

empirica-Preisdatenbank

Datensatzbeschreibung „Wohnen“



empirica

Impressum

empirica ag
Büro: Berlin
Kurfürstendamm 234, 10719 Berlin
Telefon (030) 88 47 95-0
berlin@empirica-institut.de
www.empirica-institut.de

Bearbeitung

Jan Grade und Dr. Reiner Braun

Projektnummer

11017

Berlin, im August 2024

INHALTSVERZEICHNIS

1.	Kurzfassung.....	1
1.1	Datenqualität.....	1
1.2	Validität, Repräsentativität und Transparenz	1
1.3	Copyright	1
2.	Ausführliche Datensatzbeschreibung.....	2
2.1	Besonderheit	2
2.2	Validität	2
2.3	Repräsentativität.....	3
2.4	Transparenz und Datenschutz.....	3
2.5	Copyright	3

1. Kurzfassung

Die **empirica-Preisdatenbank** reicht zurück bis ins Jahr 2004 und ist die mit Abstand größte Sammlung von Immobilieninseraten für Deutschland. Seit 2012 stützen wir unsere Analysen daher auf die VALUE Marktdatenbank (die bis 2021 empirica-systeme Marktdatenbank hieß). Diese Datenbank bietet eine stichtagsunabhängige Zufallsstichprobe mit professioneller Dopplerbereinigung (im Quer- und Längsschnitt) und Experten-gestützter Plausibilitätsprüfung. Mit über 30-jähriger Erfahrung in der immobilienwirtschaftlichen Beratung von z.B. Kommunen, Banken und Investoren garantieren wir damit eine bisher einmalige Qualitätssicherung. Damit sind wir mehr als ein reiner Datenbankanbieter.

1.1 Datenqualität

Die **empirica-Preisdatenbank** stützt sich auf einen ausgewogenen Mix an Datenquellen verschiedenster Vertriebskanäle. Neben den großen Immobilienmarktplätzen fließen auch spezialisierte Internetquellen und Printmedien ein. Die Daten werden kontinuierlich, deutschlandweit und flächendeckend recherchiert, im Querschnitt (über alle Quellen) und im Längsschnitt (über die Zeit) professionell um Doppler bereinigt sowie um weitere amtliche und nicht-amtliche Statistiken ergänzt. So werden die notwendige Konsistenz, eine sonst unerreichbare regionale Abdeckung sowie ein hoher Besetzungsgrad der Angaben zu Wohnwertmerkmalen sichergestellt.

1.2 Validität, Repräsentativität und Transparenz

Die **empirica-Preisdatenbank** ist valide, weil systematische Fehler durch eine Experten-gestützte Plausibilitätsprüfung ausgeschlossen werden. Unsere Stichprobe ist repräsentativ, weil es sich um eine Zufallsauswahl handelt. Aufgrund der kontinuierlichen (nicht-stichtagsbezogenen) Stichprobenziehung gilt dies bei uns zusätzlich im Hinblick auf die Erfassungszeitpunkte. Insbesondere werden dadurch auch sehr kurzzeitig eingestellte Angebote erfasst. Die Erhebung und Verarbeitung der Daten wird dokumentiert und damit höchste Transparenz gewährleistet.

1.3 Copyright

Die Daten der empirica-Preisdatenbank bleiben geistiges Eigentum der empirica ag. Der Erwerb berechtigt nur zum Gebrauch innerhalb Ihrer Firma. Eine Veröffentlichung oder Weitergabe der Daten an Dritte (auch verbundene Unternehmen) – ob in Originalform oder weiterverarbeiteter Form, ob entgeltlich, unentgeltlich oder im Tausch – bedürfen der Zustimmung der empirica ag und erfordern zwingend die Quellenangabe „**empirica-Preisdatenbank** (Basis: VALUE Marktdaten / bis 2011 IDN ImmoDaten GmbH)“.

2. Ausführliche Datensatzbeschreibung

Im Januar 2012 wurde die **empirica-Preisdatenbank** von Rohdaten der IDN ImmoDaten GmbH auf die VALUE Marktdatenbank umgestellt. Die resultierende Preisdatensammlung vereint unser immobilienwirtschaftliches Know-how mit dem für die Erhebung und automatisierte Verarbeitung notwendigen technischen Know-how der VALUE. Dies ermöglicht unter methodischen und inhaltlichen Aspekten eine weitere Qualitätssteigerung und die vollständige Transparenz bei allen Arbeitsschritten.

Die **empirica-Preisdatenbank** (Basis: VALUE Marktdaten; bis 2011: IDN ImmoDaten GmbH) speist sich aus kontinuierlich, deutschlandweit und flächendeckend recherchierten Immobilieninseraten (Kaufangebote, Mietangebote) und umfasst detaillierte Informationen zu den Objekten (Größe, Qualität, Zustand usw.), deren Standort (Gemeinde, Postleitzahlbezirk, Adresse), der Preisvorstellung des Anbieters und dem Erscheinungsdatum.

