

# Wohnungsjagd in der Schweiz: Ein WebScraping-Abenteuer durch Bern, Basel und Zürich

Nabil Mikhael | Alessandro Gregori

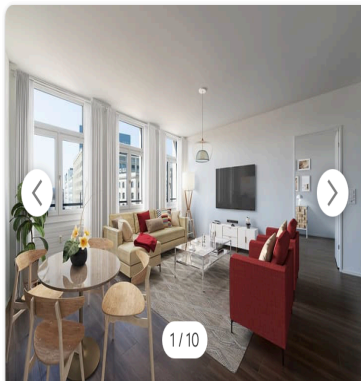
# Einleitung

- ▶ **Ziel:**  
Untersuchung von 3.5-Zimmerwohnungen in Zürich, Basel und Bern
- ▶ **Vorgehen:**  
Webscraping der Plattform immoscout24.ch mit RVEST (Stand 21.05)
- ▶ **Was ist Webscraping?**  
Webscraping ist eine Technik zur automatisierten Extraktion von Daten aus Webseiten.

# Kernfunktionen von RVEST:

- ▶ `read_html()`: Lädt HTML-Inhalte von Webseiten.
- ▶ `html_elements()`: Extrahiert HTML-Elemente aus HTML-Dokument mittels CSS-Selektoren
- ▶ `html_text()`: Extrahiert den Textinhalt von HTML-Elementen

# Implementierung - Funktion Extraktion letzte Seite (1/2)

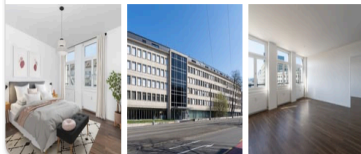


**3.5 Zimmer, 71m<sup>2</sup>, CHF 1'820.-**

 Klybeckstrasse 190, 4057 Basel

**"Einzigartige Lage! Ihr neues Zuhause?"**

An der Klybeckstrasse 190 in Basel vermieten wir moderne 3.5 Zimmerwohnungen sind auf der Webseite [www.klybeck190.ch](http://www.klybeck190.ch) ersichtlich. An Ihnen die folgende Ausstattung: • Moderner Ausbau • Wohnräume: klassisch (Farbkonzept) • Stilvolle Küche mit hochwertiger Ausstattung, dunkel



**Extraktion letztes  
Zeichen (Seitenzahl)  
aus HTML - Element**

# Implementierung - Funktion Extraktion letzte Seite (2/2)



```
# Funktion zur Extraktion der Nummer der letzten Seite aus einer URL
get_last_page <- function(url) {

  # Lädt das HTML-Dokument von der angegebenen URL
  page <- read_html(url)

  # Verwendet CSS-Selektoren, um das vorletzte Element innerhalb der Paginator-Box
  last_page_number <- page %>%
    html_elements(".HgPaginationSelector_paginatorBox_nNiMu a:nth-last-of-type(2)")
  # Extrahiert den Textinhalt des HTML-Elements (dies sollte die Seitenzahl sein)
  html_text() %>%
  # Schneidet das letzte Zeichen des Textes ab (Seitenzahl als letztes Zeichen)
  str_sub(.,-1) %>%
  # Konvertiert den extrahierten Text in eine numerische Form
  as.numeric()

  # Druckt die extrahierte Nummer der letzten Seite zur Überprüfung
  print(paste("Extracted last page number:", last_page_number))

  # Gibt die Nummer der letzten Seite zurück
  return(last_page_number)
}
```

# Web-Scraping Preis, Wohnungsgrösse und PLZ pro Stadt

```
size <- page %>%  
  html_elements("span+strong") %>%  
  html_text() %>%  
  gsub("m²", "", .) %>% # mit gsub m2 entfernen aus Fläche  
  as.numeric() # in numerische Zahl umwandeln
```

**85m2 zu 85**

```
price <- page %>%  
  html_elements(".HgListingRoomsLivingSpacePrice_price_u9Vee") %>%  
  html_text() %>%  
  gsub(pattern = "[^0-9]", replace = "", .) %>% # Alle Werte ersetzen die keine  
  as.numeric()
```

