ECONSTOR

Make Your Publications Visible.

A Service of

ZBW

Leibniz-Informationszentrum Wirtschaft Leibniz Information Centre for Economics

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-Section)- Version V3.1: Advertisements on the Internet Platform lmmobilienScout24 for teaching purposes

RWI Datenbeschreibung

Provided in Cooperation with:

RWI – Leibniz-Institut für Wirtschaftsforschung, Essen

Suggested Citation: Schaffner, Sandra; Thiel, Patrick (2022) : FDZ Data description: Real-Estate Data for Germany Campus Files (RWI-GEO-RED Panel and RWI-GEO-RED Cross-Section)- Version V3.1: 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/10419/274653

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.



WWW.ECONSTOR.EU



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-Section)- Version V3.1: Advertisements on the Internet Platform ImmobilienScout24 for teaching purposes

August 2022

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

© RWI 2022

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

RWI Datenbeschreibung

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

FDZ Data description: Real-Estate Data for Germany Campus Files (RWI-GEO-RED Panel and RWI-GEO-RED Cross-Section)- Version V3.1: Advertisements on the Internet Platform ImmobilienScout24 for teaching purposes August 2022

Sandra Schaffner Patrick Thiel

List of contents

| List of ta | List of tables4 | | |
|-------------------------------|--|--|--|
| Abstract | 5 | | |
| 1 1.1 | Introduction and Short Description | | |
| 1.2 2 2.1 2.2 2.3 | Original data base | | |
| 3 3.1 3.2 | Variable description | | |
| 3.3 3.4 3.5 | Object features 11 General Object Information 13 Area Information 14 | | |
| 3.6 3.7 3.8 | Energy and Structure Information | | |
| 3.9 3.10 | Meta-Information of Advertisement | | |
| 4 | Changes between versions 19 | | |

List of tables

| Table 1 | (Rough) Number of observations per city and year in the Panel Campus File | 7 |
|----------|--|------|
| Table 2 | Variables covered in the different campus files | 8 |
| Table 3 | Number of municipalities in the Cross-Sectional Campus file 2019 that can be identified | . 10 |
| Table 4 | Detailed Variable Descriptions of Identifiers | . 10 |
| Table 5 | Detailed Variable Descriptions of Variables Describing Time | . 11 |
| Table 6 | Detailed Variable Descriptions of Variables Describing Object Features | . 11 |
| Table 7 | Detailed Variable Descriptions of General Object Information | . 13 |
| Table 8 | Detailed Variable Descriptions of Area Information | . 14 |
| Table 9 | Detailed Variable Descriptions of Energy and Structure Information | . 14 |
| Table 10 | Detailed Variable Descriptions of Price Information | . 16 |
| Table 11 | Detailed Variable Descriptions of Regional Information | . 16 |
| Table 12 | Detailed Variable Descriptions of Variables Containing Meta-information of the Advertise | . 17 |
| Table 13 | Detailed Variable Descriptions of Technical Variables | . 19 |

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 biggest cities in Germany over the whole time period whereas the Cross-Section File covers all Germany within one year (2021). 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 a sample drawn and not the comprehensive data base 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.

The data report refers to V3.1 of both data sets.

1 Introduction and Short Description

1.1 Introduction

The campus files of the RWI-GEO-RED data base 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 ImobilienScout24 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 zip-codes 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 regarding 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 Germany.

This data description gives a brief overview on 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 V3.1.

1.2 Original data base

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. The present dataset covers the timespan from January 2007 to June 2022. The campus files only contain information for completed years, i.e., the last observed month is December 2021. 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 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 and Barkow 2010). Due to the high market share of ImmobilienScout24, the dataset consists of a substantial number of observations.

