

A Service of

ZBW

Leibniz-Informationszentrum Wirtschaft Leibniz Information Centre for Economics

Schaffner, Sandra

# Research Report FDZ data description: Real-estate data for Germany (RWI-GEO-RED v3) - Advertisements on the internet platform ImmobilienScout24. 2007-06/2020

**RWI Projektberichte** 

**Provided in Cooperation with:** RWI – Leibniz-Institut für Wirtschaftsforschung, Essen

*Suggested Citation:* Schaffner, Sandra (2020) : FDZ data description: Real-estate data for Germany (RWI-GEO-RED v3) - Advertisements on the internet platform ImmobilienScout24. 2007-06/2020, RWI Projektberichte, RWI - Leibniz-Institut für Wirtschaftsforschung, Essen

This Version is available at: https://hdl.handle.net/10419/242996

#### 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 (RWI-GEO-RED v3) - Advertisements on the Internet Platform ImmobilienScout24 2007-06/2020

November 2020

Sandra Schaffner



# 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 2020

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 (RWI-GEO-RED v3) -Advertisements on the Internet Platform ImmobilienScout24 2007-06/2020 November 2020 Sandra Schaffner

# Datenbeschreibung

RWI - Leibniz-Institut für Wirtschaftsforschung

# FDZ Data description: Real-Estate Data for Germany (RWI-GEO-RED v3) - Advertisements on the Internet Platform ImmobilienScout24 2007-06/2020

November 2020

Sandra Schaffner



# List of contents

| List of tables4  |  |  |  |
|--|--|--|--|
| Abstract   |  | .5   |  |
| 1<br>1.1<br>1.2<br>1.3<br>1.4  | Introduction and Short Description<br>Introduction<br>Short Description<br>List of Variables<br>Data Access  | .6<br>.6<br>.7<br>.7                               |  |
| 2<br>2.1<br>2.2<br>2.3<br>2.4<br>2.5                                     | Data preparation 1   Data source 1   Georeferencing 1   Anonymization 1   Missings 1   Data problems 1   | LO<br>LO<br>L1<br>L1<br>L1                         |  |
| 3<br>3.1<br>3.2<br>3.3<br>3.4<br>3.5<br>3.6<br>3.7<br>3.8<br>3.9<br>3.10 | Variable description1Identifier1Time period1Object features1General Object Information1Area Information2Energy and Structure Information2Price Information2Regional information2Meta-Information of Advertisement2Generated technical variables2 | 15<br>16<br>16<br>18<br>20<br>22<br>23<br>25<br>26 |  |
| 4<br>4.1<br>4.2<br>4.3<br>4.4  | Changes between deliveries2Houses for rent2Houses for sale2Flats for Rent2Flats for Sale2  | 26<br>26<br>27<br>27<br>27<br>27                   |  |
| 5<br>5.1<br>5.2  | Further Material 2   Value Labels 2   Summary Statistics 2   | 28<br>28<br>28<br>28                               |  |
| 6  | References   | 28   |  |

# List of tables

| Table 1  | Data overview   | 7         |
|----------|---|-----------|
| Table 2  | List of variables   | 8         |
| Table 3  | Missing Values  | 11        |
| Table 4  | Censoring thresholds  | 12        |
| Table 5  | Categorization of duplicate object identifiers                          | 14        |
| Table 6  | Detailed variable descriptions of identifiers                           | 15        |
| Table 7  | Detailed variable descriptions of variables describing time             | 16        |
| Table 8  | Detailed variable descriptions of variables describing object features  | 16        |
| Table 9  | Detailed Variable Descriptions of general object information            |           |
| Table 10 | Detailed Variable Descriptions of Area Information                      | 20        |
| Table 11 | Detailed Variable Descriptions of Energy and Structure Information      | 20        |
| Table 12 | Detailed Variable Descriptions of Price Information                     | 22        |
| Table 13 | Detailed Variable Descriptions of Regional Information                  | 23        |
| Table 14 | Detailed Variable Descriptions of Variables Containing Meta-information | on of the |
|          | Advertisement   | 25        |
| Table 15 | Detailed Variable Descriptions of Technical Variables                   |           |
| Table 16 | Changes over Deliveries: Houses for Rent                                |           |
| Table 17 | Changes over Deliveries: Houses for Sale                                | 27        |
| Table 18 | Changes over Deliveries: Flats for Rent                                 | 27        |
| Table 19 | Changes over Deliveries: Flats for Sale                                 | 27        |

# Abstract

The FDZ Ruhr provides a dataset on real estate advertisements in Germany for the years 2007 to 2020 (up to June). The data are based on information from the internet platform ImmobilienScout24 and covers both apartments and houses for sale and for rent. The data are available for scientific research only. The provided dataset covers detailed regional information and a rich set of housing characteristics. This data report gives a brief overview on the data as well as its limitations and specifics. The data report is addressed to (potential) users of the data as support for their data preparation.

# 1 Introduction and Short Description

#### 1.1 Introduction

The research data center (FDZ) Ruhr at the RWI provides a unique dataset on German real estate prices, obtained by ImmobilienScout24. The dataset entails information of real estate offerings on prices as well as on various observable characteristics that determine the value of a property. It is provided on a monthly basis. The present dataset covers January 2007 until June 2020. The data are organized into four separate datasets: houses for sale, houses for rent, flats for sale, and flats for rent.