**CHF2'360 zu 2360**

```
plz <- page %>%  
  html_elements("div address") %>%  
  html_text() %>%  
  sapply(function(x) {  
    parts <- strsplit(x, " ")[[1]] # Teilt die Adresse an den Leerzeichen  
    if (length(parts) >= 2) {  
      sub_part <- trimws(parts[length(parts)-1]) # Wählt vorletzten Teil und  
      return(sub_part)  
    } else {  
      return(NA) # Rückgabe NA, wenn weniger als zwei Teile vorhanden sind  
    }  
  })
```

**Ackerstrasse 38,4057 Basel zu 4057**

# Datenzusammenführung im DataFrame:

Show 10 entries

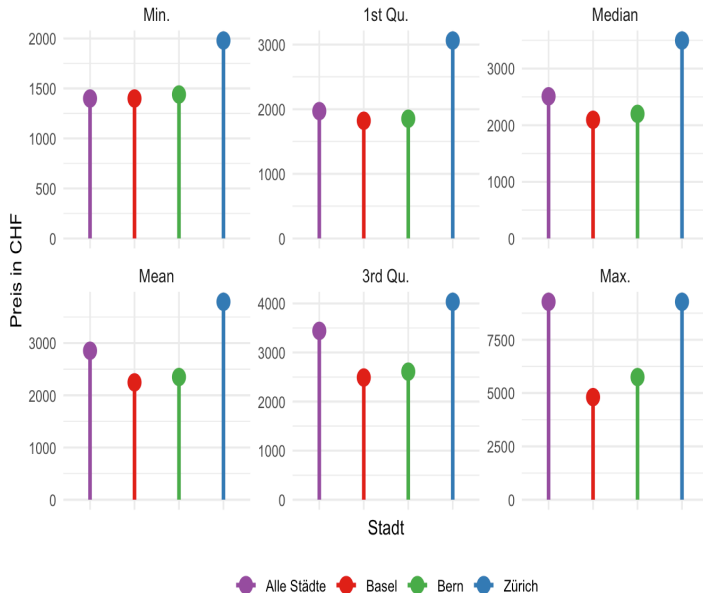
Search:

	Price [CHF]	Size [m2]	PLZ	City
1	3470	155	4001	basel
2	4810	171	4056	basel
3	4360	140	4001	basel
4	3460	114	4058	basel
5	3430	147	4057	basel
6	2380	136	4053	basel
7	2670	117	4058	basel
8	2910	170	4051	basel
9	3420	166	4051	basel
10	4900	150	4057	basel

