

Neue Datenquellen und Methoden für die Stadtbeobachtung

Link, Tobias

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Empfohlene Zitierung / Suggested Citation:

Link, T. (2023). Neue Datenquellen und Methoden für die Stadtbeobachtung. *Stadtforschung und Statistik : Zeitschrift des Verbandes Deutscher Städtestatistiker*, 36(1), 72-78. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-86677-0>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY-NC-SA Lizenz (Namensnennung-Nicht-kommerziell-Weitergabe unter gleichen Bedingungen) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier: <https://creativecommons.org/licenses/by-nc-sa/4.0/deed.de>

Terms of use:

This document is made available under a CC BY-NC-SA Licence (Attribution-NonCommercial-ShareAlike). For more information see: <https://creativecommons.org/licenses/by-nc-sa/4.0>

Tobias Link

Neue Datenquellen und Methoden für die Stadtbeobachtung

Vor dem Hintergrund der Expansion digitaler Technik, die fast alle Lebensbereiche umfasst, sind als Folge oder Ziel dieser neuen Prozesse Datenakkumulationen entstanden, die es in diesem Umfang bisher nicht gab. Viele Unternehmen bzw. Gemeinschaften haben digitale Plattformen erschaffen, die mehr oder weniger themenspezifisch zugeschnitten sind und deren Zweck darin liegt, Menschen untereinander zu vernetzen – entweder um analoge Transaktionen effizienter und einfacher zu gestalten oder um gemeinschaftlich eine solche Datensammlung zu betreiben. Der Zugriff auf solche Datensammlungen soll an dieser Stelle beispielhaft mit Auswertungen von Daten aus OpenStreetMap, Immobilienportalen und Airbnb vorgestellt werden.

Einleitung

Dieser Beitrag beruht auf Analysen, die im Rahmen des von der EU geförderten City Statistics-Projekts durchgeführt wurden. In Deutschland liegt die Projektkoordination beim Statistischen Bundesamt, das die vertraglichen Vorgaben in Zusammenarbeit mit der KOSIS-Gemeinschaft Urban Audit umsetzt¹. Die EU erwartet von der Bearbeitung dieses Arbeitspakets Vorschläge zur Verbesserung oder Ermöglichung der Nutzung von Verwaltungsdaten, Erhebungsdaten, Geodaten, offenen Daten, Big Data usw., die bei der Erstellung subnationaler Statistiken von Nutzen sein könnten. Insbesondere geht es dabei um Datenanforderungen, die nicht über die amtliche Statistik abgedeckt werden können. Bei den Analysen handelt es sich um Erprobungen von digitalen Datenquellen, die in den Bereich Big Data fallen.

Wenn also in diesem Beitrag von neuen Methoden und Datenquellen für die Raumbewertung gesprochen wird, so geschieht dies vor dem Hintergrund des Primats digitaler Technik, die der größte Innovationstreiber der letzten Jahrzehnte war und immer noch ist. Die Präsenz digitaler Technik ist in der modernen Gesellschaft kaum mehr wegzudenken und vor allem in der Arbeitswelt zu einem Faktum geworden. Die wichtigsten Aspekte dabei sind die flächendeckende globale wie lokale Vernetzung, die stetige Zunahme an Geschwindigkeit bei der Datenverarbeitung, die Expansion bei den Speicherkapazitäten und nicht zuletzt die immer kleiner werdende Strukturgröße bei der Halbleiterfertigung, die es erlaubt, immer mächtigere digitale Gerätschaft in bürotauglicher Größe bereitzustellen. Die Skaleneffekte der globalisierten Massenproduktion führten schließlich zur weitreichenden Verbreitung von digitalen Hilfs- und Konsumgeräten. Es ist wichtig, sich dieser grundlegenden Komponenten der Digitalisierung bewusst zu sein, wenn man die daraus resultierenden Innovationen verstehen und nicht den Fehler begehen möchte, neue digitale Verfahren als bloße Fortführung ihrer analogen Vorgänger zu betrachten.

Die in diesem Artikel verwendete Differenzierung zwischen nutzergenerierten, transaktionsgenerierten und sensorgenerierten Daten wurde bereits in der 2019 veröffentlichten Urban Audit-Broschüre² verwendet und basiert auf dem von der Berliner Senatsverwaltung für Stadtentwicklung und Wohnen in Auftrag gegebenen Bericht „Big Data und Crowd Data für die Berliner Stadtentwicklungsplanung“ aus dem Jahr 2017. Damit lassen sich die untersuchten Datenquellen

Tobias Link

Dipl.-Sozialwissenschaftler, Stadt Mannheim, Fachbereich Geoinformation und Stadtplanung

 tobias.link@mannheim.de

Schlüsselwörter:

Big Data – City Statistics-Projekt – nutzergenerierte Daten – transaktionsgenerierte Daten – Urban Audit

sinnvoll klassifizieren, wenn auch eine überschneidungsfreie Abgrenzung kaum möglich ist:

„**Nutzergenerierte Daten** entstehen durch Aktivitäten und oft insbesondere durch die Interaktion von Personen“ (Tonndorf 2019, 10). Die bekanntesten Repräsentanten dieser Art von Datengenerierung dürften OpenStreetMap (OSM) oder Wikipedia sein, deren Inhalte von der Nutzergemeinschaft beigesteuert werden. Für das City Statistics-Projekt wurden OSM-Daten bereits in der Vergangenheit erfolgreich genutzt, um die Länge des Radverkehrsnetzes in deutschen Städten zu berechnen (Schmidt 2017). Diesmal haben wir darüber hinaus im Zusammenhang mit der Generierung von City Statistics-Daten untersucht, ob OpenStreetMap eine alternative Datenquelle für die Berechnung der Anzahl von Theatern, öffentlichen Bibliotheken und öffentlichen Schwimmbädern sein kann.