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

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

This unique dataset has many advantages compared to other house price data available for the German real estate market. First, due to the high market share of ImmobilienScout24, the dataset consists of a substantial number of observations. This allows for analyzing small-scale housing markets, such as the investigation of the effect of policy intervention on local house prices (e.g. Bauer et al. 2015) and house price responses to location-specific shocks (e.g. Bauer et al. 2017). Second, the FDZ Ruhr generates regional price indices for apartments and houses based on these data which are updated regularly (see DOI: <u>http://doi.org/10.7807/immo:redx:V4</u>.). Those price indices are provided for districts (*Kreise*) and municipalities (*Verbandsgemeinden*). See Klick and Schaffner (2019) for more details. Both the raw data and the price indices are made available for scientific research.

Besides the described scientific use file and the on-site use, there exist two versions of campus files that are available for teaching purposes. For more details see Schaffner (2020).

This data description is an updated version of Boelmann and Schaffner (2019). Therefore, there are several text parts that are the same.

# 1.2 Short Description

# Table 1

# Data overview

| Data unit                  | Residential objects offered for rent or sale in Germany on Immo-<br>bilienscout24.de                                 |
|----------------------------|--|
| Period covered             | January 2007 until June 2020   |
| Time reference             | Monthly  |
| Regional structure         | Geocoded addresses (on-site), as well as 1 sqm grid identifier and larger administrative areas (scientific use file) |
| Date of territorial status | End of 2015 (for administrative boundaries)  |
| Sample                     | Full population of all objects offered on platform   |
| Update frequency           | Continuous   |
| Data access                | Available for scientific research only   |
| Anonymization              | 1 sqm grid identifier anonymized for scarcely populated cells in scientific use file                                 |
| Current version            | 01/2007 –06/2020 V3.0  |

# 1.3 List of Variables

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

# Table 2 List of variables

| Category    | Variable                        | Description  | House | House | Flat | Flat |
|-------------|---------------------------------|--|-------|-------|------|------|
|             | name                            |  | rent  | sale  | rent | sale |
| Identifier  | obid                            | Object identifier  | 1     | 1     | 1    | 1    |
|             | uniqueID_gen                    | Unique object identifier (generated)                       | 1     | 1     | 1    | 1    |
| Time period | ajahr                           | Beginning of ad, year                                      | 1     | 1     | 1    | 1    |
|             | amonat                          | Beginning of ad, month                                     | 1     | 1     | 1    | 1    |
|             | ejahr                           | Ending of ad, year   | 1     | 1     | 1    | 1    |
|             | emonat                          | Ending of ad, month  | 1     | 1     | 1    | 1    |
| Object      | aufzug                          | Elevator in object   | 1     | 0     | 1    | 1    |
| features    | ausstattung                     | Facilities of object                                       | 1     | 1     | 1    | 1    |
|             | badezimmer                      | Number of bathrooms  | 1     | 1     | 1    | 1    |
|             | balkon                          | Balcony at object  | 0     | 0     | 1    | 1    |
|             | denkmalobjekt                   | Protected historic building                                | 0     | 1     | 0    | 1    |
|             | einbaukueche                    | Kitchenette in object                                      | 1     | 0     | 1    | 1    |
|             | etage                           | Floor on which object is located                           | 0     | 0     | 1    | 1    |
|             | ferienhaus                      | Usable as holiday home                                     | 0     | 1     | 0    | 1    |
|             | freiab                          | Available from   | 1     | 1     | 1    | 1    |
|             | gaestewc                        | Guest toilet in object                                     | 1     | 1     | 1    | 1    |
|             | garten                          | (Shared) garden available                                  | 0     | 0     | 1    | 1    |
|             | haustier_er-<br>laubt           | Pets allowed   | 1     | 0     | 1    | 0    |
|             | kategorie_Haus                  | House type   | 1     | 1     | 0    | 0    |
|             | kategorie_Woh-<br>nung          | Flat type  | 0     | 0     | 1    | 1    |
|             | keller                          | Cellar in object   | 1     | 1     | 1    | 1    |
|             | parkplatz                       | Garage/parking space available                             | 1     | 1     | 1    | 1    |
|             | zimmeranzahl                    | Number of rooms  | 1     | 1     | 1    | 1    |
|             | anzahletagen                    | Number of floors   | 1     | 1     | 1    | 1    |
|             | bauphase                        | Construction phase   | 0     | 1     | 0    | 0    |
|             | betreut                         | Assisted living for the elderly                            | 0     | 0     | 0    | 0    |
|             | einliegerwoh-<br>nung           | Granny flat in object                                      | 0     | 1     | 0    | 0    |
|             | foerderung                      | Public housing – certificate of eligibil-<br>ity is needed | 0     | 0     | 1    | 0    |
|             | immobilientyp                   | Type of real estate  | 1     | 1     | 1    | 1    |
|             | kaufvermietet                   | Rented when sold   | 0     | 1     | 0    | 1    |
|             | mietein-<br>nahmen-<br>promonat | Rental income per month in EUR                             | 0     | 1     | 0    | 1    |
|             | nebenraeume                     | Number of ancillary rooms                                  | 0     | 0     | 0    | 0    |
|             | rollstuhlgerecht                | Accessible, no steps                                       | 1     | 1     | 1    | 1    |
|             | schlafzimmer                    | Number of bedrooms   | 1     | 1     | 1    | 1    |
|             | wohngeld                        | Common charge for community asso-                          | 0     | 0     | 0    | 1    |
|             | grundstuecks-<br>flaeche        | Plot area  | 1     | 1     | 0    | 0    |
|             | nutzflaeche                     | Usable floor space   | 1     | 1     | 1    | 1    |
|             | wohnflaeche                     | Living area  | 1     | 1     | 1    | 1    |
|             | baujahr                         | Year that object was built                                 | 1     | 1     | 1    | 1    |

