

# Optimierung der Preisstrategien bei Airbnb: Eine Analyse zur Maximierung der Einnahmen

Cedric Gisler & Jovan Pajic

## A. 1. Abstrakt

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

## B. 2. Einleitung (mit Forschungsfrage (d.h. Geschäftsfrage) am Ende)

In der heutigen, schnelllebigen Welt des Online-Tourismus spielen Plattformen wie Airbnb eine zentrale Rolle bei der Art und Weise, wie Menschen reisen und Unterkünfte buchen. Airbnb bietet eine Vielzahl von Unterkünften an, von einfachen Zimmern bis hin zu luxuriösen Villen. So vielfältig wie das Angebot sind auch die Vorlieben und Erwartungen der Gäste. Vor diesem Hintergrund stellt sich die Forschungsfrage: **Welche Eigenschaften einer Airbnb-Unterkunft ziehen Gäste an und ermöglichen es, einen höheren Preis pro Apartment zu erzielen?**

Der grösste Unterschied eines Airbnbs ist deren Grösse. Eine der wichtigsten Eigenschaften eines Apartments ist aber die Lage, die Bewertung und der Gastgeber, somit definieren wir folgende Nullhypothesen:

### 1. Bewertung der Airbnb-Unterkünften:

- **Nullhypothese (H0):** Die Höhe der Gesamtbewertung (review\_scores\_rating) hat keinen signifikanten Einfluss auf den Preis pro Person der Unterkunft.
- **Alternativhypothese (H1):** Die Höhe der Gesamtbewertung (review\_scores\_rating) hat einen signifikanten Einfluss auf den Preis pro Person der Unterkunft.
- **Fragen:**
  - Hat die Gesamtbewertung der Unterkunft (review\_scores\_rating) einen signifikanten Einfluss auf den Preis pro Person?
  - Gibt es bestimmte Bewertungsmetriken (z.B. Sauberkeit, Kommunikation), die den Preis pro Person stärker beeinflussen als andere?

### 2. Gastgeber der Airbnb-Unterkunft:

- **Nullhypothese (H0):** Die Attribute des Hosts wie “host\_is\_superhost”, “host\_identity\_verified” und “host\_listings\_count” haben keinen signifikanten Einfluss auf den Preis pro Person der Unterkunft.
- **Alternativhypothese (H1):** Die Attribute des Hosts wie “host\_is\_superhost”, “host\_identity\_verified” und “host\_listings\_count” haben einen signifikanten Einfluss auf den Preis pro Person der Unterkunft.
- **Fragen:**
  - Beeinflusst der Status “Superhost” (host\_is\_superhost) den Preis pro Person?
  - Hat die Verifizierung des Hosts (host\_identity\_verified) einen Einfluss auf den Preis pro Person?
  - Wie beeinflusst die Anzahl der Listings eines Hosts (host\_listings\_count) den Preis pro Person?

### 3. Lage des Airbnb:

- **Nullhypothese (H0):** Die Entfernung zum Stadtzentrum (berechnet durch geografische Koordinaten) hat keinen signifikanten Einfluss auf den Preis pro Person der Unterkunft.
- **Alternativhypothese (H1):** Die Entfernung zum Stadtzentrum (berechnet durch geografische Koordinaten) hat einen signifikanten Einfluss auf den Preis pro Person der Unterkunft.
- **Fragen:**
  - Beeinflusst die Entfernung zum Stadtzentrum den Preis pro Person der Unterkunft?
  - Gibt es andere geografische Faktoren (z.B. Nähe zu bestimmten Attraktionen), die den Preis pro Person beeinflussen?

Weitere wichtige Eigenschaften sind Ereignisse (z.B. Zürich: Streetparade, Zürich Film Festival) in der betreffenden Stadt. Da wir aber die Daten nur zu einem einzelnen Zeitpunkt haben und keine Zeitserie, können wir den Einfluss dieser Ereignisse auf den Preis pro Person der Unterkunft nicht untersuchen.