Transaktionsgenerierte Daten „entsteh(en) quasi als Nebenprodukt im Rahmen von (digitalen) ökonomischen Transaktionen“ (Tonndorf 2019, 10), die zumeist in Form von Angebotsauflistungen auf themenspezifischen Internetportalen vorzufinden sind. Mithilfe der Methode des Web Scrapings haben wir Immobilienportale als alternative Quelle für die Erstellung von Statistiken zum Wohnungswesen herangezogen und Airbnb-Daten als Alternative für die Variablen zum Tourismus untersucht. Diese Portale decken einen großen Teil des deutschen Marktes ab.

Sensorgenerierte Daten werden „durch verschiedenartige Sensoren oder Sensornetzwerke“ (Tonndorf 2019, 10) erzeugt. Ein Beispiel für diese Art von Daten wurde in der Urban Audit-Broschüre 2019 gegeben, in der Günther Bachmann erläuterte, wie Mobilfunkdaten bestehende Datenquellen ergänzen können, um detaillierte Informationen über das Mobilitätsverhalten von Pendlern sowie Besuchern und Touristen zu liefern (Bachmann 2019). In diesem Beitrag spielen sensorgenerierte Daten jedoch keine Rolle und sollen an dieser Stelle nur der Vollständigkeit halber erwähnt werden.

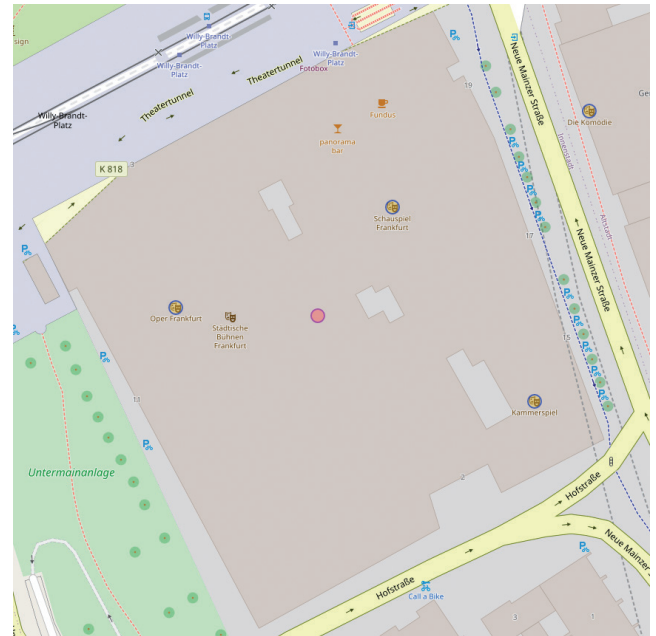
OpenStreetMap als Beispiel für nutzergenerierte Daten

Ermutigt durch unsere früheren Erfahrungen mit der Messung der Länge von Radwegenetzen versuchten wir, die Variablen „Anzahl der Theater“, „Anzahl der öffentlichen Bibliotheken“ und „Anzahl der öffentlichen Schwimmbäder“ mit Daten von OSM für die 127 deutschen Städte³ in der City Statistics-Datensammlung zu berechnen. Alle drei Variablen beziehen sich auf geografische Orte, die im Prinzip durch das OSM-Tagging-System erfasst werden können. Die Ergebnisse pro Stadt können mit bestehenden Daten für diese Variablen verglichen werden. Die folgenden Analysen wurden im April 2022 durchgeführt.

Theater

Die Anzahl der Theater wurde in OSM über die Schlüssel-Wert-Kombination „amenity“=“theatre“ operationalisiert, welche die einzelnen thematischen Einheiten identifiziert (Oper, Schauspiel, Musical, etc.). Da es bei dieser Variable um die Anzahl der Gebäude- bzw. Einrichtungsstandorte innerhalb der Stadt und nicht um die einzelnen Säle oder Bühnen geht, mussten wir

Abbildung 1: Mehrere unterschiedliche Bühnen innerhalb der Städtischen Bühnen Frankfurt, Frankfurt am Main



Quelle: OpenStreetMap

Standorte identifizieren, an denen sich mehrere verschiedene Arten von Bühnen ein Gebäude teilen, d. h. beispielsweise eine Opernbühne und eine Kammertheaterbühne. Dies wurde gelöst, indem wir die verschiedenen Geometrien getrennt betrachteten, d. h. wir suchten nach Relationen oder Wegen, die als Theater getaggt waren und die mehrere separate Knoten enthielten, die ebenfalls als Theater getaggt waren (Abb. 1).

Diese OSM-Ergebnisse wichen erheblich von den Daten ab, die der Deutsche Bühnenverein für das Referenzjahr 2019 zur Verfügung stellte: In 28 Städten gibt es eine positive Differenz von 10 oder mehr Theatern in den OSM-Daten; an der Spitze steht Berlin mit +156 Theatern, +53 in Hamburg, +46 in München und +43 in Köln. Die absolute Abweichung ist in den sehr großen Städten am größten, allerdings ändert sich das Bild und weist kein klares Muster mehr auf, wenn man die relativen Abweichungen betrachtet. Nur für einige wenige kleinere Städte ergeben die beiden unterschiedlichen Datenquellen das gleiche Ergebnis, wie im Fall von Celle, Landshut, Neu-Ulm und Pforzheim. Interessanterweise gab es kein Ergebnis auf der Grundlage der OSM-Daten, das niedrigere Werte aufwies als die Werte aus unserer bislang verwendeten Datenquelle. Eine Ursache für diese Differenzen ist ein grundlegender Unterschied in der Definition von Theater. Unsere derzeitige Quelle, der Deutsche Bühnenverein, ist eine Organisation, die ihre Statistiken auf der Grundlage aller ihrer Mitglieder erstellt, die aus kommunalen und staatlichen Theatern einschließlich aller Opernhäuser, den staatlichen Bühnen, zahlreichen privaten Theatern und den Orchestern bestehen. Nicht berücksichtigt werden dabei die meisten lokalen Kleinbühnen aller Art sowie Open Air-Locations, aber auch private Großveranstaltungsorte, die keinen festen, zum Standort gehörenden Betrieb haben (Konzertthallen). Der Schwerpunkt liegt auf einer traditionelleren Interpretation dessen, was ein Theater ist, wobei der

