

Online Wohnungssuche und die Analyse individueller Wohnortpräferenzen mithilfe von Big Data

Timo Schnepf

Master of Science Student at the University of Cologne, Germany

✉ schnepf@mailbox.org



26. Oktober 2019

Motivation Webseiten wie [“wg-gesucht.de“](http://wg-gesucht.de), [“quoka.de“](http://quoka.de), [“immowelt.de“](http://immowelt.de) oder [“ebay-kleinanzeigen.de“](http://ebay-kleinanzeigen.de) sind moderne Formen von schwarzen Brettern mit Informationen über Wohnungssuchende, deren Ressourcen und Präferenzen. Sie bieten eine Datenquelle für die Analyse der Nachfrageseite von Wohnraum. Für Köln beispielsweise wurden im August 2019 im Schnitt am Tag 8,8 Anzeigen allein auf Ebay-Kleinanzeigen geschaltet. Für Berlin 35,6, was in etwa dem proportionalen Faktor der Wohnbevölkerung entspricht. Online Wohnungsgesuche bieten für die Stadtforschung eine wertvolle und meines Wissens bisher nicht genutzte Datenquelle. Die Daten bieten sich aus zwei Gründen an. Zum einen aufgrund des Datenumfangs. Im Zeitraum von Juli und August 2019 schalteten Kölner um die 600, Berliner nur innerhalb des Augusts um die 1200 Anzeigen auf [“ebay-kleinanzeigen.de“](http://ebay-kleinanzeigen.de). Durch die automatisierte Sammlung über beispielsweise ein Jahr lässt sich damit ein beachtlicher Datensatz erstellen. Zum anderen deuten erste Analysen der Gesuche-Texte darauf hin, dass die Portale von Individuen mit verschiedenen sozioökonomischen Hintergründen gleichermaßen genutzt werden (bspw. nach Einkommenssituation und Familienstand).

Das hier vorgestellte Messinstrument erlaubt die Untersuchung *individueller Wohnortpräferenzen in Echtzeit für verschiedene sozioökonomische Gruppen*. Welche Stadtteile präferieren Studenten und in welchen Monaten? Welche Stadtteile Familien mit hohem, welche mit geringem Einkommen? Wie entwickelt sich die Wichtigkeit der Anbindung zum ÖPNV als Kriterium des Wohnortwunsches? Wiederholte Inserate derselben Personen sowie deren Verweildauer könnten Rückschlüsse auf Probleme für bestimmte Personengruppen erlauben, Wohnraum zu finden. Dies ist für alle deutschen Städte mit genügend Aktivität auf Online-Wohnungsgesuche-Portalen durchführbar.

In Korrespondenz mit Prof. Dr. Marcel Helbig vom WZB merkte er an, dass sich - hinsichtlich der Repräsentativität - vor allem jüngere Altersgruppen online exponieren und die Portale von Stadt zu Stadt unterschiedlich intensiv genutzt werden. Er beurteilte diesen neuen Ansatz dennoch positiv als ein mögliches *“Tool um [...] zukünftige Gentrifizierungsprozesse zu messen“*.

Anwendungsbeispiel Der Datensatz wird mithilfe der sogenannten Web-Scraping Methode erstellt. Dabei werden spezifische Informationen automatisiert von bestimmten Webseiten heruntergeladen. In diesem Fall lädt ein geschriebenes Python-Script sys-

tematisch Informationen aus allen veröffentlichten Wohnungsgesuchen der Stadt Köln von [“ebay-kleinanzeigen.de“](https://www.ebay-kleinanzeigen.de) im Zeitraum Juli und August 2019 herunter und speichert sie in einem beliebigen Datenformat ab. Jede Beobachtung (Zeile) entspricht einer Gesuche-Seite. Von zentralem Interesse ist der Gesuche-Text, da dieser Stichwörter zu u.a. Wohnortpräferenzen, dem Arbeitsverhältnis oder dem Familienstand *enthalten kann*. Sobald beispielsweise das Wort *“Ehrenfeld“* oder *“Sülz“* im Anzeigentext erscheint, erhält eine Indikatorvariable für den jeweiligen Stadtteil den Wert 1.