| <u>.</u>                                 | Variable                            | <b>-</b> · · ·   | House | House | Flat | Flat |
|--|-------------------------------------|--|-------|-------|------|------|
| Category                                 | name                                | Description  | rent  | sale  | rent | sale |
| Energy and<br>structure in-<br>formation | energieausweis-<br>typ              | Type of Energy Performance Certificates (EPCs)           | 1     | 1     | 1    | 1    |
|  | energieeffizien-<br>zklasse         | Energy Efficiency Rating                                 | 1     | 1     | 1    | 1    |
|  | ev_kennwert                         | Energy consumption per year and square meter             | 1     | 1     | 1    | 1    |
|  | ev_wwenthal-<br>ten                 | Warm water consumption included in energy consumption    | 1     | 1     | 1    | 1    |
|  | heizkosten                          | Heating costs  | 1     | 0     | 1    | 0    |
|  | heizungsart                         | Type of heating  | 1     | 1     | 1    | 1    |
|  | bef1-bef10                          | firing type  | 1     | 1     | 1    | 1    |
|  | letzte_modern-<br>isierung          | Year of last modernisation of object                     | 1     | 1     | 1    | 1    |
|  | objektzustand                       | Condition of object                                      | 1     | 1     | 1    | 1    |
| Price infor-<br>mation                   | courtage                            | Brokerage at contract conclusion                         | 1     | 1     | 1    | 1    |
|  | heiz-<br>kosten_in_wm_<br>enthalten | Heating costs covered by inclusive rent                  | 1     | 0     | 1    | 0    |
|  | kaufpreis                           | Purchasing price in EUR                                  | 0     | 1     | 0    | 1    |
|  | mietekalt                           | Exclusive rent in EUR                                    | 1     | 0     | 1    | 0    |
|  | mietekaution                        | Security deposit   | 1     | 1     | 1    | 1    |
|  | mietewarm                           | Inclusive rent in EUR                                    | 1     | 0     | 1    | 0    |
|  | nebenkosten                         | Utilities in EUR   | 1     | 0     | 1    | 0    |
|  | parkplatzpreis                      | Price of parking space in EUR                            | 1     | 1     | 1    | 1    |
| Regional in-                             | blid                                | German state   | 1     | 1     | 1    | 1    |
| formation                                | erg_amd                             | Local labour market (Kosfeld and Werner, 2012)           | 1     | 1     | 1    | 1    |
|  | ergg_1km                            | 1-skm raster cell following INSPIRE                      | 1     | 1     | 1    | 1    |
|  | gid2015                             | Municipality Identifier (AGS, 2015)                      | 1     | 1     | 1    | 1    |
|  | ags2019                             | Municipality Identifier (AGS, 2019)                      | 1     | 1     | 1    | 1    |
|  | kid2015                             | District identifier (AGS, 2015)                          | 1     | 1     | 1    | 1    |
|  | kid2019                             | District identifier (AGS, 2019)                          | 1     | 1     | 1    | 1    |
|  | plz                                 | Address: postcode  | 1     | 1     | 1    | 1    |
| Meta-infor-                              | click_customer                      | Number of clicks on customer profile                     | 1     | 1     | 1    | 1    |
| mation of ad                             | click_schnellkon<br>takte           | Number of clicks on contact button                       | 1     | 1     | 1    | 1    |
|  | click_url                           | Number of clicks on customer URL                         | 1     | 1     | 1    | 1    |
|  | click_weiter-<br>sagen              | Number of clicks on share button                         | 1     | 1     | 1    | 1    |
|  | hits                                | Number of hits of ad                                     | 1     | 1     | 1    | 1    |
|  | laufzeittage                        | Days of availability of ad                               | 1     | 1     | 1    | 1    |
|  | lieferung                           | Date of data retrieval                                   | 1     | 1     | 1    | 1    |
| Generated technical var-                 | dupID_gen                           | Classification of object identifiers used more than once | 1     | 1     | 1    | 1    |
| iables                                   | duplicate_id                        | obid of duplicate  | 1     | 1     | 1    | 1    |
|  | spell                               | Spell counter within object identifier                   | 1     | 1     | 1    | 1    |

#### 1.4 Data Access

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

Data access is provided by the Research Data Centre Ruhr at the RWI – Leibniz-Institute for Economic Research (FDZ Ruhr). Data access can be applied for online at http://fdz.rwi-es-sen.de/application.html. 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 V3 of the Scientific Use Files are:

| 10.7807/immo:red:hk:suf:v3 | for houses for sales |
|----------------------------|----------------------|
| 10.7807/immo:red:hm:suf:v3 | for houses for rent  |
| 10.7807/immo:red:wk:suf:v3 | for flats for sales  |
| 10.7807/immo:red:wm:suf:v3 | for flats for rent   |

# 2 Data preparation

# 2.1 Data source

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

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

The use of ImmobilienScout24 is not free for users offering real estates. The price of an advertisement varies by type of the real estate, its regions, its value, and the duration of the advertisement. For example, the price for an advertisement looking for a new tenant as the current tenant currently starts from 6.05 EUR per fortnight. In addition, users can pay for premium features, such as special highlighting in the search results. We do not obtain the information whether the advertisement was such a premium advertisement.