Schwerpunkt auf öffentlich finanzierten Einrichtungen liegt. Die OSM-Daten haben ein breiteres Verständnis davon, was ein Theater ist, und leider ist es kaum möglich, zu einer gezielteren Abfrage unter Verwendung der Schlüssel „theatre:type“ oder „theatre:genre“ überzugehen, da diese nicht konsequent genug verwendet werden, um eine bessere Annäherung an die bestehenden Werte der Variablen zu ermöglichen.

Bibliotheken

Die Anzahl der Bibliotheken aus OSM zu extrahieren ist einfacher, weil die Anzahl der Ausgabestellen gezählt werden kann, was bedeutet, dass mehrere Punkte pro Gebäude zugelassen sind. Das Schlüssel-Wert-Paar „amenity“ = „library“ macht genau das. Eine größere Herausforderung stellt die Unterscheidung zwischen öffentlichen und nicht öffentlichen Bibliotheken dar. Normalerweise sollte hierbei der Schlüssel „access“ helfen, nicht öffentliche Bibliotheken zu identifizieren, aber wie bei den Theatern wird dieser Schlüssel nicht konsequent genug verwendet, um sinnvoll eingesetzt werden zu können. Er würde eher den relativen Unterschied in der Anzahl der Bibliotheken verzerren, als dass er die Zahlen mehr in Einklang mit der Definition bringen würde.

Die Spanne der absoluten Abweichungen von den aktuellen Werten (Referenzjahr 2020) dieser Variable reicht von -24 (Essen) bis +133 (Hamburg), so dass die OSM-Daten diesmal in beide Richtungen abweichen. Auch die Stadtgröße scheint weder mit den absoluten noch mit den relativen Abweichungen stark zu korrelieren. Im Allgemeinen ist die Übereinstimmung der Ergebnisse aus den verschiedenen Datenquellen besser als bei den Theatern. Für 11 Städte stimmen die Ergebnisse überein, bei einer Toleranz von bis zu 10% erhöht sich diese Zahl auf 19 Städte. Die Gründe für die Unterschiede zwischen den Datenquellen dürften ähnlich wie bei der Anzahl der Theater sein: Unsere offizielle Datenquelle ist der Deutsche Bibliotheksverband e.V., der größte Bibliotheksverband in Deutschland, der mehr als 9.000 Bibliotheken vertritt. Es ist sehr wahrscheinlich, dass es auch hier eine Verzerrung zugunsten der öffentlich geförderten Bibliotheken gibt, während kleine private Bibliotheken, manchmal mit einem speziellen Themenschwerpunkt, unterrepräsentiert sein könnten. Ein Grund für die Überschätzung der Zahl der Bibliotheken in Universitätsstädten ist die Zählung der wissenschaftlichen Fachbibliotheken, die an diesen Universitäten angesiedelt sind, wenn OSM-Daten verwendet werden. Wir konnten noch keine zuverlässige Methode finden, um diese herauszufiltern.

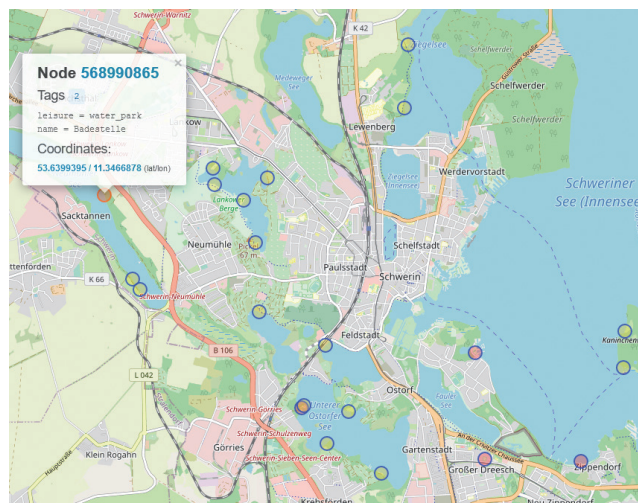
Öffentliche Schwimmbäder

Die letzte der drei Variablen, die wir mit OSM-Daten replizieren wollten, ist die Anzahl der öffentlichen Schwimmbäder in den Städten. Laut Glossar⁴ sind nur solche Schwimmbäder gemeint, die von der Allgemeinheit genutzt werden können, also nicht Schwimmbäder, die zu Fitnessstudios, Hotels oder anderen exklusiven Privateinrichtungen gehören. Auch sollen Strände und Badestellen ausgeschlossen werden. Es zählen Hallen- und Freibäder (wenn mindestens sechs Monate im Jahr geöffnet), die auch Teil von größeren Freizeitzentren oder Erholungskomplexen sein können. Einrichtungen mit mehreren Wasserbecken zählen als ein Schwimmbad. Diese Variable war am schwierigsten mit dem OSM-Tagging-System zu operati-