Showing 1 to 10 of 378 entries

# Lollipop-Plot - Preise

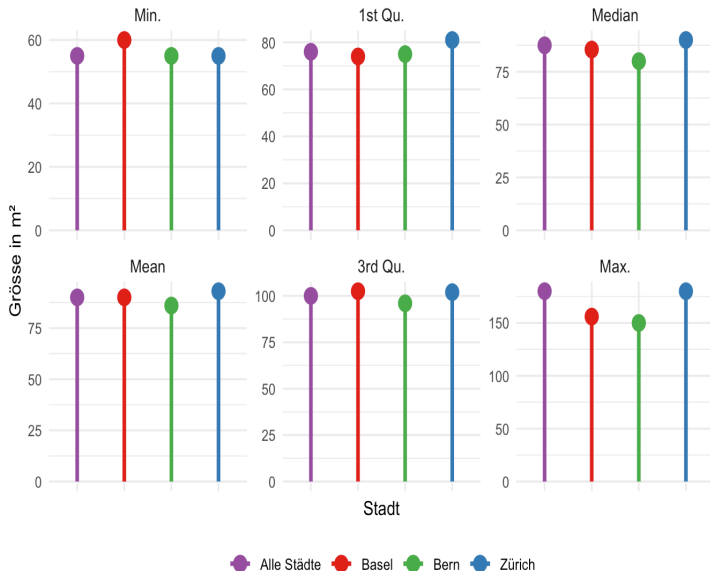
Zusammenfassung der Preise nach Stadt



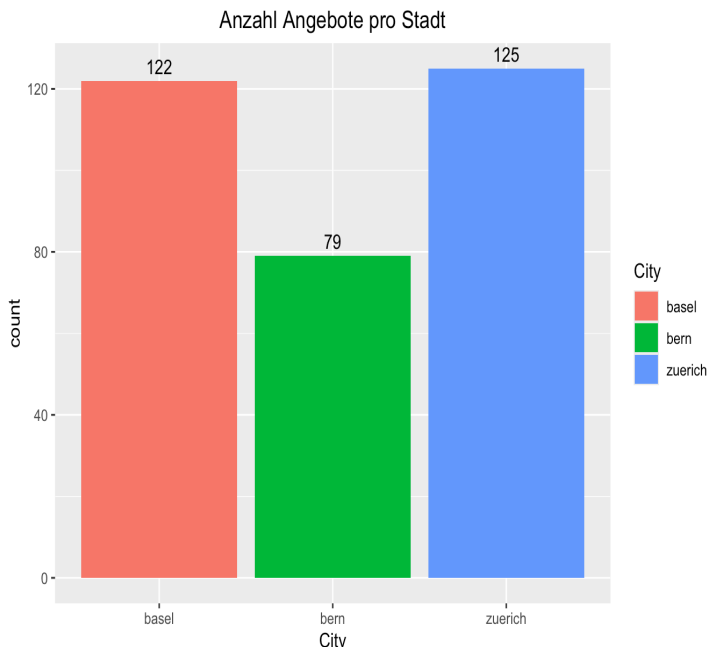


# Lollipop-Plot - Wohnungsgrösse

## Zusammenfassung der Wohnungsgrössen nach Stadt

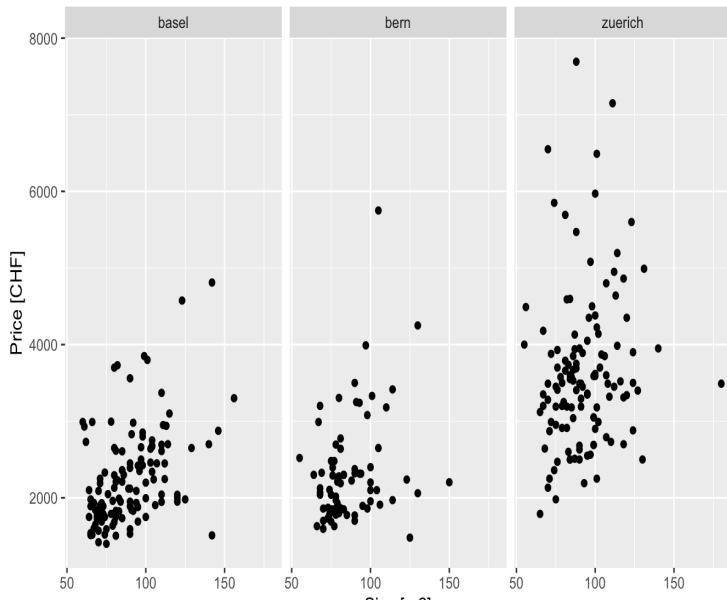


# Barplot Anzahl Angebote pro Stadt



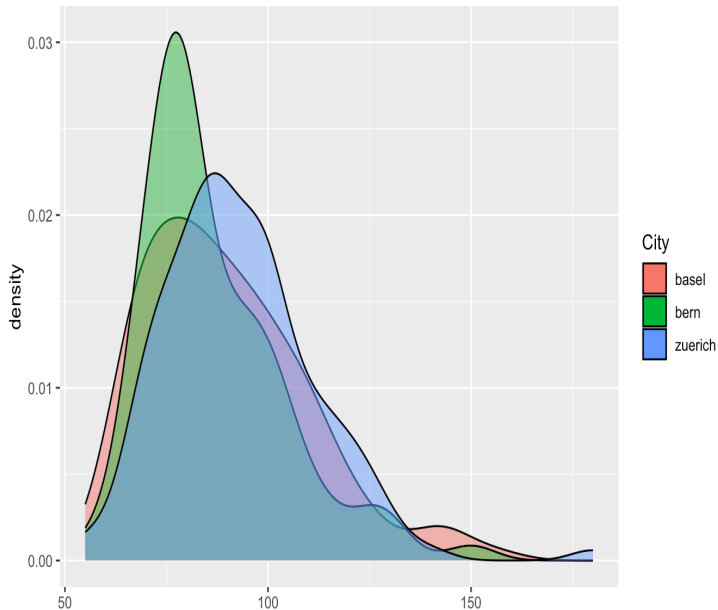
# Scatterplot Wohnungsgrösse vs Preis

Vergleich Städte: Size/Price



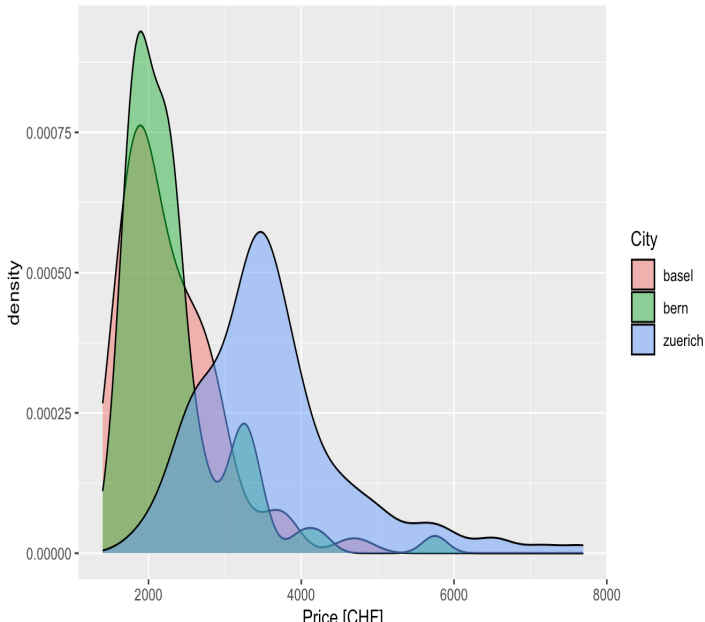
# Densityplot Wohnungsgrößenvergleich

Wohnungsgrößenvergleich nach Stadt

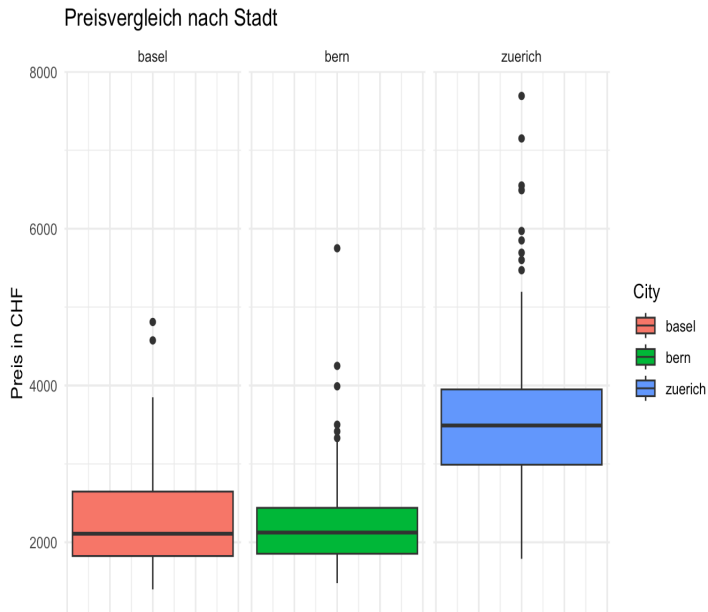


# Densityplot Preisvergleich

Preisvergleich nach Stadt

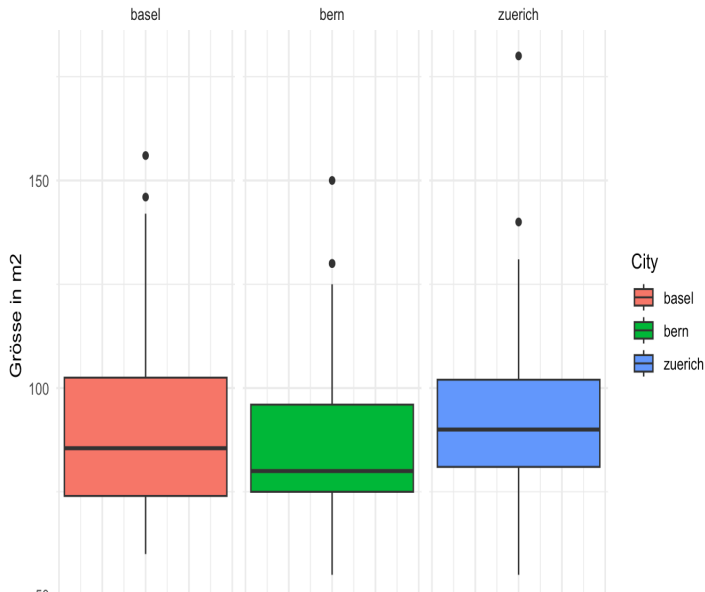


# Boxplot Preisvergleich



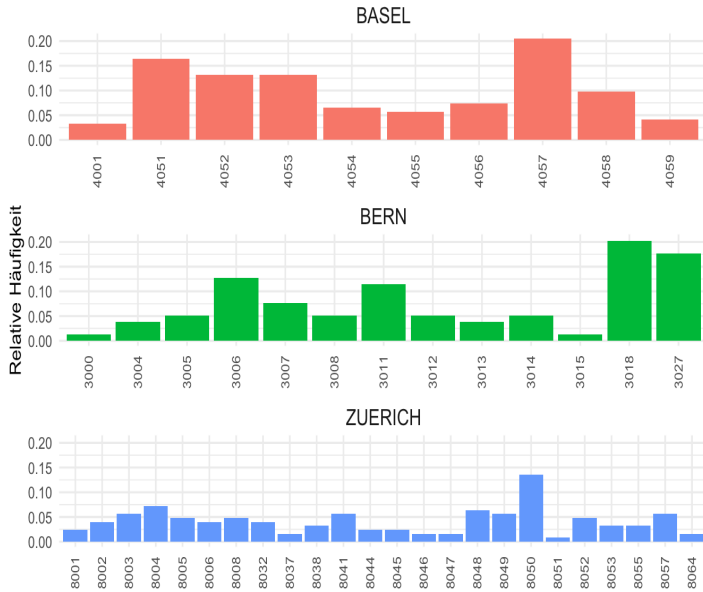
# Boxplot Wohnungsgrößenvergleich

Wohnungsgrößenvergleich nach Stadt