# 2.2 Georeferencing

Users provide the address of the real estate they offer themselves. In the most recent years, it is mandatory to put the address into the platform. However, it is possible to show only urban district or municipality for public use. ImmobilienScout24 geocodes these addresses according to their own Mercator projection. Due to inaccuracies in the first years (the non-mandatory address), several offers are not geocoded while from 2016 onwards all offers are geocoded. The research data center (FDZ) Ruhr at the RWI has access to the projection data geocoded by ImmobilienScout24 and converts the coordinates into the European standard ETRS89-LAEA according to INSPIRE (INSPIRE 2014). This is a grid of 1-square-meter raster cells covering all of Germany. Grid cells are each matched to administrative regions, in particular municipalities (*Gemeinden*) and districts (*Kreise*) in the territorial status as of the end of 2015. This is done based on the shapefiles provided by the Federal Agency for Cartography and Geodesy (BKG). Additionally, cells are matched to labor market regions. Labor market regions are based on commuting behavior between districts, and districts with strong commuter flows are subsumed to a single labor market. The definitions of labor market regions are based on the commuter flows of 2009 (Kosfeld and Werner 2012).

# 2.3 Anonymization

The exact coordinates of the property are replaced by the 1x1km grid information. Subsequently, the 1x1km grid information is also anonymized to cover the General Data Protection Regulation of the European Union 2018 for scarcely populated grids. The information is censored for house purchase and house rent offers located in grids containing (i) less than 5 houses in total or (ii) less than 10 houses and at least 50 percent of these houses offered on the website. Exceptions to (ii) are houses that are still under construction during the time of offer. The strategy for the anonymization of apartment data is analogous, the relevant benchmark given by the number of households in a grid cell (in contrast to houses). The municipality and district information are kept in the data set after anonymization on the grid level.

# 2.4 Missings

Missings are consistently encoded in the following manner:

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

#### Table 3 Missing Values

We censor implausibly high values, encoded as -5, since they seem to result from typing errors. The respective thresholds can be found in Table 2. Additionally, some obvious typing errors like year of construction 2116 instead of 2016 are corrected. However, it is likely that there are still typing errors and possibly implausible values left. Since there is no interviewer as in surveys, we

assume that the share of such mismeasured values is higher than in standard surveys. Unfortunately, it is not possible to identify them for sure.

Table 4 Censoring thresholds

| Variable name         | Censoring threshold  |
|-----------------------|--|
| badezimmer            | >15 if house, >5 if flat   |
| baujahr               | >2120 and <1000  |
| etage                 | >45  |
| ev_kennwert           | >600   |
| grundstuecksflaeche   | >5000  |
| heizkosten            | >1500  |
| kaufpreis             | >50,000,000  |
| letzte_modernisierung | >2018 and <1800  |
| mieteinnahmenpromonat | >100,000   |
| mietekalt, mietewarm  | >15,000  |
| nebenkosten           | >1000  |
| nebenraeume           | >10  |
| nutzflaeche           | >1000  |
| parkplatzpreis        | >50,000 when buying, >1000 when renting                                    |
| schlafzimmer          | >20 and when this is larger than total number of rooms (zim-<br>meranzahl) |
| wohnflaeche           | >10,000 if house, >1000 if flat  |
| wohngeld              | >50,000  |
| zimmeranzahl          | >25 if house, >10 if flat  |

# 2.5 Data problems

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

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

1. Many of the binary variables might have been entered in different ways at different points in time, namely either as a yes/no/missing scale or as a tag, which by default can only take the value 1. We suspect that for some binary variables the way of entering them changed over time. Since old advertisements are frequently re-used, there is no clear cut in the data, but the missing patterns of many of the binary variables hint at this. One way to deal with this, is to recode all missing values as 0. Since almost all of

these variables are desirable features which users are likely to advertise, the measurement error induced by this might not be too grave. In particular, this refers to the variables denkmalobjekt, enliegerwohnung, ferienhaus, gaestewc, kaufvermietet, keller, haustier\_erlaubt, aufzug, balkon, garten, einbaukueche, foerderung.

- It might not have been possible to use all variables at all times in the online questionnaire. We have adjusted missing values to indicate this when possible. However, some variables show a missing pattern which indicates that the variable was only disabled temporarily.
  - a. betreut: This variable indicating assisted living for the elderly seems to have been deprecated in later years.
  - b. ev\_wwenthalten: This variable indicates whether the warm water consumption is included in the energy consumption. It shows an increasingly large share of missings from 2015 onwards, whereas before it was constantly increasing.
  - c. Variable parkplatz, indicating the availability of a parking space, seems to have been deactivated for the years 2009 to 2015 in which we observe hardly any values. Yet, the price for a parking lot (parkplatzpreis) does not mirror this pattern. This could be used by data users to infer the availability of the parking lot.
  - d. The variable rollstuhl, which indicates the accessibility of the object, has a large share missing values for the years 2009 to 2015. Potentially, it was not always possible to use this feature when creating a new advertisement, but old advertisements could still use it.
- 3. Since 2007, an Energy Performance Certificate (EPC) is mandatory for everybody who wants to rent or sell a property in Germany. Since May 2014, it is further mandatory to report the following information in a real estate advertisement: type of Energy Performance Certificate (energieausweistyp), the Energy Efficiency Rating for new EPCs (energieeffizienzklasse), the energy consumption (ev\_kennwert). For these variables, we see a substantial increase in the share filled from 2014 onwards. It is further required to report the year the building was built (baujahr) and the fuel source for the heating system. However, many advertisements already reported these before the new law, and hence we do not observe a change in the share of missings for these variables.
- 4. The variable kategorie\_haus, indicating the type of house, has changed categories over time. In 2008, the category "single family home" was introduced in addition to the category "single family home (detached)", the latter of which was discontinued from 2009 onwards. Given the shares of the two categories in 2008, we recommend combining both categories to get a consistent variable over time. Likewise, the category "block of flats" was introduced in 2008 in addition to the category "two-family home" which was not used after 2008. It seems reasonable to combine both categories. The same applies for the category "terraced house" was split up in middle unit and end unit. However, the overall category continued to be used for some time. If the distinction between middle and end unit is not important for the analysis, one might consider merging all three categories for consistency purposes. The categories "bungalow" and "castle" were only introduced in 2008.