onalisieren, da es viele verschiedene Arten von öffentlichen Räumen gibt, die zum Schwimmen gedacht sind. Unter Berücksichtigung der Empfehlungen des OSM-Wikis entschieden wir uns für die folgende Kombination von Schlüssel-Wert-Paaren: Wasserparks haben in der Regel Einrichtungen in ihrem Angebot, die der Definition der Variable entsprechen und in OSM als „leisure“ = „water_park“ getaggt sind. Es gibt auch viele Sportzentren, die sich auf Schwimmsportarten spezialisiert haben, denen aber der vergnügungsorientierte Teil eines Wasserparks fehlt. Diese haben wir mit den Schlüssel-Wert-Paaren „leisure“ = „sports_centre“ und „sport“ = „swimming“ identifiziert. Darüber hinaus mussten wir auch einfache öffentliche Schwimmbäder berücksichtigen, die keine weiteren Angebote haben. Diese werden mit „leisure“ = „swimming_pool“ gekennzeichnet, und um mehrere Treffer für eine Einrichtung zu vermeiden, wenn es mehr als ein Schwimmbecken gibt, haben wir nach dem umgebenden Gebäude gefragt, indem wir den Schlüssel [„building“] hinzugefügt haben (Freibäder scheinen immer Teil eines Wasserparks oder Ähnlichem zu sein). Schließlich fügten wir auch öffentliche Bäder hinzu, die sich nicht im Freien befinden, da es sich bei öffentlichen Freibädern meist um nicht offizielle Badestellen in der Nähe von Seen oder Flüssen zu handeln scheint. Für diese haben wir die Schlüssel-Wert-Paare „amenity“ = „public_bath“ und „bath:type“ = „!~“ ^ (lake|river|sea)\$“ verwendet. Zu allen hier erwähnten Abfragen fügten wir auch das Schlüssel-Wert-Paar „access“ = „!~“ ^ private“ hinzu, um sicherzustellen, dass wir nur öffentliche Einrichtungen einbeziehen, und den Schlüssel „name“ um eine gewisse Sicherheit zu haben, dass es sich um einen offiziellen Ort handelt. Ein Problem trat in der Stadt Schwerin auf, wo viele öffentliche Strände als „leisure“ = „water_park“ gekennzeichnet sind (Abb. 2) und nur der Name (Badestelle) einen Hinweis darauf enthielt, dass es sich nicht um eine offizielle Badeeinrichtung handelt. Aus diesem Grund mussten wir diese explizit ausschließen, indem wir das Schlüssel-Wert-Paar „name“ = „!~“ ^ *Badestelle.*\$“ verwendeten.

Im Vergleich zu den bisher untersuchten Variablen scheint die Anzahl der Schwimmbäder recht gut mit OSM-Daten

Abbildung 2: Badestellen in Schwerin



Quelle: OpenStreetMap

reproduzierbar zu sein. Für die 104 Städte mit verfügbaren Werten aus unserem Datenbestand (wir haben den letzten verfügbaren Wert zwischen 2017 und 2020 verwendet, der von den Städten zur Verfügung gestellt wurde), erhielten wir 28 perfekte Übereinstimmungen, 32 Abweichungen um eins und 21 Abweichungen um zwei, was bedeutet, dass 68 % der Städte eine maximale Abweichung von zwei haben, 58 % haben eine maximale Abweichung von eins und 27 % sind eine perfekte Übereinstimmung. Die größte Abweichung beobachten wir für Berlin, wo wir 79 Schwimmbäder nach unserer aktuellen Quelle und 61 Schwimmbäder nach OSM haben, was eine Differenz von 18 Schwimmbädern bedeutet. Im Allgemeinen besteht die größte Herausforderung darin, öffentliche Freibäder auszuschließen, die an Stränden errichtet wurden und keine künstlichen Schwimmbecken haben (Abb. 3). Dies könnte auch den großen Unterschied in den Zahlen für Berlin erklären, wo es viele Seen mit öffentlichen Stränden gibt.

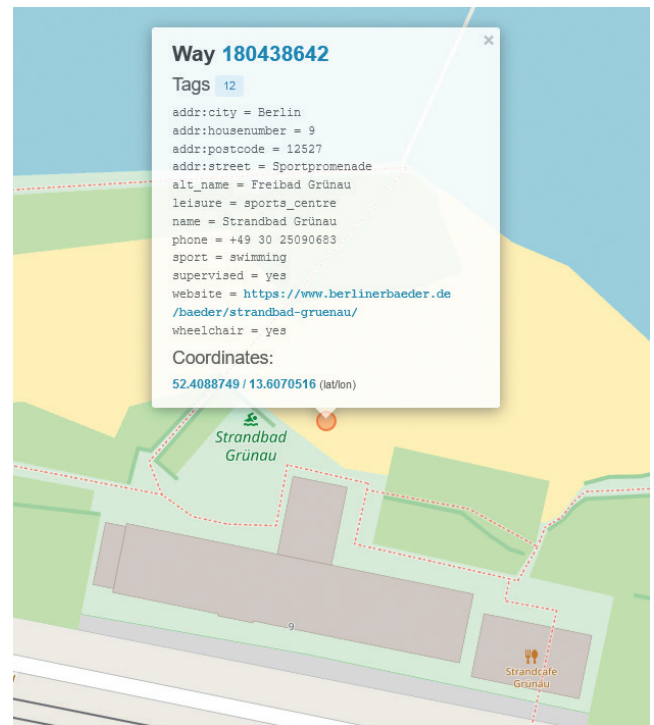
Immobilienportale und Airbnb als Beispiele für transaktionsgenerierte Daten

Immobilienportale

Für unsere aktuelle Urban Audit-Broschüre 2021⁵ haben wir Daten von Immobilienportalen im Internet getestet, die durch Web Scraping erhoben wurden. Die Daten zu den aktuellen Angebotspreisen auf dem Mietmarkt stammen von der Metasuchmaschine Immosuchmaschine.de, einem Service der Immobilien Scout Österreich GmbH (Teil der Scout24-Gruppe). Sie umfasst Mietangebote auf verschiedenen deutschen Immobilienportalen und Kleinanzeigenseiten, die über Makler oder privat provisionsfrei angeboten werden. Die Angebote reichen von Wohnimmobilien wie Mietwohnungen, Eigentumswohnungen und Häusern über Grundstücke bis hin zu Gewerbeimmobilien wie Büros, Geschäftsräume und Anlageimmobilien. Es handelt sich hier also um transaktionsgenerierte Daten, die mehrmals täglich automatisch von verschiedenen Online-Portalen abgerufen und den Nutzern der Website in aggregierter Form zur Verfügung gestellt werden. Die gesammelten Daten basieren auf allen Wohnungsinseraten, die an den Stichtagen 5. Oktober, 6. Oktober und 11. Oktober 2021 für die befragten Städte der Urban-Audit-Befragung 2018/19 („Perception survey on quality of life“ der EU⁶ und die Koordinierte Umfrage zur Lebensqualität in deutschen Städten der AG Umfragen des VDSt⁷) angezeigt wurden.

