

Measuring construction activities using advertisements from real estate portals. ESSnet WIN Work Package 3, Use Case 2

WIN conference, Gdansk, February 4<sup>th</sup>-5<sup>th</sup> 2025

Tobias Gramlich (Hesse State Statistical Office, Germany)

**Trusted Smart Statistics – Web Intelligence Network**

Grant Agreement: 101035829



**Web Intelligence**  
Network



**Funded by**  
**the European Union**

# Overall goal

From ads appearing at online real estate portals: can **we** produce some **meaningful early estimate** of the number of **newly constructed buildings** that have **become available** in a **specific year** of reference?

Implicit limitation by the way of data collection:

... become available **on the (online) market** in a specific year ....

**We?** SE-SCB, DE-AfS, DE-HSL



Web Intelligence  
Network



Funded by  
the European Union

**STATISTISCHE ÄMTER DES BUNDES UND DER LÄNDER** | **Neubauatlas**  
Baufertigstellungen von Wohngebäuden

Themen

Jahresauswahl: 2018, 2019, 2020, 2021, 2022

**Wohngebäude**

- Fertiggestellte Wohngebäude nach Kreisen
- Fertiggestellte Wohngebäude nach Gemeinden
- Fertiggestellte Wohngebäude je Tsd. EW nach Gemeinden
- Vorherrschender Haustyp nach Gemeinden
- Vorherrschende Energiequelle der Heizung nach Gemeinden

**Wohnungen**

- Fertiggestellte Wohnungen nach Gemeinden
- Fertiggestellte Wohnungen je Tsd. EW nach Gemeinden

**Wohnfläche**

- Fertiggestellte Wohnfläche nach Gemeinden
- Fertiggestellte Wohnfläche nach Kreisen
- Fertiggestellte Wohnfläche je EW nach Gemeinden

**Kaufwerte**

- Kaufwert nach Kreisen

Legende

**Fertiggestellte Wohngebäude nach Kreisfreien Städten**

Anzahl der fertiggestellten Wohngebäude nach Kreisfreien Städten

Klassen

- unter 100
- 100 bis unter 200
- 200 bis unter 300
- 300 bis unter 450
- 450 oder mehr

Adresse oder Ort suchen

**STATISTISCHE ÄMTER DES BUNDES UND DER LÄNDER** | **Neubauatlas**  
Baufertigstellungen von Wohngebäuden

Themen

Jahresauswahl: 2018, 2019, 2020, 2021, 2022

**Wohngebäude**

- Fertiggestellte Wohngebäude nach Kreisen
- Fertiggestellte Wohngebäude nach Gemeinden
- Fertiggestellte Wohngebäude je Tsd. EW nach Gemeinden
- Vorherrschender Haustyp nach Gemeinden
- Vorherrschende Energiequelle der Heizung nach Gemeinden

**Wohnungen**

- Fertiggestellte Wohnungen nach Gemeinden
- Fertiggestellte Wohnungen je Tsd. EW nach Gemeinden

**Wohnfläche**

- Fertiggestellte Wohnfläche nach Gemeinden
- Fertiggestellte Wohnfläche nach Kreisen
- Fertiggestellte Wohnfläche je EW nach Gemeinden

**Kaufwerte**

- Kaufwert nach Kreisen

Legende

**Fertiggestellte Wohngebäude nach Gemeinden**

Anzahl der fertiggestellten Wohngebäude nach Gemeinden

Klassen

- unter 5
- 5 bis unter 10
- 10 bis unter 15
- 15 bis unter 30
- 30 oder mehr
- Keine

Adresse oder Ort suchen

25	Landkreis	14 918
26	Schwalm-Landkreis	
27	Landkreis	
28	Werra-Landkreis	
29	Reg.-Bez. Mittelhessen	
30	Landkreis	
31	davon kreisfreie Städte	
32	Landkreise	
33	Landkreise	

33	Landkreise	7 912	- 26,5
----	------------	-------	--------

Can we?

Yes, we can!

Well, ...

In principle.

Maybe, we couldn't in every detail.



Web Intelligence  
Network



Funded by  
the European Union



# What do we need?

- ... data
  - Ads from „all“ / several relevant portals, relevant info from each ad
  - Over a longer period of time (three years)
  - Either collect it or get it by means of an agreement
    - Both typically mean you need to spend some money
- ... staff to program and maintain scrapers
  - R and Python
- ... some infrastructure to develop, test and run scrapers, to store data
  - In our case: no production level, so no „production level“ infrastructure
- ... (gain) experience (to) make decisions



# Things to decide, experience to gain

- Which portals to choose?
  - Relevance, accesability, availabbility, reliability, stability, coverage, overlap
  - Collect or buy?
- How to define or identify „newly constructed objects“?
  - Speficiation error, completeness, validity
- When to start scraping? How often to scrape?
  - „Prospective“ scraping, scraping frequency (weekly, daily? Even shorter interval?)
  - Undercoverage
- How to identify duplicates within and between data sources?
  - Completeness, validity, stability