# Table 5

#### Categorization of duplicate object identifiers

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

A third issue is that some advertisement identifiers are not unique. This can occur for several reasons: First, since we obtain spells that have not been concluded at the time of data delivery, these will also occur in the next delivery which continues from the time of the previous delivery. Moreover, users can make small changes to the advertisement in order to attract more people. In the data, we only observe the status of the advertisement at the time of data delivery. Hence, the same advertisement might appear twice but with slightly different features in the data when a change was made after the delivery date. Fourth, users can temporarily set an object as inactive. This may be reasonable when a prospective buyer has committed to buy an object, but the deal has not yet been finalized. While inactive, objects will not be included in queries of potential buyers and will thus not be included in the dataset. However, if the potential buyer withdraws

their offer to buy, the user might decide to reactivate the advertisement. Lastly, users might decide to use an old advertisement as a template for a new advertisement, e.g. when renting out two similar flats in the same house with only a short period in between.

In order to give a guideline how to deal with duplicate spells, we have developed a routine to classify duplicate spells as displayed in Table 3 (found in the variable dupID\_gen). The routine to replicate and adjust the above definition is available for Stata and can be obtained by data users from the research data center (FDZ) Ruhr at the RWI. For further details see section 5.

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

# 3 Variable description

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

#### 3.1 Identifier

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

# 3.2 Time period

# Table 7

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

# 3.3 Object features

| Detailed variable descriptions 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  |  |
|  |  |  |

| Feature           | Description  |
|-------------------|--|
| 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.  |
| Name              |  |
| Data type         | Numeric  |
| Description       | This variable indicates the presence of a fitted kitchen   |
| Label             | Floor on which object is located   |
| Name              | etage  |
| Data type         | Numeric  |
|                   | Apartment-specific variable indicates the floor the apartment is lo-   |
| Description       | cated in.  |
| Label             | Usable as holiday home   |
| Name              | ferienhaus   |
| Data type         | Numeric  |
| Description       | This is a binary indicator for whether the object can be used as a hol-<br>iday home. It is only filled for flats. |
| Label             | Available from   |
| Name              | freiab   |
| Data type         | String   |
| Description       | This variable indicates the date from which the object is available.   |
| Label             | Guest toilet in object   |
| Name<br>Data tura | gaestewc   |
| Data type         | Numeric  |
| Description       | (Chared) garden queilable  |
| Name              | (Sildieu) galuell avallable  |
| Data type         | Numeric  |
| Description       | This variable indicates the presence of a garden   |
| Label             | Pets allowed   |
| Name              | haustier erlaubt   |
| Data type         | Numeric  |
| Description       | This binary indicator shows whether pets are allowed in the object.  |
| Label             | House type   |
| Name              | kategorie_Haus   |
| Data type         | Numeric  |

| Feature     | Description  |
|-------------|--|
| Description | The artificial category number indicates which object category a property belongs to. Each property is assigned exactly one category number. This variable is filled for houses only.  |
| Label       | Flat type  |
| Name        | kategorie_Wohnung  |
| Data type   | Numeric  |
| Description | The artificial category number indicates which object category a property belongs to. Each property is assigned exactly one category number. This variable is filled for flats 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   | Numeric  |
| Description | This is a binary variable indicating whether a parking space is available.   |
| Label       | Number of rooms  |
| Name        | zimmeranzahl   |
| Data type   | Numeric  |
| Description | Number of rooms, excluding kitchen, bath or corridors. In several cases, "zimmeranzahl" is not a natural number, which is not necessarily due to a faulty entry. In Germany there is the concept of half rooms. Following the DIN 283 norm, a half room is defined as a room with a size between 6 and 10 square meters. While this definition is outdated, it is still frequently in use. |

# 3.4 General Object Information

| Detailed Variable Descriptions of general object 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 object is still under construction.<br>Missings likely indicate that the object is not under construction. |  |
| Label  | Assisted living for the elderly  |  |
| Name   | betreut  |  |

| Feature     | Description  |
|-------------|--|
| Data type   | Numeric  |
| Description | This indicates whether the property is part of an assisted living for the elderly-programme.                             |
| 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 ob-  |
| Description | ject is rented out. This is on a monthly basis and in EUR rounded to   |
|             | two decimal digits.  |
| Label       | Number of ancillary rooms  |
| Name        | nebenraeume  |
| Data type   | Numeric  |
| Description | This is the number of ancillary rooms.   |
| Label       | Accessible, no steps   |
| Name        | rollstuhlgerecht   |
| Data type   | Numeric  |
| Description | This is a binary indicator for step-free access of the object.   |
| Label       | Number of bedrooms   |
| Name        | schlafzimmer   |
| Data type   | Numeric  |
| Description | This is the number of bedrooms of the object.  |
| Label       | Common charge for community association in EUR/month   |

| Feature     | Description  |
|-------------|--|
| Name        | wohngeld   |
| Data type   | Numeric  |
| Description | This variable refers to the amount of the common charge for com-<br>munity association in EUR per month. |

# 3.5 Area Information

Table 10

| <b>Detailed Variable De</b> | scriptions 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.<br>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<br>natural numbers and numbers with two decimal places. If users enter<br>more than two decimal places, numbers are rounded to two decimal<br>places. |