Um die vorliegenden Daten für einzelne Städte effizient aus Immosuchmaschine.de extrahieren zu können, wurde die Methode des Web Scraping mit dem Firefox-Add-on Web Scraper (www.webscraper.io) gewählt. Bei diesem Ansatz werden die Daten direkt aus dem Browser extrahiert und nicht, wie sonst üblich, durch in Python oder R programmierte Skripte. Der größte Vorteil dieser Methode ist, dass der Zugriff auf die Daten über einen „echten“ Webbrowser und damit für den Server kaum unterscheidbar von der Interaktion mit einem „echten“ Menschen erfolgt. Die Verwendung eines Browser-Add-ons erfordert auch weniger technisches Hintergrundwissen, da das Erlernen einer Programmiersprache wie Python oder R nicht notwendig ist, um diesen Ansatz anzuwenden. Der Nachteil dieses Ansatzes gegenüber der Programmierme-

Abbildung 3: Strandbad Grünau, Berlin



Quelle: OpenStreetMap

thode ist, dass das verwendete Tool weniger Flexibilität und weniger Funktionalität bietet. Jedoch waren beide Aspekte für diese experimentelle Anwendung nicht sonderlich von Bedeutung.

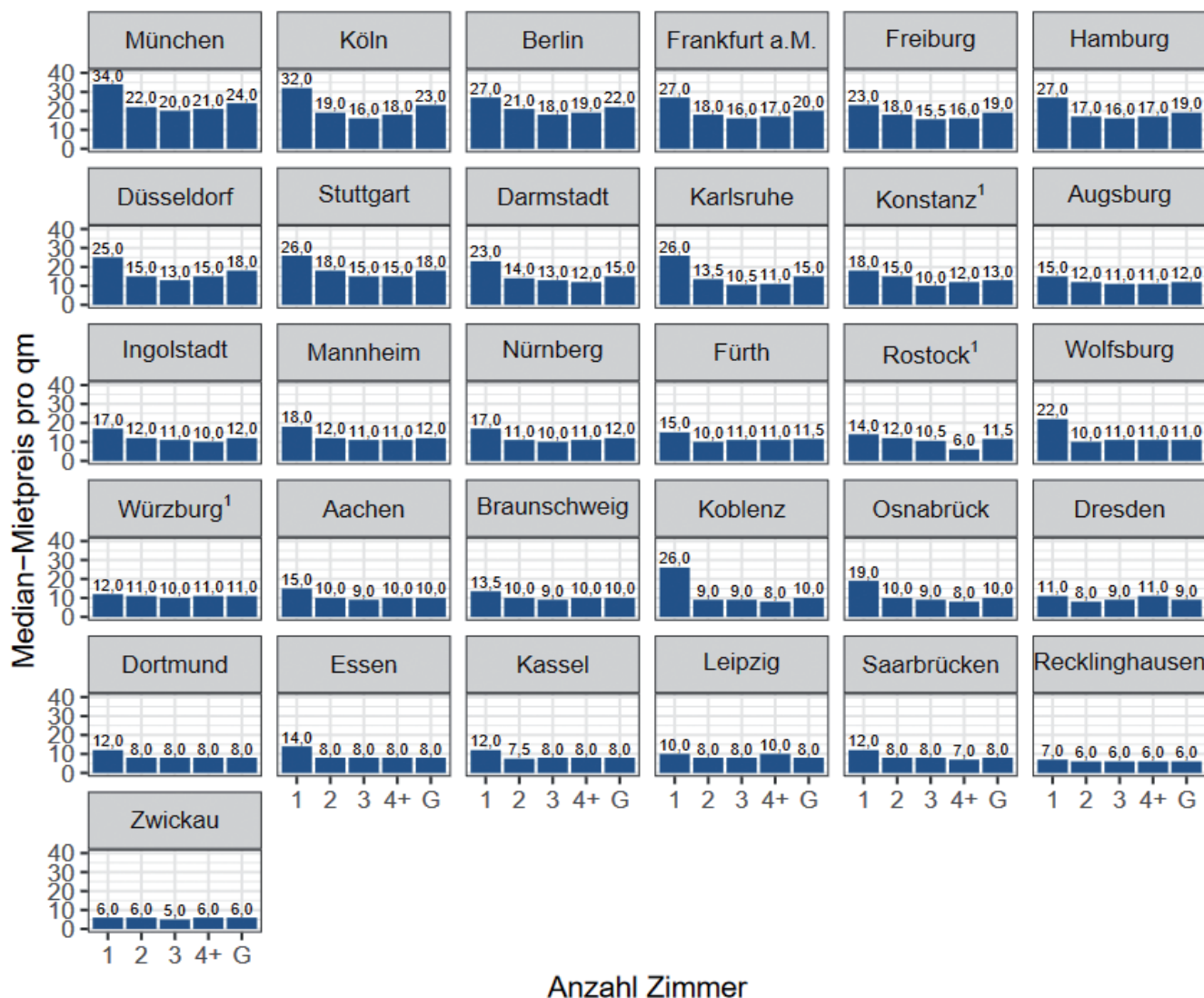
Der auf diese Weise erzeugte vollständige Datensatz besteht aus 37.344 Einzelfällen, die die Informationen „Postleitzahl“, „Miete in Euro“, „Preis pro m²“, „Wohnfläche in m²“, „Anzahl der Zimmer“ und „Datenquelle“ enthalten. Die Angaben in der Kategorie „Datenquelle“ geben an, wo das Inserat ursprünglich erschienen ist, d.h. aus welchem Online-Portal Immosuchmaschine.de die Informationen bezogen hat. Wenig überraschend ist die größte Datenquelle mit deutlichem Abstand ImmobilienScout24 (80,5 %), gefolgt von Wohnung-jetzt.de (4,1 %), Locaberlin.de (3,0 %), Regionalimmobilien24.de (2,5 %), IVD24immobilien.de (2,5 %), The-homelike.com (2,2 %), Immo4trans.de (1,5 %) und diversen anderen, die nach Rundung jeweils einen Anteil von 1 % oder weniger aufweisen.