To advertise an object, 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 object 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 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 an apartment 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 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 and 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. In most times, due to the randomization process it is not the exact number of observations. 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 fulfil some quality standards like less missing values, no outliers etc. A more detailed description of the sample drawn is described in 2.3.

| City | Houses for sale | Apartments for sale | Apartments for rent |
|------------|-----------------|---------------------|---------------------|
| Berlin | 5,300 | 10,700 | 52,200 |
| Bremen | 2,100 | 2,100 | 7,500 |
| Cologne | 2,100 | 4,300 | 16,100 |
| Dortmund | 1,100 | 2,200 | 14,400 |
| Dresden | 1,100 | 4,300 | 33,900 |
| Duisburg | 2,100 | 2,100 | 10,900 |
| Düsseldorf | 1,100 | 4,300 | 19,700 |
| Essen | 1,100 | 2,100 | 15,500 |
| Frankfurt | 1,100 | 4,300 | 16,100 |
| Hamburg | 2,100 | 4,300 | 16,100 |
| Hannover | 1,100 | 2,100 | 12,000 |
| Leipzig | 1,100 | 2,100 | 53,600 |
| Munich | 1,000 | 10,700 | 23,900 |
| Nuremberg | 1,100 | 2,100 | 7,500 |
| Stuttgart | 1,100 | 4,300 | 12,100 |

Table 1

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

The exact number of observations vary due to the randomization process to a small extent.

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 price, ...) lead to a new spell generated by the system. Therefore, the data set allows the identification of price adjustments. Table 2 gives an overview on the variables in each data set.

| | City Campus File | | Cross-sectional Campus File | | | |
|---------------------------|--------------------|------------------------|--------------------------------|--------------------|------------------------|-----------------------------|
| | houses for sale | apartments for sale | apart- ments for rent | houses for sale | apartments for sale | apart- ments for rent |
| obid | х | х | х | х | х | х |
| mietekalt | | | х | | | х |
| nebenkosten | | | х | | | х |
| kaufpreis | х | x | | x | х | |
| mieteinnahmenpromonat | х | | | х | | |
| heizkosten | Х | | | х | | |
| baujahr | х | x | х | х | x | х |
| letzte_modernisierung | Х | х | х | х | х | х |
| wohnflaeche | Х | х | х | х | х | х |
| grundstuecksfläche | Х | х | | х | х | |
| nutzflaeche | Х | х | х | х | х | х |
| etage | | x | х | | х | х |
| anzahletagen | | x | х | | х | х |
| zimmeranzahl | Х | х | х | х | х | х |
| nebenraeume | Х | х | х | х | х | х |
| schlafzimmer | Х | х | х | х | х | х |
| badezimmer | Х | х | х | х | х | х |
| parkplatzplatzpreis | X | x | | х | х | |
| wohngeld | | х | | | х | |
| ev_kennwert | Х | х | х | х | х | х |
| laufzeittage | Х | х | х | х | х | х |
| hits | Х | х | х | х | х | х |
| click_schnellkontakte | Х | х | Х | х | х | х |
| click_customer | Х | x | Х | х | X | Х |
| click_weitersagen | Х | x | Х | х | x | Х |
| click_url | X | x | х | Х | Х | х |
| hits_gen | Х | x | Х | х | x | Х |
| click_schnellkontakte_gen | X | х | X | х | X | Х |
| click_weitersagen_gen | Х | х | Х | х | Х | Х |
| liste_show | Х | x | Х | X | X | Х |
| liste_match | Х | x | Х | х | х | Х |
| immobilientyp | X | X | X | X | X | Х |
| piz | X | X | Х | X | X | Х |
| denkmalobjekt | X | X | | X | X | |
| aufzug | | x | Х | | х | Х |
| balkon | | X | X | | X | Х |
| einbaukueche | X | X | X | X | X | X |
| einliegerwonnung | X | X | | X | X | |
| ev_wwenthalten | X | Х | Х | X | X | X |
| teriennaus | X | | | X | | |
| toerderung | X | | X | | | X |
| gaestewc | х | х | х | x | х | х |