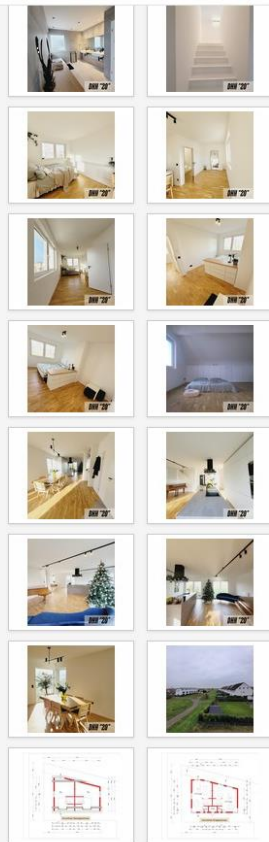


RStudio Workbench

5 Zimmer Doppelhaushälfte in | +

immoexperten.de/immobilie/om-329742\_riedstadt-erstbezug-provisionsfrei-haus-153qm-wohnen-in-parkrandlage-neubaugebiet/

Alle Lesezeichen



Hausauswahl zwei bzw. drei gut geschnittene Schlafzimmer mit schönem Ausblick auf Park und ein Tageslichtbad in Betonoptik modern deckenhoch gefliest mit Badewanne und Dusche, WC und Waschbecken ausgestattet.

Highlight im Dachgeschoss ist das großzügige Studio, sehr hell und luftig durch große Fenster und einem atemberaubenden Ausblick auf Felder. Je nach Ihren persönlichen Bedürfnissen eröffnen sich gleich mehrere Möglichkeiten der Nutzung: als Elternschlafzimmer, Wohlbereich für Ihre Gäste, Kreativstudio oder für produktives Arbeiten im Homeoffice.

Für modernen Komfort sorgen dreifach verglaste Fenster, Fußbodenheizung, elektrische Rollläden im Erdgeschoss, sowie hochwertige Bodenbeläge und Farbe an den Wänden.

Für ein intelligentes Zuhause sorgt ein Smarthome-System mit Lichtsteuerung.

Kfz-Stellplätze gehören ebenso zum Haus und befinden sich nach dem Einfahrtweg zum Grundstück.

Freuen Sie sich auf Ihr Zuhause und das alles erhalten Sie PROVISIONSFREI!

**Angaben zur Ausstattung:**  
Ausstattung Doppelhaushälfte links (Parkstraße 20):  
Wohnfläche 153,18 m<sup>2</sup> (Nutzfläche 156,16 m<sup>2</sup>)

- neuwertiger Zustand Bezug seit 07.2024
- Einbauküche mit Kochinsel seit 08.2024
- Einbauschrank im Dachstudio
- KFZ-Stellplatz
- Bodenheizung
- hochwertiger Eichenparkettboden
- Smarthome Lichtsteuerung
- Tageslichtbad Walk-in-Dusche und Badewanne sowie Doppelwaschtisch
- Kaminofenanschluss
- Gäste-WC
- großer Abstellraum / Garderobe
- 3 große Schlafzimmer
- offener Wohn-, Esszimmerbereich
- Rollläden elektrisch im EG
- Fernwärme

Kaufpreis 629000 Euro

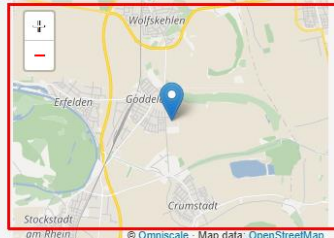
Riedstadt von Feldern und Wiesen umgeben ist. Die direkte Parkrandlage des beschriebenen Anwesens ermöglicht eine ungestörte Aussicht und schafft eine ruhige Atmosphäre. Diese Umgebung eignet sich ideal für Spaziergänge, Fahrradtouren oder entspannte Stunden im Freien.

Gleichzeitig ist die Anbindung an Verkehrswege hervorragend. Riedstadt-Goddelau liegt verkehrsgünstig in der Nähe von Autobahnen und bietet somit eine schnelle Erreichbarkeit größerer Städte in der Region. Dies ermöglicht nicht nur eine unkomplizierte Anfahrt zur Arbeit, sondern auch die Möglichkeit, die kulturellen und wirtschaftlichen Vorzüge der Umgebung zu nutzen.

Auch Kinder werden hier schnell Anschluss finden: Ein Spielplatz und eine Kindertagesstätte liegen nur wenige hundert Meter entfernt. Die örtliche Infrastruktur umfasst auch diverse Einkaufsmöglichkeiten und Freizeiteinrichtungen, was Riedstadt-Goddelau zu einer sozialfreundlichen Gemeinde macht.