Diesmal ging es weniger um die Kompatibilität mit den Variablen der Städtestatistik als vielmehr um die Methode des Web Scraping als Mittel der Datenerhebung. In unserer Broschüre haben wir die Daten auf ihre Plausibilität geprüft, indem wir sie mit den Einstellungen der Bürger zum Wohnungsmarkt verglichen haben, die in der Urban-Audit-Befragung von 2018/19 gemessen wurden (vgl. Link 2021). Die Ergebnisse waren für einen ersten Versuch recht vielversprechend und wir denken, dass diese Methode in Zukunft weiter vertieft werden kann.

Airbnb

Zwei Variablen zum Tourismus, nämlich die Gesamtzahl der Übernachtungen in Beherbergungsbetrieben und die Zahl der Schlafgelegenheiten in Beherbergungsbetrieben, wurden nä-

Abbildung 4: Median-Mietpreis pro m² in den Städten nach Anzahl der Zimmer



G = Gesamt; ¹Fallzahl unter 100

Quelle: Immosuchmaschine.de, eigene Berechnungen

her untersucht. Laut Glossar⁸ sind Beherbergungsbetriebe „ein örtlicher Leistungserbringer, der als entgeltliche Dienstleistung [...] kurzfristige oder kurzzeitige Beherbergungsdienste anbietet [...]“. Die Zahl der Schlafgelegenheiten wird „durch die Zahl der Personen bestimmt, die in den in der Unterkunft aufgestellten Betten übernachten können, wobei zusätzliche Betten, die auf Wunsch der Kunden aufgestellt werden können, nicht berücksichtigt werden“.

Die im City Statistics-Projekt verwendete Datenquelle ist die amtliche Tourismusstatistik, die für alle enthaltenen deutschen Städte eine einheitliche und valide Grundlage verspricht. Die deutsche Tourismusstatistik erfasst Beherbergungsbetriebe mit mindestens 10 Betten sowie Campingplätze mit mindestens 10 Stellplätzen. Allerdings übernachten immer mehr Touristen in kleinen, privat geführten Beherbergungsbetrieben mit weniger als 10 Betten.

Daher untersuchten wir anhand einer Fallstudie der Stadt Stuttgart, ob Airbnb-Daten die bestehenden Tourismusdaten ergänzen und verbessern könnten. Stuttgart ist die sechstgrößte Stadt in Deutschland, hat einen sehr angespannten Wohnungsmarkt sowie einen wachsenden Tourismus. Allerdings hat Stuttgart seit 2016 die Satzung gegen Zweckentfremdung umgesetzt⁹, wonach über 50 Prozent des verfügbaren Wohnraums nicht länger als 10 Wochen im Jahr für Fremdbeherbergung genutzt werden darf. Damit soll verhindert werden, dass der zunehmende Fremdenverkehr nicht zu sehr den ohnehin überhitzten regulären Wohnungsmarkt belastet. Die Vergleichbarkeit mit strukturell ähnlichen Städten, die eine solche Richtlinie bisher nicht umgesetzt haben, ist dadurch natürlich nur eingeschränkt möglich.

Die ersten beiden Zeilen von Tabelle 1 zeigen die Daten der offiziellen Tourismusstatistik für Stuttgart, die an EUROSTAT

Tabelle 1: Daten der amtlichen Tourismusstatistik und Airbnb für Stuttgart

	2015	2016	2017	2018	2019
Übernachtungen in Touristenunterkünften	3.561.490	3.706.017	3.781.564	3.911.781	4.086.683
Gästebetten insgesamt (alle Arten von Beherbergungsbetrieben)	19.080	20.418	20.712	21.494	22.122
Angebotene Privatunterkünfte über Airbnb ¹⁰	532	790	1.080	-	1.097

Quellen: Statistisches Landesamt Baden-Württemberg, Beherbergung im Reiseverkehr seit 1984 und Held et al. 2019: 1.

gemeldet wurden. Die Gesamtzahl der Übernachtungen in Beherbergungsbetrieben stieg von etwas weniger als 3,6 Millionen im Jahr 2015 auf über 4 Millionen im Jahr 2019 (+0,5 Millionen). Die Zahl der Schlafgelegenheiten stieg von etwas mehr als 19.000 auf mehr als 22.000 (+3.000). Die über Airbnb angebotenen Privatunterkünfte verdoppelten sich von etwas mehr als 500 im Jahr 2015 auf fast 1.100 im Jahr 2017, blieben aber seither nahezu unverändert. Im Durchschnitt verfügt jedes Airbnb-Objekt über 2,5 Schlafgelegenheiten, was bedeutet, dass Airbnb 2019 rund 12 Prozent der Kapazitäten der Stuttgarter Beherbergungsbetriebe ausmachen. Dies ist natürlich eine relevante Größe, allerdings ist nur ein Teil der Airbnb-Unterkünfte dauerhaft verfügbar. Die meisten Angebote sind zeitlich begrenzt, zum Beispiel in den Semesterferien oder am Wochenende.