Bei der Recherche fließen neben den bekannten großen Immobilienmarktplätzen auch kleinere, spezialisierte Internetquellen (z.B. für Privatanbieter, Genossenschaften oder Wohngemeinschaften) und Printmedien (Kleinanzeigen lokaler, regionaler und überregionaler Zeitungen) ein. Aufgrund des ausgewogenen Mix an berücksichtigten Vertriebskanälen werden Angebote hochwertiger Neubauwohnungen ebenso erfasst wie unsanierte Altbaubestände, Angebote privater Vermieter ebenso wie Angebote von Wohnungsunternehmen und Angebote von freifinanzierten Wohnungen ebenso wie von preisgebundenen Sozialwohnungen (Anzeigen mit dem Hinweis „WBS erforderlich“ o.Ä. im Freitext).

Seit 2004 werden täglich neue Daten in die Datenbank eingelesen, seit 2012 aus der VALUE Marktdatenbank. Im Laufe der Jahre ist die mit Abstand größte Sammlung von Immobilieninseraten aus Deutschland entstanden. Durch ergänzende Daten (z.B. Adressregister) und amtliche Statistiken (z.B. Baufertigstellungen) wird die notwendige Konsistenz der Datengrundlage sichergestellt.

2.1 Besonderheit

Ein Qualitätsmerkmal ist die professionelle Dopplerbereinigung im Querschnitt (über alle Internetquellen und Printmedien) und Längsschnitt (über den gesamten Angebotszeitraum): Immobilien, die über einen längeren Zeitraum im Angebot stehen, werden nur einmal in die Datenbank aufgenommen. Im Falle von Anpassungen der Inserate, findet der im jeweils betrachteten Zeitraum zuletzt genannten Preis bzw. die am häufigsten genannten Objektmerkmale Berücksichtigung.

2.2 Validität

Eine Stichprobe ist valide, wenn systematische Fehler ausgeschlossen werden können. In unserer Stichprobe werden invalide oder widersprüchliche Aussagen durch eine Experten-gestützte Plausibilitätsprüfung vermieden. Anbieterangaben aus Freitext und Angebotsformular werden gegenübergestellt und ggf. korrigiert (z.B. Nicht-Wohnungsangebote wie Büros oder Gewerberäume, die in der Kategorie „Wohnungen“ einsortiert wurden), unplausible oder preisverzerrende Beobachtungen herausgefiltert (z.B. untypisch große/kleine Wohnflächen, Angabe der Brutto- anstelle der Nettokaltmiete, Fertighäuser ohne Grundstücke etc.) und fehlende Angaben ergänzt (z.B. aktuelle Gemeindekennziffern oder Stadtteile). Nur durch eine derart umfangreiche Aufbereitung kann die hohe

Datenqualität bezogen auf Validität, Vollständigkeit der Objektmerkmale und räumliche Verortung garantiert werden.

2.3 Repräsentativität

Eine Stichprobe ist repräsentativ, wenn es sich um eine Zufallsauswahl handelt und die Merkmale in der Stichprobe in derselben Häufigkeit vorkommen wie in der Grundgesamtheit. Unserer Stichprobe zeichnet sich vor allem durch eine ausgewogene Berücksichtigung aller Immobilienmarktakteure (privat, gewerblich, öffentlich gefördert etc.) und aller öffentlich verfügbaren Quellen (Internet- und Printmedien) aus, bei der kein Marktsegment systematisch vernachlässigt wird. Zudem erhebt die VALUE AG (im Unterschied zu anderen Anbietern) tatsächlich eine Zufallsstichprobe mit Bezug auf die Erfassungszeitpunkte. Dadurch werden (anders als bei stichtagsbezogener Datenabfrage) auch sehr kurzzeitig eingestellten Angeboten erfasst. Die **empirica-Preisdatenbank** ist damit eine repräsentative Stichprobe der öffentlich inserierten Immobilien.

2.4 Transparenz und Datenschutz

In die **empirica-Preisdatenbank** (Basis: VALUE Marktdaten) fließen ausschließlich allgemein zugängliche und mit dem Ziel der Verbreitung veröffentlichte Daten ein. Diese Daten haben keinen direkten Personenbezug und dienen ausschließlich Zwecken der Immobilienmarktforschung. Erhoben werden die zur Beschreibung einer Immobilie relevanten Informationen und diese lediglich im Umfang einer repräsentativen Stichprobe aller öffentlich inserierten Immobilien. Die Erhebung und Verarbeitung der Daten wird dokumentiert, wodurch eine vollständige Transparenz gewährleistet ist.

2.5 Copyright

Die Daten der empirica-Preisdatenbank bleiben geistiges Eigentum der empirica ag. Der Erwerb berechtigt nur zum Gebrauch innerhalb Ihrer Firma. Eine Veröffentlichung oder Weitergabe der Daten an Dritte (auch verbundene Unternehmen) – ob in Originalform oder weiterverarbeiteter Form, ob entgeltlich, unentgeltlich oder im Tausch – bedürfen der Zustimmung der empirica ag und erfordern zwingend die Quellenangabe „**empirica-Preisdatenbank** (Basis: VALUE Marktdaten / bis 2011 IDN ImmoDaten GmbH)“.