Die Nähe zu größeren Städten und die gleichzeitige Ruhe des ländlichen Umfelds machen diese Lage zu einer attraktiven Wahl für Menschen, die das Beste aus beiden Welten suchen.

**Objektadresse**  
Parkstraße 20  
64560 Riedstadt  
Hessen, Deutschland



**Anbieter**  
Becker

Hilfe Preise AGB Nutzungsbedingungen Kontakt Impressum Datenschutzerklärung Partner

immobilienscout24.de/expose/151279388?referrer=RESULT\_LIST\_LISTING&searchId=0c32f99d-8ed3-382d-8508-eb259d613e9&searchUrl=%2Fde/

immobilien Scout24 Suchende EigentümerInnen Inserieren ab 0 €

# KRASS!

Haus kaufen · Hessen · Vogelsbergkreis · Schwalmtal · Exposé

23 Inhalte ansehen 360° Grundriss

## Neubau: Energieeffizientes KfW40-Einfamilienhaus Förderungsanspruch als Musterhaus

Adresse: Helmut-Schaaf-Weg 2, 36318 Schwalmtal, Vogelsbergkreis

Fahrtzeit (z.B. zur Arbeit, Kita...): Lassen Sie sich die Fahrtzeit dieses anderen Orten berechnen.

399.000 € -13% **349.000 €** Kaufpreis 1631 €/m²

Ab 1113 € mtl. finanzieren 6 Zi. 214 m² Wohnfläche

## Neubau: Energieeffizientes KfW40-Einfamilienhaus mit Förderungsanspruch als Musterhaus

Adresse: Helmut-Schaaf-Weg 2, 36318 Schwalmtal, Vogelsbergkreis

Fahrtzeit (z.B. zur Arbeit, Kita...): Lassen Sie sich die Fahrtzeit dieses anderen Orten berechnen.

399.000 € -13% **349.000 €** Kaufpreis 1631 €/m²

Ab 1113 € mtl. finanzieren 6 Zi. 214 m² Wohnfläche 556 m² Grundstücksfläche

immobilienscout24.de/expose/151279388?referrer=RESULT\_LIST\_LISTING&searchId=0c32f99d-8ed3-382d-8508-eb259d613e9&searchUrl=%2Fde/

Grundbucheintrag 0,50 % 1745 €

Gesamtkosten: 389.379 €

PLZ Objekt: 36318

Monatsrate: 1.241 €

Eff. Jahreszins: 3,72 %

Finanzierung starten

### Wer baut mein Traumhaus?

Jetzt Kataloge führender Hausbau-Anbieter bestellen.

100% Kostenlos

Gratis Kataloge anfordern

### Bausubstanz & Energieausweis

Baujahr: 2024	Energieausweis: liegt vor
Bauphase: Haus fertiggestellt	Energieausweistyp: Bedarfsausweis
Objektzustand: Erstbezug	Endenergiebedarf: 16,86 kWh/(m²*a)
Heizungstyp: Fußbodenheizung	Energieeffizienzklasse: A+
Wärmeschutz: -	Baujahr laut Energieausweis: 2024
Energieerzeuger: Umweltwärme	

Maximaler Ertrag - Weltklasse Solarmodule entwickelt in Deutschland

Solar Komplettpaket mit 10 Jahren Versicherung & 25 Jahren Leistungsgarantie. Jetzt kaufen oder flexibel in Raten zahlen - ohne Bankkredit!

powered by ocellis

### Services passend zum Haus

Stromverbrauch

Ihre Postleitzahl: 36318

Anzahl der Personen im Haushalt: 6

Home Entertainment, Stromverbrauch, Finanzierungcheck, Wohnung kündigen

Herr Stefan Heyder  
7 Bewertungen  
Telefonnummer anzeigen

Nachrichte Preis vorschlagen

Gäste-WC: 2

typ: Einfamilienhaus (freistehend) Zimmer: 6

Etageanzahl: 2 Garage/ Stellplatz: 3 Stellplätze

Wohnfläche ca.: 214 m² Internet: Verfügbarkeit prüfen

Grundstück ca.: 556 m²

### Hilfreiche Tipps zu diesem Haus

Kaufen oder doch bauen? Kostenlose Hausbaukataloge

Entdecke den neuen virtuellen Musterhauspark

Finanzieren zu günstigen Zinsen

Was ist mein Haus aktuell wert? Jetzt kostenlos berechnen

Endlich das Traumhaus gefunden? Jetzt fehlt noch MagentaTV mit Internet

Hier sparen: www.telekom.de

### Plus+ Mitglieder wissen mehr!

Erkunde deine Chancen auf diese Immobilie.