# 3.6 Energy and Structure Information

Table 11

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   |

| Feature           | Description   |
|-------------------|---|
| Data type         | Numeric   |
| Description       | The Energy Efficiency rating is represented here as an artificially cre-<br>ated 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       | sumption was included in the calculation of the energy consumption value.   |
| Label             | Heating costs   |
| Name              | heizkosten  |
| Data type         | Numeric   |
| Description       | This indicates the monthly heating costs in EUR and is rounded to two decimal digits.   |
| Label             | Type of heating   |
| Name              | heizungsart   |
| Data type         | Numeric   |
| Description       | This is an artificially created category number indicating the type of heating.   |
| Label             | Type of firing  |
| Name              | Bef1-bef10  |
| Data type         | Numeric   |
|                   | This is an artificially created category number indicating the type of  |
| Description       | firing of the heating system. The information is in bef1. If there is a   |
|                   | second type of firing it is in bet2   |
| Label             | Year of last modernisation of object  |
| Name<br>Data tuna | letzte_modernisierung   |
|                   |   |
| Description       | Indicator for the year of the last modernisation.   |
| Label             | Condition of object   |
| Name              | objektzustand   |
| рага туре         |   |
| Description       | The artificial condition number indicates the condition of a property.<br>Each property is assigned exactly one out of 11 possible numbers. |

# 3.7 Price Information

**Detailed Variable Descriptions of Price Information** 

| Feature     | Description  |
|-------------|--|
| Label       | Brokerage at contract conclusion   |
| Name        | courtage   |
| Data type   | String   |
| Description | Brokerage to be paid to the agent.   |
| Label       | Heating costs covered by inclusive rent  |
| Name        | heizkosten_in_wm_enthalten   |
| Data type   | Numeric  |
| Description | This is a binary indicator showing whether heating costs are included in the inclusive rent. |
| Label       | Purchasing price in EUR  |
| Name        | kaufpreis  |
| Data type   | Numeric  |
|             | Price at which the owner advertises to sell the object. Prices are ex-                       |
| Description | pressed in EUR and rounded to two decimal digits. Only objects of                            |
| Labol       | Tered for sale.  |
| Name        | mietekalt  |
| Data type   | Numeric  |
| Dutu type   | Drive studiek the surrow is willing to wort out the shirt. The wort                          |
|             | covers expenses for the living space only. Amenities as well as ex-                          |
| Description | penses for heating or fees such as for garbage disposal are not in-                          |
|             | cluded. Prices are expressed in EUR and rounded to two decimal dig-                          |
|             | its. This is only filled for objects offered for rent.                                       |
| Label       | Security deposit   |
| Name        | mietekaution   |
| Data type   | String   |
|             | This specifies the security deposit to be paid before renting out the                        |
| Description | object. This can either be specified in EUR, however it is also common                       |
|             | to refer to this in terms of base rent (KM/Kaltmiete/Monatsmiete) or                         |
|             | inclusive rent (WW) warmmiete).  |
| Label       | Inclusive rent in EUR  |
| Name        | mietewarm<br>Nume or in  |
| Data type   | Numeric  |
| Description | This refers to the rent inclusive of certain bills, most commonly wa-                        |
| Description | in FLIB and rounded to two decimal digits  |
| Labol       |  |
| Name        | nehenkosten  |
| Data type   | Numeric  |
| Data type   | Humeno   |

| Feature                    | Description  |
|----------------------------|--|
| Description                | This refers to the extra monthly costs that need to be paid for bills<br>on top of the base rent. This usually does not include electricity, but<br>other bills. Prices are expressed in EUR and rounded to two decimal<br>digits. |
| Label<br>Name<br>Data type | Price of parking space in EUR<br>parkplatzpreis<br>Numeric   |
| Description                | This refers to the price of the parking place, expressed in EUR.   |

# 3.8 Regional information

|  | Detailed Variable | Descriptions of | f Regional Information |  |
|--|-------------------|-----------------|------------------------|--|
|--|-------------------|-----------------|------------------------|--|

| Feature                    | Description   |
|----------------------------|---|
| Label                      | German state  |
| Name                       | blid  |
| Data type                  | Numeric   |
| Description                | Each German federal state is attributed to a specific number. For<br>each object, this number identifies the state it is located in. The state<br>variable is obtained from the AGS code retrieved by our georeferenc-<br>ing procedure, i.e. it is based on the variable kid2015. If we failed to<br>geocode an object, the variable reports the state retrieved by Im-<br>moscout's internal georeferencing. You can find these cases by filter-<br>ing to those observations where there is no AGS identifier. |
| Label                      | Local labour market (Kosfeld and Werner, 2012)  |
| Name                       | erg_amd   |
| Data type                  | Numeric   |
| Description                | This variable indicates the local labour market following definitions<br>of Kosfeld and Werner, 2012. Labour market regions are based on<br>commuting behaviour between districts.  |
| Label<br>Name<br>Data type | 1-skm raster cell following INSPIRE<br>ergg_1km<br>String   |
| Description                | This variable indicates the grid cell of a 1-square-km raster of Ger-<br>many according to the INSPIRE guideline. Addresses are matched to<br>this raster based on their geocoded location.   |
| Label<br>Name<br>Data type | Municipality Identifier (AGS, 2015)<br>gid2015<br>Numeric   |
| Description                | This is the municipality identifier according to the German Official Municipality Key (Amtlicher Gemeindeschluessel). It is based on the territoral definition of 2015 (end of year).   |
| Label<br>Name<br>Data type | Municipality Identifier (AGS, 2019)<br>ags2019<br>Numeric   |

