	X	X	X	X
energieausweistyp	X	X	X	X	X	X
haustier_erlaubt	X	X	X	X	X	X
heizungsart	X	X	X	X	X	X
kategorie_Haus	X	X	X	X	X	X
objektzustand	X	X	X	X	X	X
lieferung			X			X
kid2019	X	X	X	X	X	X
ags2019	X	X	X	X	X	X
bef1	X	X	X	X	X	X
bef2	X	X	X	X	X	X
bef3	X	X	X	X	X	X
bef4	X	X	X	X	X	X
bef5	X	X	X	X	X	X
bef6	X	X	X	X	X	X
bef7	X	X	X	X	X	X
bef8	X	X	X	X	X	X
bef9	X	X	X	X	X	X
bef10	X	X	X	X	X	X
anbieter	X	X	X	X	X	X
duplicateid	X	X	X			

Continued on next page

Table 2 – Continued from previous page

	House sales (HS)	Apartment sales (AS)	Apartment rents (AR)	House sales (HS)	Apartment sales (AS)	Apartment rents (AR)
spell	x	x	x			
adat	x	x	x	x	x	x
edat	x	x	x	x	x	x
price_sqm	x	x		x	x	
rent_sqm			x			x
ergg_1km	x	x	x	x	x	x

Source: Author's table.

Table 3

Number of municipalities in the Cross-Sectional Campus file 2023 that can be identified

NOBS	House sales (HS)	Apartment sales (AS)	Apartment rent (AR)
50	1,519	439	598
200	1,562	445	728
1,000	386	117	265
5,000	18	18	51
5,000+	3	7	17
All municipalities	3,488	1,026	1,659

Notes: All other municipalities are censored due to less than 50 observations after data cleaning. For these observations only district information is available.

Source: Author's table.

2.3. Data Preparation

Both samples are based on the Scientific Use File (SUF) and have therefore gone through some preparation. Further, we will clean the samples before they are drawn. Some variables that are not relevant for the respective category or are suffering from high shares of missing variables are deleted from the SUF. See Table 2 for an overview of all variables included. Further, those spells with a missing price/ rent and missing information on the lot size are deleted. Further, some outliers regarding the price/rent, lot size, and price per sqm/ rent per sqm have been deleted. Those spells below the 1st percentile and those above the 99th percentile are not included in the sampling. The same is true for housing units that are available after 2023 (houses/ apartments under construction for more than 1 year). Further, construction years before 1500 are set to a missing value.

3. Variable Description

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

3.1. Identifier

Table 4

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 object 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).

Source: Author's table.

3.2. Time Period

Table 5

Detailed Variable Descriptions of Variables Describing Time

Feature	Description
Label	Beginning of ad, month
Name	adat
Data type	Numeric
Description	This is a numerical variable, which refers to the month during which a 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, month
Name	edat
Data type	Numeric
Description	This numeric variable refers to the month of the end of the advertisement. This can be misleading if the advertisement is split into two spells due to the timing of the data extraction from the database (see Section 2 for more information).

Source: Author's table.

3.3. Housing Unit Features

Table 6
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 an 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 apartments.
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

Continued on next page

Table 6 – Continued from previous page

Feature	Description
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	Cellar in housing unit
Name	keller
Data type	Numeric
Description	This variable indicates if an 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.

Source: Author’s table.

3.4. General Information of the Housing Unit

Table 7
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	Granny flat in housing unit
Name	einliegerwohnung

Continued on next page

Table 7 – Continued from previous page

Feature	Description
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 to 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.

Source: Author's table.

3.5. Area Information

Table 8
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	String
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.

Source: Author's table.

3.6. Energy and Structure Information

Table 9
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 an 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

Continued on next page

Table 9 – Continued from previous page

Feature	Description
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	Year of last modernization of housing unit
Name	letzte_modernisierung
Data type	Numeric
Description	Indicator for the year of the last modernisation.
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.

Source: Author's table.

3.7. Price Information

Table 10

Detailed Variable Descriptions of Price Information

Feature	Description
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. This is only filled for 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	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 per sqm in EUR
Name	price_sqm
Data type	Numeric
Description	Calculated price per square meter by price and size of house/size of apartment
Label	Monthly Rent per sqm in EUR
Name	rent_sqm
Data type	Numeric
Description	Calculated rent per square meter by rent and size of apartment
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.

Source: Author's table.

3.8. Regional Information

Table 11
Detailed Variable Descriptions of Regional Information

Feature	Description
Label	Municipality identifier (AGS, 2019)
Name	gid2019
Data type	Numeric
Description	This is the municipality identifier according to the German Official Municipality Key (Amtlicher Gemeindegchlüssel). It is based on the territorial definition of 2019 (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 Gemeindegchlüssel). It is based on the territorial definition of 2019 (end of year).
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.
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.

Source: Author's table.

3.9. Meta-Information of the Advertisement

Table 12
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.
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.

Continued on next page

Table 12 – Continued from previous page

Feature	Description
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.
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.
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.
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.
Label	Number of hits of ad
Name	hits
Data type	Numeric
Description	This variable shows the number of hits that the advertisement got.
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. The value is adjusted by the availability of ad (i.e., 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. The value is adjusted by the availability of ad (i.e., 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. The value is adjusted by the availability of ad (i.e., 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 value is adjusted by the availability of ad (i.e., laufzeittage).
Label	Number of hits through result list, adjusted by ad availability
Name	liste_show_gen
Data type	Numeric

Continued on next page

Table 12 – Continued from previous page

Feature	Description
Description	The variable describes how often the advertisement appeared in the result list while searching the website. The value is adjusted by the availability of ad (i.e., laufzeittage).
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. The value is adjusted by the availability of ad (i.e., laufzeittage).
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.
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.

Source: Author's table.

3.10. Generated Technical Variables

Table 13
Detailed Variable Descriptions of Technical Variables

Feature	Description
Label	Firing type
Name	bef1 - bef10
Data type	categorical
Description	Firing type of heating system; if there are several heating types than they are listed in bef2, bef3 etc.
Label	Type of offeror
Name	anbieter
Data type	categorical
Description	Type of offeror like private, real-estate agent etc.
Label	Classification of housing unit identifiers used more than once
Name	duplicateid
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 apartment, or when an housing unit is re-rented/resold. There is also the possibility to change certain variables strategically. See more under Data preparation/Data problems. In City Campus File only
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. In City Campus File only

Source: Author's table.

4. Changes Between Versions

This section describes briefly the changes made between the current version (v5) and the previous version (v4). Other changes between earlier versions can be found in previous data reports.

Version 5 extends the observation period to 2023. Both campus files contain “ergg_1km” as a new variable, which provides the opportunity to perform analysis on a smaller regional level than before. See also

5. Data Access

The data can be obtained as Public Use File (PUF) from the FDZ Ruhr at RWI. The FDZ Ruhr is the research data center of the RWI - Leibniz Institute for Economic Research.

Data access does not require a data use agreement. Interested users should write an email to fdz@rwi-essen.de.

Users are requested to cite the source correctly. When using the dataset RWI-GEO-RED Campus File, please cite the data as, depending on the received version, RWI; ImmobilienScout24 (2024): RWI Real Estate Data - Campus File Panel. Version: 1. RWI - Leibniz Institute for Economic Research. Dataset. <https://doi.org/10.7807/immo:red:panel:v5>. Further, we recommend citing this data description.

6. References

Georgi, S., & Barkow, P. (2010). Wohnimmobilien-indizes: Vergleich deutschland-großbritannien [residential real estate indices-a comparison between germany and the uk]. *ZIA Projektbericht*.



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