Veröffentlicht: Aufrufe:

Gemerkt: Kontaktiert:

Konkurrenz-Analyse

Infos zum Einkommen, Haushaltsgröße...

Jetzt Plus+ aktivieren

### Kosten

Kaufpreis: 349.000 € Provision für Käufer: 3,57 % inkl. ges. MwSt.

Kann ich mir dieses Haus leisten? Immobilienwert: Wertermittlung als KäuferPlus Nutzer inklusive

Preis/m²: 1631 €/m²

Du benötigst eine Baufinanzierung?

Gesponsert

500€ Startguthaben

Jetzt neu: Die Business Platinum Card

Mit noch mehr Vorteilen für dich und dein Business.

Jetzt beantragen

Es gelten Bedingungen.

5 Kostenfallen beim Immobilienkauf

Ob Neubau oder Bestand: Diese typischen Kostenfallen kalkulieren Sie besser ein.

Mehr erfahren



# Examples: projects / deduplication

neubaukompass.de/neubau/hainer-weg-50-50-a-b-frankfurt-am-main/

neubaukompass

Standorte für A...

< Zurück

Region Frankfurt > Frankfurt am Main > Sachsenhausen

BILDER DETAILS MERKMALE LAGE

Frankfurt, Hainer Weg 50

Neubau von 45 Eigentumswohnungen

\* Schlüsselfertig

Projektdetails

ADRESSE  
Hainer Weg 50 und 50a-b, 60599 Frankfurt am Main / Sachsenhausen

PREIS  
889.800 € - 1.337.100 €

Frankfurt, Hainer Weg 50 und 50a-b

neubaukompass.de/neubau/hainer-weg-50-50-a-b-frankfurt-am-main/

BILDER DETAILS MERKMALE LAGE **WOHNBEISPIELE** PROJEKTBESCHREIBUNG ANBIETER

3 Zi. ( 4 Zi. )

	Hainer Weg 50a - 2.2.3	158 IS	WOHNFLÄCHE:	ZIMMERN:	>
		889.800 €	98,6 m <sup>2</sup>	3	
	Hainer Weg 50a - 2.1.3	158 IS	WOHNFLÄCHE:	ZIMMERN:	>
		919.500 €	107,05 m <sup>2</sup>	3	
	Hainer Weg 50a - 2.4.3	158 IS	WOHNFLÄCHE:	ZIMMERN:	>
		921.400 €	98,6 m <sup>2</sup>	3	
	Hainer WEG 50 - 2.1.1	158 IS	WOHNFLÄCHE:	ZIMMERN:	>
		929.900 €	102,85 m <sup>2</sup>	3	
	Hainer Weg 50b - 2.1.5	158 IS	WOHNFLÄCHE:	ZIMMERN:	>
		939.800 €	104,4 m <sup>2</sup>	3	
	Hainer WEG 50 - 2.1.2	158 IS	WOHNFLÄCHE:	ZIMMERN:	>
		944.700 €	106,05 m <sup>2</sup>	3	
	Hainer WEG 50 - 2.2.1	158 IS	WOHNFLÄCHE:	ZIMMERN:	>
		944.700 €	106,05 m <sup>2</sup>	3	

Anbieter kontaktieren



# Example: „prospective“ scraping?

The screenshot shows a search results page on the website immobilienscout24.de. The search criteria are set to 'Hessen' (Hesse) and 'Wohnfläche' (Living Area). The results list several properties with their respective prices and features. The top result is '3 Passende Wohneinheiten in diesem Projekt:' with a price of 399.900 € and 761 m². Below it, another result is '509.900 € 1001 m² 4 ZL.' and '2 Passende Wohneinheiten in diesem Projekt:' with a price of 528.218,50 € and 93,49 m². The page also features a 'Wer baut mein Traumhaus?' section with a call to action 'Jetzt Kataloge führender Hausbau-Anbieter bestellen'.

The screenshot shows a detailed project page for '288 WESTCOURT Neubauprojekt - Sossenheim' on the website immobilienscout24.de. The page features a large 3D architectural rendering of the building complex. Below the rendering, the project name '288 WESTCOURT Neubauprojekt - Sossenheim' is displayed. The price is listed as 'Kaufpreis 271.833,40 € - 622.920 €' with a price per square meter of 'Preis/m² ab 5.300 €'. The page also includes a '50% vermarktet' badge and a 'Bezugstermin Q1 2027' badge. The address is 'Westerbacherstrasse 286-288 65936 Frankfurt Sossenheim'. The page also features a '24 Einheiten 24 Wohnungen' badge and a 'Bezugstermin Q1 2027' badge. The page includes a 'Anbieter kontaktieren' button and a 'Das 288 WESTCOURT vereint scheinbar überwindbare Gegensätze: Urbanität und Nähe zur Natur, langjährige Tradition und top modernen Wohnraum.' text.