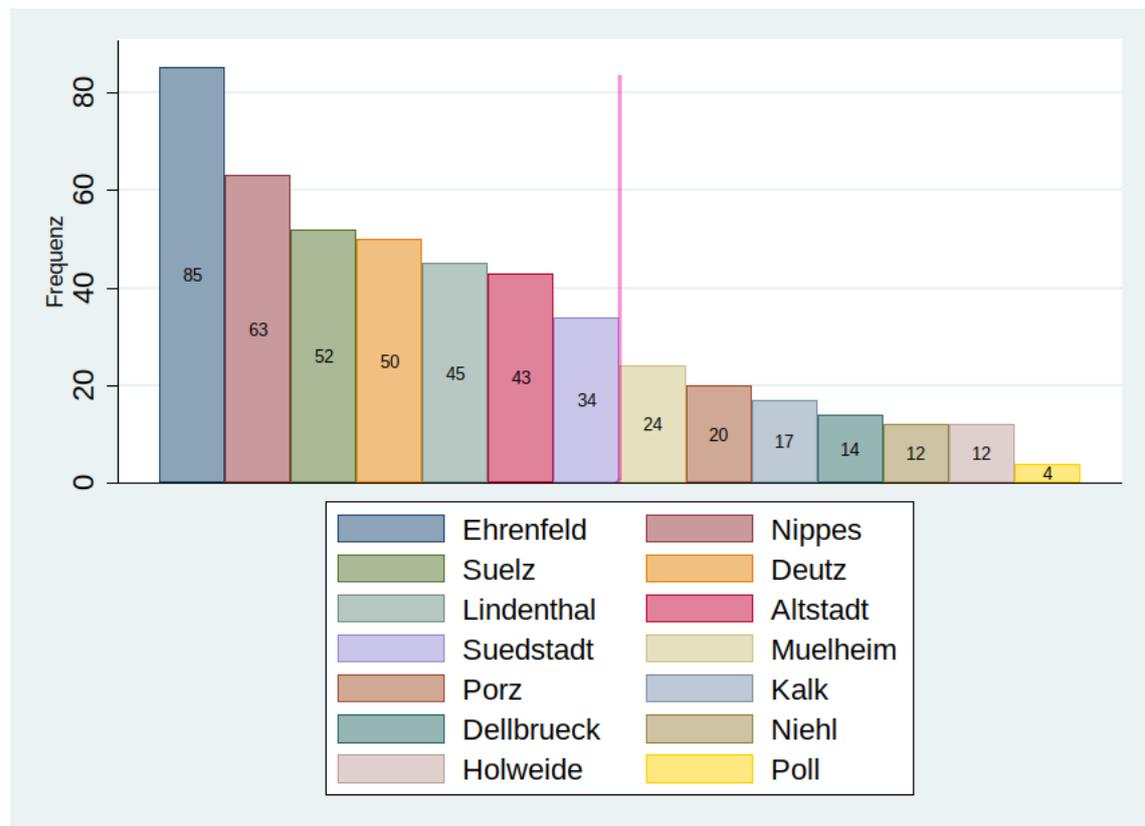


Abbildung 1: Verteilung der Wohnortpräferenzen nach Stadtteilen in Köln. Mehrfachangaben möglich, nicht alle Stadtteile enthalten.

Abbildung 1 zeigt die Häufigkeit der Nennungen für einige Stadtteile. Ehrenfeld wird mit 85 Nennungen am häufigsten genannt. Alle Striche rechts der eingezeichneten roten Linie sind rechtstheinische Stadtteile. Deutz ist der einzige rechtsrheinische Stadtteil (mit 50 Nennungen) links der roten Linie.

In unserem Beispiel wurde mithilfe entsprechender Indikatorwörter eine einfache Varia-

ble zur Messung des sozio-ökonomischen Status (SES) erstellt. Beispiele für geringen SES waren z.B. “arbeitssuchend“, “WBS“, “Amt“. Für hohen SES bspw. “Akademiker“, “promoviert“ oder “oeffentlicher Dienst“. Außerdem konnte mit guter Qualität eine Indikatorvariable für Einzelpersonen, Paare und Familien erstellt werden (nicht im Beispiel).

ses	Wohnungswunsch				Total
	Rechts	Links	Beides	Keine Ang	
Gering	10 15.38	4 6.15	0 0.00	51 78.46	65 100.00
KA	50 10.46	107 22.38	23 4.81	298 62.34	478 100.00
Hoch	5 3.76	43 32.33	19 14.29	66 49.62	133 100.00
Total	65 9.62	154 22.78	42 6.21	415 61.39	676 100.00

Abbildung 2: Kreuztabelle für Wohnungswünsche nach SES-Indikatorwörtern. Frequenz und Spaltenprozent.

Abbildung 2 zeigt eine Kreuztabelle auf Basis von 676 Wohnungsgesuchetexten aus Köln für den Zeitraum von Anfang Juli bis Ende August 2019. Es lässt sich erkennen, dass Individuen mit hohem SES deutlich häufiger ausschließlich linksrheinische Wohnortpräferenzen besitzen (32%) als Individuen mit geringem SES (6%). Interessant ist auch die Beobachtung, dass Personen mit geringem SES sehr häufig keine Angaben zum Wohnortwunsch tätigen (78%) im Vergleich zu Individuen mit hohem SES (50%).

Es sei darauf hingewiesen, dass die geringen Fallzahlen im Beispiel mit der Zeit der Datenerhebung zunehmen und hier nur als Anwendungsbeispiel dienen sollen. Weitere extrahierbare Informationen sind beispielsweise: maximaler Miethöhenwunsch, Anzahl gewünschter Zimmer und Quadratmeter, Namen und Alter der Inserenten. Ebenfalls denkbar ist eine Approximation des Bildungsniveaus anhand maschineller Textanalyse (NLTK).

Für weitere Rückfragen stehe ich Ihnen gerne zur Verfügung.