Table 2 Variables covered in the different campus files

| | City Campus File | | Cross-sectional Campus File | | | |
|------------------------|--------------------|------------------------|--------------------------------|--------------------|------------------------|-----------------------------|
| | houses for sale | apartments for sale | apart- ments for rent | houses for sale | apartments for sale | apart- ments for rent |
| garten | х | х | х | х | х | х |
| kaufvermietet | х | | | x | | |
| heizkosten | | | х | | | х |
| keller | х | x | х | x | х | х |
| parkplatz | х | х | х | х | x | х |
| rollstuhlgerecht | х | x | х | x | х | х |
| bauphase | х | x | х | х | х | х |
| ausstattung | х | x | х | х | х | х |
| energieeffizienzklasse | х | x | х | х | х | х |
| energieausweistyp | х | x | х | х | х | х |
| haustier_erlaubt | х | x | х | х | х | х |
| heizungsart | x | x | х | х | х | х |
| kategorie_Haus | х | x | х | х | х | х |
| objektzustand | х | x | х | х | х | х |
| lieferung | | | х | | | х |
| kid2019 | х | x | х | x | х | х |
| ags2019 | х | х | х | х | x | х |
| bef1 | х | x | х | x | х | х |
| bef2 | х | х | х | х | x | х |
| bef3 | х | x | х | x | х | х |
| bef4 | х | х | х | х | х | х |
| bef5 | х | x | х | x | х | х |
| bef6 | х | х | х | х | х | х |
| bef7 | х | x | х | x | х | х |
| bef8 | х | х | х | х | x | х |
| bef9 | х | x | х | х | х | х |
| bef10 | х | x | х | х | х | х |
| anbieter | х | x | х | х | х | х |
| duplicateid | х | х | х | | | |
| spell | х | x | х | | | |
| adat | x | х | х | х | х | х |
| edat | x | x | х | | | |
| price_sqm | x | х | | х | x | |
| rent_sqm | | | x | | | x |

2.2 Cross-Sectional Campus File

The Cross-Sectional Campus File is a sample drawn from one recent year (2021) for all 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. Afterwards a sample of 50, 200, 1000 or 5000 observations is drawn dependent 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. For these observations only the district is available.

| NOBS | houses for sale | apartments for sale | apartments for rent |
|--------------------|-----------------|---------------------|---------------------|
| 50 | 1,068 | 326 | 612 |
| 200 | 1,081 | 339 | 748 |
| 1000 | 1,86 | 74 | 277 |
| 5000 | 5 | 14 | 75 |
| all municipalities | 2,340 | 753 | 1,712 |

Table 3Number of municipalities in the Cross-Sectional Campus file 2019 that can be identified

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

2.3 Data preparation

Both samples are based on the Scientific Use File and have therefore gone through some preparation. Further, we implement some additional cleaning before the samples 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. Further, the information on the 1km² grid cell is not included. See Table 2 for an overview of all variables included. Further, those spells with a missing price/ rent, missing information on the lot size are deleted. Further, some outliers are deleted regarding the price/rent, lot size, and price per sqm/ rent per sqm. Those spells below the 1st percentile and those above the 99th percentile are not included in the sampling. The same is true for objects that are available after 2021 (houses/ apartments under construction for more than 1 year). Further, construction years before 1500 are set to 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 | Object 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 tempo- rarily 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 previ- ous advertisements. For duplicate spells, we provide a classification of these (see dupID_gen). | | |

Detailed Variable Descriptions of Identifiers

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 an object is first advertised. If an object 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 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). |
| | In Panel Campus File only |