Web Intelligence  
Network



Funded by  
the European Union

# Results for DE-HE: 2023

- Comparison of scraped ads vs. official statistics (completed buildings, 2023)
- Overall coverage (ads to buildings): 210%
- Overall coverage (ads to apartments): 70%
- Coverage varies between NUTS3 areas: [45%, 1400%]
- Coverage higher / „better“ in urban areas



# Summary

- Successfully collected data over a longer period of time (up to three years weekly / monthly data)
- ... meaningful number... ? Yes!
- Comparable to official statistics?
  - Only aggregates, no linkage of microdata
  - Different concepts / „specification error“
  - Undercoverage, missing data
  - Duplicates
  - Bias (higher coverage of ads in urban areas than in rural areas)





# Since this is also our **final** WIN conference

**Thank you! Dziękuję bardzo! Tack så mycket! Hartelijk dank! ¡Muchas gracias!**

- ... for several years of very good and very friendly exchange and cooperation
- ... for input, reviewing reports, support, advice, patience
- ... organization of online and F2F meetings, conferences, ...
- ... for financing.



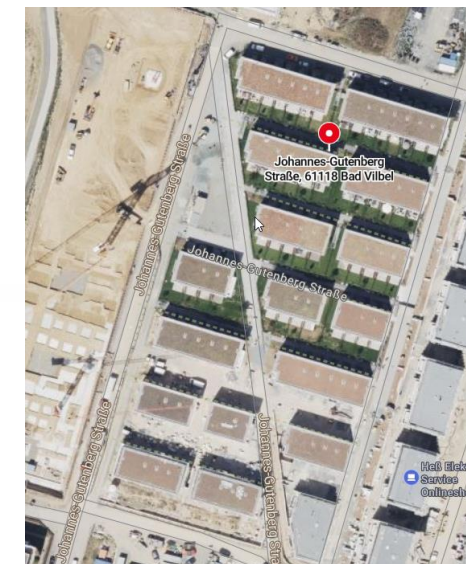
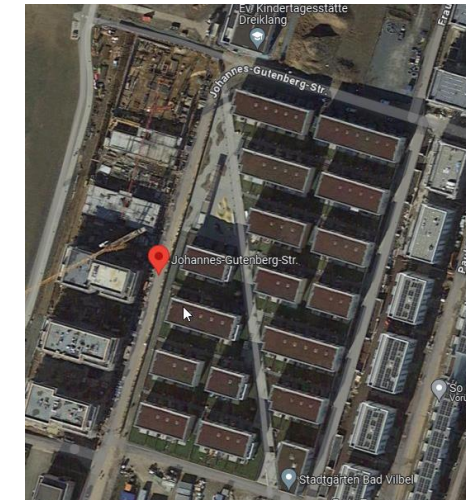
**Web Intelligence**  
Network