**Deskriptive Analyse:** Wir möchten überprüfen, ob unsere Hypothesen so stimmen und wir mit unserer Annahme über die Einflüsse auf den Preis pro Person der Unterkunft richtig liegen.

**Prädiktive Analyse:** Wir möchten den Trend des Preises pro Person der Unterkunft aufzeigen und untersuchen, ob wir mithilfe der Bewertung und den wichtigsten Eigenschaften einer Unterkunft eine Prognose über den Preis machen können.

**Präskriptive Analyse:** Wir möchten zeigen, welche Eigenschaften wie verbessert werden müssen, um die Preise pro

Person der Unterkunft signifikant erhöhen zu können, um einen höheren Preis erzielen zu können.

### C. 3. Datenquelle (mit Angaben zu Quelle, Qualität und Bereinigungsschritten der Daten)

Die Daten für diese Analyse stammen von der Inside Airbnb Organisation [1], die sich dafür einsetzt, ihre Gemeinden vor den negativen Auswirkungen von Kurzzeitvermietungen zu schützen. Diese Organisation sammelt und veröffentlicht regelmässig aktualisierte Datensätze, die aus öffentlich verfügbaren Informationen auf der Airbnb-Website stammen. Diese Datensätze würden wir als Vertrauenswürdig einstufen.

Die extrahierten Datensätze umfassen Informationen aus drei bedeutenden Regionen in der Schweiz: Zürich (27. Dezember 2023), Genéve (27. Dezember 2023) und Vaud (10. März 2024).

Die Daten umfassen verschiedene Dateien für jede Stadt, wobei für die Analyse hauptsächlich das "listings\_long.csv"-File verwendet wird, da es für die Geschäftsfragen relevant ist. Die Qualität der Daten in diesem File ist insgesamt sehr hoch, mit wenigen leeren Feldern und einer konsistenten Struktur innerhalb der Spalten.

Es werden insgesamt 3 Datensätze verwendet, die dieselbe Struktur aufweisen, jedoch aus drei verschiedenen Regionen stammen, die wir hier analysieren möchten.

Einige Spalten, wie "description", "neighborhood\_overview", "host\_neighborhood", "neighborhood", und die Beschreibung der Liegenschaft, wie "bathrooms" und "bedrooms", weisen eine beträchtliche Anzahl leerer Felder auf. Es wird vermutet, dass diese Felder optional für die Gastgeber sind und daher nicht immer ausgefüllt werden. Ebenso fehlt bei einigen Einträgen der Preis, was eine Analyse erfordert, um mögliche Korrelationen mit anderen Feldern, wie dem ersten Review, zu identifizieren.

Trotz dieser kleinen Unregelmässigkeiten ist die Datenqualität insgesamt hoch, und die Beschreibung der Datenfelder wird durch das Data Dictionary [2] gut unterstützt.

#### Datenbereinigungsschritte:

1. **Entfernung irrelevanter oder leerer Spalten:** Vor der Analyse wurden alle Spalten entfernt, die für die Fragestellungen nicht relevant sind oder leere Felder enthalten.
2. **Überprüfung der Einheitlichkeit und Konsistenz der Werte:** Die verbleibenden Spalten wurden auf Einheitlichkeit der Werte und Konsistenz der "N/A"-Kennzeichnungen überprüft, um sicherzustellen, dass die Daten konsistent und interpretierbar sind.
3. **Analyse von Einträgen ohne Preisangabe:** Einige Einträge weisen keine Preisangabe auf, was eine Analyse erfordert, um mögliche Korrelationen mit anderen Feldern, wie dem ersten Review, zu identifizieren. Je nach Ergebnis dieser Analyse könnten Einträge ohne Preisangabe entfernt oder anderweitig behandelt werden, um die Datenintegrität zu gewährleisten.