Die Gesamtzahl der in Airbnb-Unterkünften verbrachten Nächte ist nicht verfügbar, so dass die konkrete Anwendung für den Datenkatalog des City Statistics-Projekts ohnehin nicht geprüft werden kann.

Im Vergleich zu Stuttgart werden nur in wenigen anderen Großstädten wie Berlin, München und Hamburg deutlich mehr Unterkünfte über Airbnb angeboten. Auch in diesen Städten sind Richtlinien zur Zweckentfremdung von Wohnraum umgesetzt worden, jedoch handelt es sich dabei um die drei größten Städte Deutschlands mit jeweils einem Vielfachen der Bevölkerung Stuttgarts und nochmals deutlich umfangreichem Tourismus. Private Kleinunterkünfte scheinen also nur für einen sehr kleinen Teil der 127 Städte des Städtestatistikprojekts ein relevanter Faktor des Beherbergungsmarktes zu sein. Zum jetzigen Zeitpunkt scheinen die vom marktführenden Airbnb-Portal generierten Transaktionsdaten im Allgemeinen nicht in der Lage zu sein, die bestehenden Daten der Tourismusstatistik zu verbessern.

Fazit

Die Nutzung neuer Datenquellen aus dem Bereich Big Data geht vielfach einher mit der Aneignung neuer Methoden. Die kaum zu überschauende Vielfalt an digitalen Werkzeugen, die einem die Datenakquirierung und das Datenmanagement erleichtern bzw. überhaupt ermöglichen sollen, nimmt kontinuierlich zu. Dabei ist es einerseits vor allem eine Herausforderung, die Orientierung und Übersicht zu behalten und sich nicht in den technischen Details zu verlieren. Andererseits sind die Möglichkeiten der Partizipation in diesem Bereich

so einfach wie nie, denn sowohl die Dokumentation als auch die Software selbst sind in den meisten Fällen kostenfrei und quelloffen erhältlich – oder es gibt zumindest fast immer eine Alternative aus diesem Bereich. Ob sich eine Einarbeitung lohnt, hängt natürlich immer von den individuellen Anforderungen und zeitlichen Einschränkungen am Arbeitsplatz ab. Um diese Entscheidung zu vereinfachen, gibt es mittlerweile auch im Bereich der Städtestatistik zunehmend Erfahrungsberichte und Studien (z. B. Bachmann 2020, Dosch 2021 oder Plennert 2016), die versuchen, komplexe Sachverhalte mit bisher schwieriger Datenlage neu zu erschließen und dabei auf große digitale Datenquellen zurückgreifen. Auch dieser Beitrag soll eine kleine Übersicht zu den Möglichkeiten der Nutzung großer digitaler Datenquellen bieten. Nicht alle dieser Datenquellen entsprechen den Anforderungen der bisherigen Datenkataloge oder bieten eine sinnvolle Ergänzung, wenn man viele unterschiedliche Städte betrachten möchte. Es lohnt sich aber, offen für neue Perspektiven zu sein, die diese Daten meistens auch enthalten und damit Entwicklungen im Blick zu behalten, die womöglich zukünftig eine größere Rolle spielen werden.

- 1 <https://www.staedtestatistik.de/arbeitsgemeinschaften/kosis/urban-audit>
- 2 https://www.staedtestatistik.de/fileadmin/media/Kosis/Urban_Audit/PDF/Broschueren/UA_Broschuere_2019.pdf
- 3 Es handelt sich um Städte mit einer Mindesteinwohnerzahl von 50.000, die den von der EU geforderten Kriterien genügen. Nähere Informationen zur Definition einer Stadt nach dem Urbanisierungsgrad sind in der Urban Audit-Broschüre von 2017 nachzulesen (https://www.staedtestatistik.de/fileadmin/media/Kosis/Urban_Audit/PDF/UA_Broschuere_2017_de.pdf).
- 4 https://ec.europa.eu/eurostat/statistics-explained/index.php?title=City_statistics_%E2%80%93_culture_and_recreation
- 5 https://www.staedtestatistik.de/fileadmin/media/Kosis/Urban_Audit/PDF/Broschueren/UA_Broschuere_2021_DE.pdf
- 6 https://ec.europa.eu/regional_policy/information-sources/maps/quality-of-life_en
- 7 <https://www.staedtestatistik.de/arbeitsgemeinschaften/vdst/ag-umfragen/koordinierte-umfrage-zur-lebensqualitaet>
- 8 https://ec.europa.eu/eurostat/statistics-explained/index.php?title=City_statistics_-_tourism
- 9 <https://www.stuttgart.de/zweckentfremdung>
- 10 Daten zu Angeboten in der Airbnb-Datenbank jeweils zum Stand Juli der Jahre 2015 bis 2017 und 2019 wurden untersucht