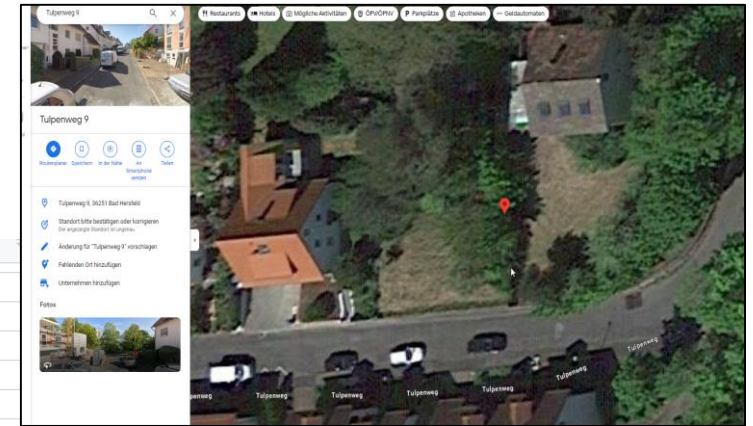
**Funded by**  
**the European Union**

# Challenges: deduplication

	adresse2	ad_provider.x	ad_id.x	surface.x	floor.x	rooms.x	price.x	ad_provider.y	ad_id.y	surface.y	floor.y	rooms.y	price.y	quelle	strasse
251	Bad Vilbel - 61118 - friedberger - 77 - NA - 5 - 171.74	NA	NA	NA	NA	NA	NA	Immoscout	1926	171.74	NA	5	1450000	nur Immoscout	TRUE
252	Bad Vilbel - 61118 - johannesgutenberg - 07 - 1 - 2 - 64	NA	NA	NA	NA	NA	NA	Immoscout	5313	64.00	1	2	1090	nur Immoscout	TRUE
253	Bad Vilbel - 61118 - johannesgutenberg - 1 - 2 - 2 - 63	ImmoWelt	27Y6R56	63.00	2	2	NA	NA	NA	NA	NA	NA	NA	nur ImmoWelt	TRUE
254	Bad Vilbel - 61118 - johannesgutenberg - 1 - 3 - 3 - 93	NA	NA	NA	NA	NA	NA	Immoscout	823	93.00	3	3	1400	nur Immoscout	TRUE
255	Bad Vilbel - 61118 - johannesgutenberg - 11 - 3 - 3 - 93	NA	NA	NA	NA	NA	NA	Immoscout	8258	93.00	3	3	1500	nur Immoscout	TRUE
256	Bad Vilbel - 61118 - johannesgutenberg - 11 - 3 - 4 - 112.9	NA	NA	NA	NA	NA	NA	Immoscout	2142	112.90	3	4	1860	nur Immoscout	TRUE
257	Bad Vilbel - 61118 - johannesgutenberg - 3 - 1 - 2 - 61	NA	NA	NA	NA	NA	NA	Immoscout	3081	61.00	1	2	1020	nur Immoscout	TRUE
258	Bad Vilbel - 61118 - johannesgutenberg - 3 - 1 - 2 - 61.2	NA	NA	NA	NA	NA	NA	Immoscout	3681	61.20	1	2	980	nur Immoscout	TRUE
259	Bad Vilbel - 61118 - johannesgutenberg - 3 - 1 - 3 - 84.7	NA	NA	NA	NA	NA	NA	Immoscout	3253	84.70	1	3	1420	nur Immoscout	TRUE
260	Bad Vilbel - 61118 - johannesgutenberg - 3 - 2 - 2 - 62	NA	NA	NA	NA	NA	NA	Immoscout	896	62.00	2	2	975	nur Immoscout	TRUE
261	Bad Vilbel - 61118 - johannesgutenberg - 3 - 2 - 3 - 83.73	NA	NA	NA	NA	NA	NA	Immoscout	4763	83.73	2	3	1260	nur Immoscout	TRUE
262	Bad Vilbel - 61118 - johannesgutenberg - 3 - 2 - 3 - 84.7	NA	NA	NA	NA	NA	NA	Immoscout	5365	84.70	2	3	1355	nur Immoscout	TRUE
263	Bad Vilbel - 61118 - johannesgutenberg - 3 - 4 - 2 - 56.8	NA	NA	NA	NA	NA	NA	Immoscout	2067	56.80	4	2	960	nur Immoscout	TRUE
264	Bad Vilbel - 61118 - johannesgutenberg - 3 - 4 - 2 - 56.8	NA	NA	NA	NA	NA	NA	Immoscout	1389	56.80	4	2	960	nur Immoscout	TRUE
265	Bad Vilbel - 61118 - johannesgutenberg - 5 - 3 - 2 - 62.1	NA	NA	NA	NA	NA	NA	Immoscout	2798	62.10	3	2	950	nur Immoscout	TRUE
266	Bad Vilbel - 61118 - johannesgutenberg - 7 - 0 - 2 - 64	NA	NA	NA	NA	NA	NA	Immoscout	4508	64.00	0	2	1090	nur Immoscout	TRUE
267	Bad Vilbel - 61118 - johannesgutenberg - 7 - 2 - 2 - 63.4	NA	NA	NA	NA	NA	NA	Immoscout	1304	63.40	2	2	1005	nur Immoscout	TRUE
268	Bad Vilbel - 61118 - johannesgutenberg - 7 - 4 - 2 - 58	NA	NA	NA	NA	NA	NA	Immoscout	6001	58.00	4	2	970	nur Immoscout	TRUE
269	Bad Vilbel - 61118 - johannesgutenberg - 7 - 4 - 3 - 99	NA	NA	NA	NA	NA	NA	Immoscout	5517	99.00	4	3	1680	nur Immoscout	TRUE
270	Bad Vilbel - 61118 - johannesgutenberg - 7 - 4 - 3 - 99	NA	NA	NA	NA	NA	NA	Immoscout	8277	99.00	4	3	1520	nur Immoscout	TRUE
271	Bad Vilbel - 61118 - johannesgutenberg - 7 - 4 - 3 - 99	NA	NA	NA	NA	NA	NA	Immoscout	7831	99.00	4	3	1680	nur Immoscout	TRUE
272	Bad Vilbel - 61118 - johannesgutenberg - 9 - 0 - 2 - 61.8	ImmoWelt	27MK75N	61.80	0	2	NA	NA	NA	NA	NA	NA	NA	nur ImmoWelt	TRUE
273	Bad Vilbel - 61118 - johannesgutenberg - 9 - 0 - 2 - 62	NA	NA	NA	NA	NA	NA	Immoscout	6992	62.00	0	2	975	nur Immoscout	TRUE
274	Bad Vilbel - 61118 - johannesgutenberg - 9 - 1 - 2 - 61.1	NA	NA	NA	NA	NA	NA	Immoscout	352	61.10	1	2	945	nur Immoscout	TRUE
275	Bad Vilbel - 61118 - johannesgutenberg - 9 - 2 - 2 - 61	ImmoWelt	27LVT5T	61.00	2	2	NA	Immoscout	7992	61.00	2	2	1000	beide Portale	TRUE
276	Bad Vilbel - 61118 - johannesgutenberg - 9 - 2 - 3 - 83.8	NA	NA	NA	NA	NA	NA	Immoscout	7611	83.80	2	3	1400	nur Immoscout	TRUE
277	Bad Vilbel - 61118 - johannesgutenberg - 9 - 2 - 3 - 86	NA	NA	NA	NA	NA	NA	Immoscout	731	86.00	2	3	1500	nur Immoscout	TRUE
278	Bad Vilbel - 61118 - johannesgutenberg - 9 - NA - 2 - 61.8	NA	NA	NA	NA	NA	NA	Immoscout	78	61.80	NA	2	990	nur Immoscout	TRUE
279	Bad Vilbel - 61118 - kurtmoosdorf - 52 - NA - 6 - 171	NA	NA	NA	NA	NA	NA	Immoscout	3449	171.00	NA	6	1094250	nur Immoscout	TRUE