3.3 Object features

| Table 6 Detailed Variable Descri | ptions of Variables Describing Object Features |
|-------------------------------------|--|
| Feature | Description |
| Label | Elevator in object |
| Name | aufzug |
| Data type | Numeric |
| Description | This variable indicates if an object has an elevator. |
| Label | Facilities of object |
| Name | ausstattung |
| Data type | Numeric |
| Description | This is an artificial category number indicating the facilities of the object. |
| Label | Number of bathrooms |
| Name | badezimmer |
| Data type | Numeric |
| Description | This indicates the number of bathrooms in the object. |
| Label | Balcony at object |
| 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 object |
| Name | einbaukueche |
| Data type | Numeric |

| Feature | Description |
|-------------------|--|
| Description | This variable indicates the presence of a fitted kitchen. |
| Label | Floor on which object 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 Data tura | terienhaus Numerie |
| Data type | |
| Description | home. It is only filled for apartments. |
| Label | Guest toilet in object |
| Name | gaestewc |
| Data type | Numeric |
| Description | This binary variable indicates the presence of a guest toilet. |
| Label | (Shared) garden available |
| Name Data type | garten |
| Description | This variable indicates the presence of a garden |
| Label | Pote allowed |
| Name | haustier erlaubt |
| Data type | Numeric |
| Description | This binary indicator shows whether pets are allowed in the object. |
| Label | House type |
| Name | kategorie_Haus |
| Data type | Numeric |
| Description | The artificial category number indicates which object category a property |
| Description | variable is filled for houses only. |
| Label | Cellar in object |
| Name | keller |
| Data type | Numeric |
| Description | This variable indicates if an object has a cellar or a cellar room. |
| Label | Garage/parking space available |
| Name | parkplatz |
| Data type | |
| Description | This is a binary variable indicating whether a parking space is available. |
| Label | Number of rooms |
| Data type | Numeric |
| butu type | Number of rooms, excluding kitchen, bath or corridors. In several cases, |
| | "zimmeranzahl" is not a natural number, which is not necessarily due to a |
| Description | faulty entry. In Germany there is the concept of half rooms. Following the |
| | 10 square meters. While this definition is outdated. it is still frequently in |
| | use. |

3.4 General Object Information

Table 7

| 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 object is still under construction. Miss- ings likely indicate that the object is not under construction. |
| Label | Granny flat in object |
| Name | einliegerwohnung |
| Data type | Numeric |
| Description | This variable indicates whether a granny flat is present in the given object. |
| 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 an object for sale is already rented out. |
| Label | Rental income per month in EUR |
| Name | mieteinnahmenpromonat |
| Data type | Numeric |
| | For objects offered for sale, this indicates the rent income if the object is |
| Description | 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 object. |
| Label | Number of bedrooms |
| Name | schlafzimmer |

Detailed Variable Descriptions of General Object Information

| Feature | Description |
|-------------|---|
| Data type | Numeric |
| Description | This is the number of bedrooms of the object. |
| 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 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 object 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 natu- ral 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 9

Detailed Variable Descriptions of Energy and Structure Information

| Feature | Description |
|-------------|--|
| Label | Year that object was built |
| Name | baujahr |
| Data type | Numeric |
| Description | Year in which the object was built. Observations that lie in the future are not necessarily faulty entries, potentially indicating that an object 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 object. |
| Label | Energy Efficiency Rating |
| Name | energieeffizienzklasse |

Campus Files RWI-GEO-RED Panel and RWI-GEO-RED Cross-Section v3.1

| Feature | Description |
|---------------|--|
| Data type | Numeric |
| Description | The Energy Efficiency rating is represented here as an artificially created cat- egorical variable. |
| Label Name | Energy consumption per year and square meter ev_kennwert |
| Data type | Numeric |
| Description | This indicated the energy consumption per year and square meter in kWh. |
| Label Name | Warm water consumption included in energy consumption 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 deci- mal 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 modernisation of object |
| Name | letzte_modernisierung |
| Data type | Numeric |
| Description | Indicator for the year of the last modernisation. |
| Label | Condition of object |
| 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 10