# Barplot Anzahl PLZ

Relative PLZ-Häufigkeit pro Stadt





## Zusammenfassung (1/2)

- ▶ **Mietpreise:** Zürich hat die höchsten Median - Mietpreise für 3.5-Zimmerwohnungen, gefolgt von Basel und Bern.
- ▶ **Preisspanne:** Zürich zeigt eine breitere Verteilung der Mietpreise, während Basel und Bern eine konzentriertere Verteilung aufweisen.
- ▶ **Durchschnittliche Wohnungsgrößen:** Die Verteilung der Wohnungsgrößen ist in allen drei Städten ähnlich, mit einem Schwerpunkt auf Wohnungen zwischen 70 und 100 Quadratmetern. In Bern dominieren kleinere Wohnungen.

## Zusammenfassung (2/2)

- ▶ **Anzahl der Angebote:** Zürich und Basel haben fast identisch viele Angebote für 3.5-Zimmerwohnungen und bieten jeweils etwa 40% mehr Wohnungen als Bern.
- ▶ **Korrelation Preis / Wohnungsgrösse :** Es gibt eine positive Korrelation zwischen Wohnungsgrösse und Mietpreis in allen drei Städten. Zürich zeigt häufigere und extremere Ausreisser, bei denen grosse Wohnungen extrem hohe Preise haben.
- ▶ **Relative Häufigkeit PLZ:** Die freien Wohnungen konzentrieren sich in Basel auf die PLZ 4057 (ca. 20 %), in Zürich auf die PLZ 8050 (ca. 15 %) sowie Bern auf die PLZ 3018 (ca. 20 %)

- ▶ **Problematik:** Stichprobengrösse mit  $n = 1$ , dadurch keine verlässliche Interpretation möglich

## Weiterführende Analysen:

- ▶ **Preisentwicklung:** Überwachung und Analyse der Mietpreisentwicklung über einen längeren Zeitraum, um Trends und Muster zu erkennen.
- ▶ **Prognosen:** Erstellung von Modellen zur Vorhersage zukünftiger Mietpreise basierend auf langfristiger Preisentwicklung
- ▶ **Stadtteilanalyse:** Detaillierte Analyse der Mietpreise auf Stadtteilebene, um Unterschiede innerhalb einer Stadt zu identifizieren.