# Challenges: deduplication



	adresse2	ad_provider.x	ad_id.x	surface.x	floor.x	rooms.x	price_rent.x	price_sell.x	ad_provider.y	ad_id.y	surface.y	floor.y	rooms.y	price_rent.y	price_sell.y
5	Bad Hersfeld - 36251 - tulpen - 9 - 1 - 101	Immowelt	26Q285G	101.0	1	4	NA	329900	Immowelt	26N9E5A	101.0	1	4	NA	319900
6	Bad Hersfeld - 36251 - tulpen - 9 - 1 - 101	Immowelt	26Q285G	101.0	1	4	NA	329900	Immowelt	2633755	101.0	1	4	NA	329900
7	Bad Hersfeld - 36251 - tulpen - 9 - 1 - 101	Immowelt	26Q285G	101.0	1	4	NA	329900	Immowelt	26DJ85H	101.0	1	4	NA	319900
8	Bad Hersfeld - 36251 - tulpen - 9 - 1 - 101	Immowelt	26Q285G	101.0	1	4	NA	329900	Immowelt	27GUP57	101.0	1	4	NA	324900
9	Bad Hersfeld - 36251 - tulpen - 9 - 1 - 101	Immowelt	26N9E5A	101.0	1	4	NA	319900	Immowelt	26Q285G	101.0	1	4	NA	329900
0	Bad Hersfeld - 36251 - tulpen - 9 - 1 - 101	Immowelt	26N9E5A	101.0	1	4	NA	319900	Immowelt	2633755	101.0	1	4	NA	329900
1	Bad Hersfeld - 36251 - tulpen - 9 - 1 - 101	Immowelt	26N9E5A	101.0	1	4	NA	319900	Immowelt	26DJ85H	101.0	1	4	NA	319900
2	Bad Hersfeld - 36251 - tulpen - 9 - 1 - 101	Immowelt	26N9E5A	101.0	1	4	NA	319900	Immowelt	27GUP57	101.0	1	4	NA	324900
3	Bad Hersfeld - 36251 - tulpen - 9 - 1 - 101	Immowelt	2633755	101.0	1	4	NA	329900	Immowelt	26Q285G	101.0	1	4	NA	329900
4	Bad Hersfeld - 36251 - tulpen - 9 - 1 - 101	Immowelt	2633755	101.0	1	4	NA	329900	Immowelt	26N9E5A	101.0	1	4	NA	319900
5	Bad Hersfeld - 36251 - tulpen - 9 - 1 - 101	Immowelt	2633755	101.0	1	4	NA	329900	Immowelt	26DJ85H	101.0	1	4	NA	319900
6	Bad Hersfeld - 36251 - tulpen - 9 - 1 - 101	Immowelt	2633755	101.0	1	4	NA	329900	Immowelt	27GUP57	101.0	1	4	NA	324900
7	Bad Hersfeld - 36251 - tulpen - 9 - 1 - 101	Immowelt	26DJ85H	101.0	1	4	NA	319900	Immowelt	26Q285G	101.0	1	4	NA	329900
8	Bad Hersfeld - 36251 - tulpen - 9 - 1 - 101	Immowelt	26DJ85H	101.0	1	4	NA	319900	Immowelt	26N9E5A	101.0	1	4	NA	319900
9	Bad Hersfeld - 36251 - tulpen - 9 - 1 - 101	Immowelt	26DJ85H	101.0	1	4	NA	319900	Immowelt	2633755	101.0	1	4	NA	329900
0	Bad Hersfeld - 36251 - tulpen - 9 - 1 - 101	Immowelt	26DJ85H	101.0	1	4	NA	319900	Immowelt	27GUP57	101.0	1	4	NA	324900
1	Bad Hersfeld - 36251 - tulpen - 9 - 1 - 101	Immowelt	27GUP57	101.0	1	4	NA	324900	Immowelt	26Q285G	101.0	1	4	NA	329900
2	Bad Hersfeld - 36251 - tulpen - 9 - 1 - 101	Immowelt	27GUP57	101.0	1	4	NA	324900	Immowelt	26N9E5A	101.0	1	4	NA	319900
3	Bad Hersfeld - 36251 - tulpen - 9 - 1 - 101	Immowelt	27GUP57	101.0	1	4	NA	324900	Immowelt	2633755	101.0	1	4	NA	329900
4	Bad Hersfeld - 36251 - tulpen - 9 - 1 - 101	Immowelt	27GUP57	101.0	1	4	NA	324900	Immowelt	26DJ85H	101.0	1	4	NA	319900
5	Bad Zwesten - 34596 - kasseler - 9 - 1 - 78	Immowelt	25X895H	78.0	1	3	890	700	Immowelt	2545H5F	78.0	1	3	890	700