| Feature     | Description  |
|-------------|--|
| Description | This is the municipality identifier according to the German Official Municipality Key (Amtlicher Gemeindeschluessel). It is based on the territoral definition of 2019 (end of year).  |
| Label       | Municipality identifier updated by IS24 information  |
| Name        | gid_updated  |
| Data type   | Numeric<br>This identifier is 1 for those observations with no information on the<br>1km <sup>2</sup> grid cell but some information on the municipality provided by<br>ImmobilienScout24 (IS24). In this case the municipality identifier ac- |
| Description | cording to the German Official Municipality Key (Amtlicher Gemein-<br>deschluessel) (see gid2019) is implemented based on IS24 infor-<br>mation.   |
| Label       | District identifier (AGS, 2015)  |
| Name        | kid2015  |
| Data type   | Numeric  |
| Description | This is the district identifier according to the German Official Munic-<br>ipality Key (Amtlicher Gemeindeschluessel). It is based on the terri-<br>torial definition of 2015 (end of year).   |
| Label       | District identifier (AGS. 2019)  |
| Name        | kid2019  |
| Data type   | Numeric  |
| Dutu type   |  |
| Description | ipality Key (Amtlicher Gemeindeschluessel). It is based on the terri-<br>torial definition of 2019 (end of year).  |
| Label       | District identifier updated by IS24 information  |
| Name        | kid_updated  |
| Data type   | Numeric  |
| Description | This identifier is 1 for those observations with no information on the 1km <sup>2</sup> grid cell but some information on the district provided by ImmobilienScout24 (IS24). In this case the district identifier according                    |
| Description | to the German Official Municipality Key (Amtlicher Gemein-<br>deschluessel) (see kid2019) is implemented based on IS24 infor-<br>mation.   |
| Label       | Latitude (UTM) [on-site access only]   |
| Name        | lat_utm  |
| Data type   | Numeric  |
| Description | This is the latitude of the objects according to UTM projection. Access restricted to on-site use only.  |
| Label       | Longitude (UTM) [on-site access only]  |
| Name        | lon_utm  |
| Data type   | Numeric  |
| Description | This is the longitude of the object's address according to UTM pro-<br>jection. Access restricted to on-site use only.   |
| Label       | Address: postcode  |
| Name        | plz  |

| Feature     | Description   |
|-------------|---|
| 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 georeferencing procedure. |

# 3.9 Meta-Information of Advertisement

| <b>Detailed Variable</b> | e Descriptions of Variables Cont | aining 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 of ad  |
| Name              | hits  |
| Data type         | Numeric   |
| Description       | This variable shows the number of hits that the advertisement got.    |
| Label             | Days of availability of ad  |
| Name<br>Data typo | lautzeittage  |
| Data type         |   |
| Description       | This variable indicates the number of days the respective advertise-  |
| Lahal             | Dete of data rational   |
| Laber             | lieforung   |
| Data type         | Numeric   |
| Data type         | This weights above the orea and we with in which the date was an      |
|                   | tracted from the database and delivered to PW/L Advertisements        |
| Description       | which overlap with this cut-off date will hence be split and dupli-   |
|                   | cated. See more under dupID_gen.                                      |
| Label             | Type of offeror   |
| Name              | Anbieter  |
| Data type         | Numeric   |

| Feature                    | Description  |
|----------------------------|--|
| Description                | This variable shows the type of offeror.   |
| 3.10 Generated tech        | nical variables  |
| Table 15                   |  |
| Detailed Variable Descr    | iptions of Technical Variables   |
| Feature                    | Description  |
| Label                      | Classification of object identifiers used more than once   |
| Name                       | dupID_gen  |
| Data type                  | Numeric  |
| Description                | This is a generated variable, indicating whether a duplicate spell is<br>likely to be the same object, or a new object. New objects can have<br>the same ID when customers re-use a previous advertisement for an-<br>other flat, or when an object is re-rented/resold. There is also the<br>possibility to strategically change certain variables. See more under<br>Data preparation/Data problems. |
| Label<br>Name<br>Data type | Spell counter within object identifier<br>spell<br>Numeric   |
| Description                | This is an artificially generated variable indicating the spell within each object ID, should an ID occur more than once. Spells are ordered chronologically.  |
| Label                      | Obid of duplicate  |
| Name                       | duplicateid  |
| Data type                  | Numeric  |
| Description                | This is a variable generated by ImmobilienScout24, indicating which is a duplicate obid that is very likely to be the same object.   |

# 4 Changes between deliveries

ImmobilienScout24 sometimes makes changes to the online questionnaire which means that some new variables become available, and some variables are no longer available. In the following, we give an overview over these changes by real estate type for each variable in the dataset.

# 4.1 Houses for rent

| <b>Changes over</b> | <b>Deliveries:</b> | Houses | for | Rent |
|---------------------|--------------------|--------|-----|------|
|---------------------|--------------------|--------|-----|------|

| Category                         | Variable    | Dez<br>15 | Apr<br>16 | Dez<br>16 | Jun<br>17 | Okt<br>17 | Mrz<br>19 | Jun<br>20 |
|----------------------------------|-------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| General object information       | betreut     | 1         | 1         | 0         | 0         | 0         | 0         | 0         |
|                                  | nebenraeume | 1         | 0         | 0         | 0         | 0         | 0         | 0         |
| Energy and structure information | bef1-bef10  | 0         | 0         | 0         | 0         | 0         | 0         | 1         |
| Meta-information of ad           | anbieter    | 0         | 0         | 0         | 0         | 0         | 0         | 1         |
| Conoral object information       | betreut     | 1         | 1         | 0         | 0         | 0         | 0         | 0         |
| General object information       | nebenraeume | 1         | 0         | 0         | 0         | 0         | 0         | 0         |