Detailed Variable Descriptions of Price Information

| Feature | Description |
|-------------|---|
| Label | Purchasing price in EUR |
| Name | kaufpreis |
| Data type | Numeric |
| | Price at which the owner advertises to sell the object. Prices are expressed |
| Description | in EUR and rounded to two decimal digits. This is only filled for objects of- |
| | tered for sale. |
| Label | Exclusive rent in EUR |
| Name | mietekalt |
| Data type | |
| | Price at which the owner is willing to rent out the object. The rent covers |
| Description | ing or fees such as for garbage disposal are not included. Prices are ex- |
| Description | pressed in EUR and rounded to two decimal digits. This is only filled for ob- |
| | jects offered for rent. |
| Label | Utilities in EUR |
| Name | nebenkosten |
| Data type | Numeric |
| | This refers to the extra monthly costs that need to be paid for bills on top |
| Description | 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 sgm in EUR |
| Name | price sqm |
| Data type | Numeric |
| Description | Calculated price per square meter by price and size of housesize of apart- |
| Description | ment |
| 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. |

3.8 Regional information

Table 11

| Feature | Description |
|-------------|--|
| Label | Municipality identifier (AGS, 2019) |
| Name | gid2019 |
| Data type | Numeric |
| | This is the municipality identifier according to the German Official Munici- |
| Description | pality Key (Amtlicher Gemeindeschluessel). 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 Gemeindeschluessel). It is based on the territorial defini- tion of 2019 (end of year). |
| Label | Address: postcode |
| Name | plz |
| Data type | Numeric |
| Description | It gives the postal code of the city the object is located in. This variable is obtained by the original address and in some cases updated by our georef- erencing procedure. |

3.9 Meta-Information of 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. |
| 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 |
| Lahel | Number of hits of ad |
| Name | hits |
| Data type | Numeric |
| 11 | |

| Description | This variable shows the number of hits that the advertisement got. |
|-------------------|---|
| Label Name | Number of clicks on contact button, adjusted by ad availability 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 Name | Number of clicks on share button, adjusted by ad availability 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 Name | Number of hits through search request, adjusted by ad availability 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., laufzeit-tage). |
| Label Name | Number of hits through result list, adjusted by ad availability liste_show_gen |
| Data type | Numeric |
| 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 Data tura | lautzeittage |
| Data type | Numeric |
| Description | 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 du- pID_gen. |

3.10 Generated technical variables

Table 13

| Feature | Description |
|---|---|
| Label Name Data type | Firing type bef1 – bef10 categorial |
| Description | Firing type of heating system; if there are several heating types than they are listed in bef2, bef3 etc. |
| Label Name Data type Description | Type of offeror anbieter categorial Type of offeror like private, real-estate agent etc. |
| Label | Classification of object identifiers used more than once |
| Name Data type Description | duplicateid Numeric This is a generated variable, indicating whether a duplicate spell is likely to be the same object, or a new object. New objects can have the same ID when customers re-use a previous advertisement for another apartment, or when an object is re-rented/resold. There is also the possibility to strategi- cally change certain variables. See more under Data preparation/Data prob- lems. In City Campus File only |
| Label | Spell counter within object identifier |
| Name | spell |
| Data type Description | Numeric This is an artificially generated variable indicating the spell within each ob- ject ID, should an ID occur more than once. Spells are ordered chronologi- cally. |
| | In City Campus File Only |

Detailed Variable Descriptions of Technical Variables

4 Changes between versions

Version 3 and Version 3.1 extend the observation period to 2021. The cross-sectional version of V3 and V3.1 covers 2021. V2 was limited to 2020.

Version 3.1 of both campus files contains "plz" as a new variable which provides the opportunity to perform analysis on a smaller regional level than before.

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-Section)- Version V3.1: Advertisements on the Internet Platform ImmobilienScout24 for teaching purposes

August 2022

Sandra Schaffner Patrick Thiel





Leibniz-Institut für Wirtschaftsforschung

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