Web Intelligence  
Network



Funded by  
the European Union



# Results for DE-HE (I): 2022

- Comparison of scraped ads vs. official statistics (completed buildings, 2022)
- Overall coverage (ads to buildings): 130%
- Overall coverage (ads to apartments): 46%
- Coverage varies between NUTS3 areas: [10%,1000%]
- Coverage higher / „better“ in urban areas





# Results from DE-Hesse (for 2022)

NUTS3 Name	NUTS3 Code	No of Buildings	No of Apartments	No of Ads	No of Ads (#4)	No of Addresses	Share of Buildings by Ads	Share of Apartments by Ads	Share of Buildings by Addresses	Share of Apartments by Addresses
Darmstadt, Wissenschaftsstadt	06411	39	195	213	139	85	546,2	109,2	217,3	43,6
Frankfurt am Main, Stadt	06412	261	2498	1645	937	266	630,3	65,9	101,9	10,6
Offenbach am Main, Stadt	06413	81	812	852	473	134	1051,9	104,9	165,4	16,5
Wiesbaden, Landeshauptstadt	06414	377	1839	570	418	195	151,2	31,0	51,7	10,6
Bergstraße	06431	380	814	222	125	108	58,4	27,3	28,4	13,3
Darmstadt-Dieburg	06432	351	670	272	172	103	77,5	40,6	29,3	15,4
Groß-Gerau	06433	295	1049	368	268	124	124,7	35,1	42,0	11,8
Hochtaunuskreis	06434	267	604	312	188	113	116,9	51,7	42,3	18,7
Main-Kinzig-Kreis	06435	604	1565	581	310	178	96,2	37,1	29,5	11,4
Main-Taunus-Kreis	06436	214	678	420	310	133	196,3	61,9	62,1	19,6
Odenwaldkreis	06437	112	227	57	26	28	50,9	25,1	25,0	12,3
Offenbach	06438	151	846	637	416	193	421,9	75,3	127,8	22,8
Rheingau-Taunus-Kreis	06439	185	538	226	189	67	122,2	42,0	36,2	12,5
Wetteraukreis	06440	464	1242	588	222	131	126,7	47,3	28,2	10,5
Gießen	06531	284	758	208	123	87	73,2	27,4	30,6	11,5
Lahn-Dill-Kreis	06532	261	526	234	131	75	89,7	44,5	28,7	14,3
Limburg-Weilburg	06533	319	568	158	79	52	49,5	27,8	16,3	9,2
Marburg-Biedenkopf	06534	287	477	224	98	93	78,0	47,0	32,4	19,5
Vogelsbergkreis	06535	144	158	14	7	22	9,7	8,9	15,3	13,9
Kassel, documenta-Stadt	06611	56	286	171	33	28	305,4	59,8	50,0	9,8
Fulda	06631	331	522	170	91	60	51,4	32,6	18,1	11,5
Hersfeld-Rotenburg	06632	90	117	50	15	33	55,6	42,7	36,7	28,2
Kassel	06633	312	570	121	5	48	38,8	21,2	15,4	8,4
Schwalm-Eder-Kreis	06634	250	396	91	17	53	36,4	23,0	21,2	13,4
Waldeck-Frankenberg	06635	315	455	94	15	38	29,8	20,7	12,1	8,4
Werra-Meißner-Kreis	06636	109	121	38	1	23	34,9	31,4	21,1	19,0
HESSEN	06	6539	18531	8536	4808	2470	130,5	46,1	37,8	13,3