Literatur

- Alter, H.; Feuerhake, J.; Jacob, S. (2021): Insolvenzstatistik in der Corona-Pandemie – aktuellere Ergebnisse durch Webscraping. In: Statistisches Bundesamt (Destatis) (Hrsg.): WISTA 3/2021. Wiesbaden, S. 58–70.
- Bachmann, G. (2019): Wissenschaftsstadt Darmstadt: Neue Verkehrsanalysen mit Mobilfunkdaten – ein Zwischenbericht. In: KOSIS-Gemeinschaft Urban Audit (Hrsg.): Das deutsche Urban Audit – Lebensqualität: Erschließung neuer Datenquellen (2019). Mannheim, S. 34–50.
- Bachmann, G. (2020): Stadt und Verkehr: Neue Verkehrsanalysen mit Mobilfunkdaten – ein Zwischenbericht. Stadtforschung und Statistik: Zeitschrift des Verbandes Deutscher Städtestatistiker, 33(1), 52–60.
- Barron, C. (2019): Universitätsstadt Mannheim: Einsatz von Crowd Data für stadtklimatische Fragestellungen. In: KOSIS-Gemeinschaft Urban Audit (Hrsg.): Das deutsche Urban Audit – Lebensqualität: Erschließung neuer Datenquellen (2019). Mannheim, S. 51–53.
- Baur, S. (2019): Ausgewählte deutsche Städte: Verknüpfung von Crowd Data und Befragungsdaten im Bereich Verkehr. In: KOSIS-Gemeinschaft Urban Audit (Hrsg.): Das deutsche Urban Audit – Lebensqualität: Erschließung neuer Datenquellen (2019). Mannheim, S. 27–33.
- Baur, S. (2020): Erschließung von Crowd Data und Verknüpfung mit Befragungsdaten im Bereich Verkehr. Stadtforschung und Statistik: Zeitschrift des Verbandes Deutscher Städtestatistiker 33(2), S. 30–35.
- Berliner Senatsverwaltung für Stadtentwicklung und Wohnen (Hrsg.) (2017): Big Data und Crowd Data für die Berliner Stadtentwicklungsplanung. Berlin/Zürich. (https://digital.zlb.de/viewer/metadata/16317737/1/LOG_0000/)
- De Lazzer, J.; Rengers, M. (2021): Auswirkungen der Coronakrise auf den Arbeitsmarkt: experimentelle Statistiken aus Daten von Online-Jobportalen. In: Statistisches Bundesamt (Destatis) (Hrsg.): WISTA 3/2021. Wiesbaden. S. 71–88.
- Demunter, C. (2019): EUROSTAT: Nutzung von Big Data in der Tourismusstatistik. In: KOSIS-Gemeinschaft Urban Audit (Hrsg.): Das deutsche Urban Audit – Lebensqualität: Erschließung neuer Datenquellen (2019). Mannheim, S. 18–22.
- Dosch, F. (2021): Satellitengestützte Erfassung des Stadtgrüns. Stadtforschung und Statistik: Zeitschrift des Verbandes Deutscher Städtestatistiker 34(2), S. 8–16.
- Hausmann, P.; von Bodelschwingh, A. (2016): Wohnungsmarktanalyse mit Daten von „Immobilienscout24“. Stadtforschung und Statistik: Zeitschrift des Verbandes Deutscher Städtestatistiker 29(1), S. 43–52.
- Held, T.; Strauß, M.; Veller, M. (2019): Kein Airbnb-Boom in Stuttgart. In: Landeshauptstadt Stuttgart, Statistisches Amt (Hrsg.): Statistik und Informationsmanagement, Monatsheft 12/2019. Stuttgart. S. 364–366.
- Kaczorowski, W. (2017). Neue digitale Daten für die Entwicklung smarter Städte und Regionen. Bertelsmann Stiftung (Hrsg.). Gütersloh.
- Link, T. (2019): Urban Audit: Nutzung von Crowd Data für die Berechnung der Länge des Radwegenetzes in deutschen Städten. In: KOSIS-Gemeinschaft Urban Audit (Hrsg.): Das deutsche Urban Audit – Lebensqualität: Erschließung neuer Datenquellen (2019). Mannheim, S. 23–26.
- Link, T. (2021): Wohnungsmarktsituation in den deutschen Städten der Urban Audit-Befragung 2018/19: Eine Gegenüberstellung mit den Angebotsmieten aus Onlineportalen. In: KOSIS-Gemeinschaft Urban Audit (Hrsg.): Das deutsche Urban Audit - Subjektive Einschätzungen zur Lebensqualität in europäischen Städten (2021). Mannheim, S. 38–48.
- Plennert, M. (2016): Anwendungsreif? Nutzung und Potential von digitalen Geodaten für Stadtforschung und Raumbearbeitung am Fallbeispiel OpenStreetMap. Stadtforschung und Statistik: Zeitschrift des Verbandes Deutscher Städtestatistiker 29(1), S. 27–34.
- Saidani, Y.; Bohnensteffen, S.; Hadam, S. (2022): Qualität von Mobilfunkdaten – Projekterfahrungen und Anwendungsfälle aus der amtlichen Statistik. In: Statistisches Bundesamt (Destatis) (Hrsg.): WISTA 5/2022. Wiesbaden. S. 55–67.
- Schmidt, S. (2017): Messung der Gesamtlänge des Radwegenetzes in Urban Audit-Städten auf Basis von OpenStreetMap-Daten. In: KOSIS-Gemeinschaft Urban Audit (Hrsg.): Das deutsche Urban Audit – Lebensqualität in Stadt und Umland (2017). Mannheim, S. 34–56.
- Schnorr-Bäcker, S. (2016). Big Data als neuer Weg der Datengewinnung für die amtliche Statistik. Stadtforschung und Statistik: Zeitschrift des Verbandes Deutscher Städtestatistiker 29(1), S. 2–10.
- Tonndorf, T. (2019): Big Data und Crowd Data. In: KOSIS-Gemeinschaft Urban Audit (Hrsg.): Das deutsche Urban Audit - Lebensqualität: Erschließung neuer Datenquellen (2019). Mannheim, S. 9–17.
- Wiengarten, L.; Zwick, M. (2017). Neue digitale Daten in der amtlichen Statistik. In: Statistisches Bundesamt (Destatis) (Hrsg.): WISTA 5/2017. Wiesbaden, S. 19–30.
- Zwick, M. (2015). Big Data in der amtlichen Statistik. In: Bundesgesundheitsblatt - Gesundheitsforschung – Gesundheitsschutz 58(8).