# 4.2 Houses for sale

# Table 17

# **Changes over Deliveries: Houses for Sale**

| Category                         | Variable    | Dez | Apr | Dez | Jun | Okt | Mrz | Jun |
|----------------------------------|-------------|-----|-----|-----|-----|-----|-----|-----|
| category                         | variable    | 15  | 16  | 16  | 17  | 17  | 19  | 20  |
| Object features                  | aufzug      | 1   | 1   | 1   | 1   | 1   | 1   | 1   |
|                                  | nebenraeume | 1   | 0   | 0   | 0   | 0   | 0   | 0   |
| Energy and structure information | heizkosten  | 1   | 1   | 1   | 1   | 1   | 1   | 1   |
|                                  | bef1-bef10  | 0   | 0   | 0   | 0   | 0   | 0   | 1   |
| Meta-information of ad           | anbieter    | 0   | 0   | 0   | 0   | 0   | 0   | 1   |
| General object information       | betreut     | 1   | 1   | 0   | 0   | 0   | 0   | 0   |

# 4.3 Flats for Rent

# Table 18

# **Changes over Deliveries: Flats for Rent**

| Catagory                         | Variable      | Dez | Apr | Dez | Jun | Okt | Mrz | Jun |
|----------------------------------|---------------|-----|-----|-----|-----|-----|-----|-----|
| Category                         | Vallable      | 15  | 16  | 16  | 17  | 17  | 19  | 20  |
| Object features                  | denkmalobjekt | 0   | 0   | 0   | 0   | 0   | 0   |     |
|                                  | betreut       | 1   | 1   | 0   | 0   | 0   | 0   |     |
|                                  | nebenraeume   | 1   | 0   | 0   | 0   | 0   | 0   |     |
| Energy and structure information | bef1-bef10    | 0   | 0   | 0   | 0   | 0   | 0   | 1   |
| Meta-information of ad           | anbieter      | 0   | 0   | 0   | 0   | 0   | 0   | 1   |

# 4.4 Flats for Sale

# Table 19

# **Changes over Deliveries: Flats for Sale**

| Category                         | Variable    | Dez | Apr | Dez | Jun | Okt | Mrz | Jun |
|----------------------------------|-------------|-----|-----|-----|-----|-----|-----|-----|
| categoly                         | Variable    | 15  | 16  | 16  | 17  | 17  | 19  | 20  |
| General object information       | betreut     | 1   | 1   | 0   | 0   | 0   | 0   |     |
|                                  | foerderung  | 0   | 0   | 0   | 0   | 0   | 0   |     |
|                                  | nebenraeume | 1   | 0   | 0   | 0   | 0   | 0   |     |
| Energy and structure information | heizkosten  | 1   | 1   | 1   | 1   | 1   | 1   |     |
|                                  | bef1-bef10  | 0   | 0   | 0   | 0   | 0   | 0   | 1   |
| Meta-information of ad           | anbieter    | 0   | 0   | 0   | 0   | 0   | 0   | 1   |

# 5 Further Material

# 5.1 Value Labels

This dataset is fully labelled in both English and German. To review value labels, please download the complementary labels description file from http://www.rwi-essen.de/forschung-undberatung/fdz-ruhr/datenangebot/regionaldaten/rwi-geo-red.

The code for the generation of the variable dupID\_gen can also be downloaded from http://www.rwi-essen.de/forschung-und-beratung/fdz-ruhr/datenangebot/regionaldaten/rwi-geo-red.

# 5.2 Summary Statistics

To get an overview of the data, we provide summary statistics of the variables by real estate type and year. These statistics can be downloaded from <u>http://www.rwi-essen.de/forschung-und-beratung/fdz-ruhr/datenangebot/regionaldaten/rwi-geo-red</u>.

# 6 References

Bauer, T.K., Braun, S. and M. Kvasnicka (2017), Nuclear power plant closures and local housing values: Evidence from Fukushima and the German housing market. *Journal of Urban Economics* 99: 94-106.

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

Boelmann, B. und S. Schaffner (2019), FDZ Data description: Real-Estate Data for Germany (RWI-GEO-RED v1) - Advertisements on the Internet Platform ImmobilienScout24 2007-03/2019. RWI Projektberichte.

Georgi, S. and P. Barkow (2010), Wohnimmobilien-Indizes: Vergleich Deutschland – Großbritannien [Residential real estate indices – A comparison between Germany and the UK]. ZIA Projektbericht.

INSPIRE Thematic Working Group Coordinate reference systems and Geo- graphical grid systems (2014), D2.8.I.2 Data Specification on Geographical Grid Systems – Technical Guidelines, <a href="http://inspire.ec.europa.eu/documents/Data\_Specifications/INSPIRE\_DataSpecifica-tion\_GG\_v3.1.pdf">http://inspire.ec.europa.eu/documents/Data\_Specifications/INSPIRE\_DataSpecifica-tion\_GG\_v3.1.pdf</a>.

Klick, L. and S.Schaffner (2019), Regional Real Estate Price Indices for Germany (RWI-GEO-REDX), RWI Projektberichte.

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

Schaffner, S. (2020), FDZ Data description: Real-Estate Data for Germany Campus Files (RWI-GEO-RED city and RWI-GEO-RED cross) - Advertisements on the Internet Platform ImmobilienScout24 for teaching purposes. RWI Projektberichte.



# Leibniz-Institut für Wirtschaftsforschung

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

Leibniz Gemeinschaft