### D. 4. Datenqualität (Analyse im Hinblick auf Datenqualitätsaspekte)

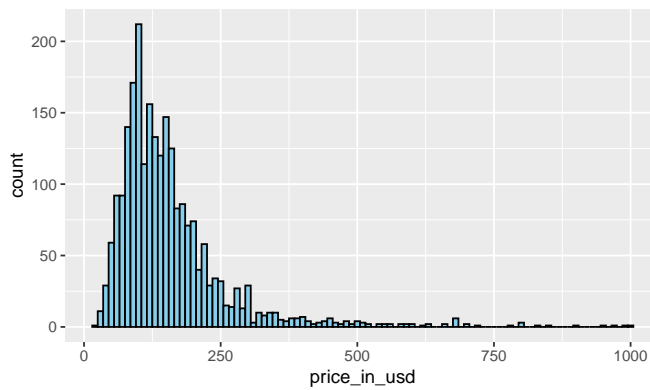
Nachdem wir die Zusammenfassung der Daten betrachtet und eine Analyse der Datensätze durchgeführt haben, haben wir die folgenden Datenanpassungen vorgenommen:

1. Konvertiere **host\_response\_rate** von Zeichenfolge (chr) in Ganzzahl (integer), wobei "N/A" durch NA ersetzt wird. Entferne das Prozentzeichen (%) und benenne die Spalte in **host\_response\_rate\_in\_%** um.
2. Konvertiere **host\_acceptance\_rate** von Zeichenfolge (chr) in Ganzzahl (integer), wobei "N/A" durch NA ersetzt wird. Entferne das Prozentzeichen (%) und benenne die Spalte in **host\_acceptance\_rate\_in\_%** um.
3. Konvertiere **price** von Zeichenfolge (chr) in Dezimalzahl (double), wobei das Dollarzeichen (\$) entfernt wird. Benenne die Spalte in **price\_in\_\$** um.
4. Lösche die folgenden Spalten: **description**, **neighborhood\_overview**, **host\_location**, **host\_about**, **host\_neighbourhood**, **host\_verifications**, **neighbourhood**, **neighbourhood\_cleansed**, **bathrooms**, **bedrooms**, **amenities**, **calendar\_updated**, **license**.
5. Analysiere fehlende Werte (NA) oder leere Felder und ersetze sie gegebenenfalls.

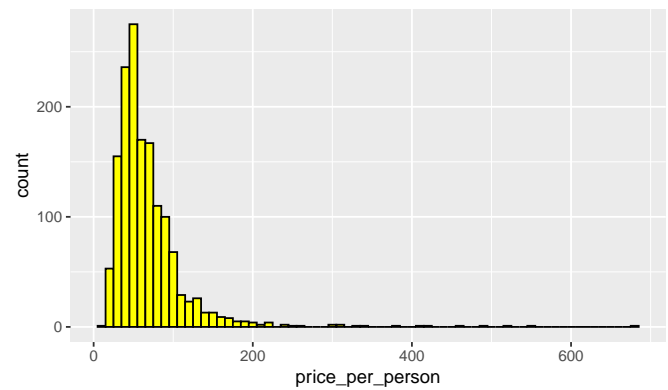
Nach diesen Anpassungen haben wir die Daten weiter analysiert und die folgenden Schritte durchgeführt:

1. **Überprüfung auf Duplikate:** Es wurde überprüft, ob es doppelte **ids** in den Daten gibt.
2. **Visualisierung der Verteilungen:** Es wurden Histogramme erstellt, um die Verteilungen der Preise, der Bewertungsscores und der Bewertungen pro Monat zu visualisieren.
3. **Analyse von Anomalien und Ausreissern:** Da der Preis stark variiert, analysieren wir die Möglichkeit, den Preis pro Person zu berechnen. Dabei fällt auf, dass der Preis bei "Entire home/apt" für die maximale Anzahl der Gäste (**accommodates**) berechnet wird, was sinnvoll ist, da die gesamte Unterkunft gebucht wird. Bei "Private room" und "Shared room" hingegen wird der Preis mehrheitlich pro Person angegeben. Im Gegensatz dazu ist der Preis bei "Hotel room" wieder für die maximale Gästeanzahl festgelegt.
4. **Export der Datensätze:** Die drei bereinigten Datensätze wurden in neuen Dateien erstellt und abgespeichert.

Verteilung des Preises in USD

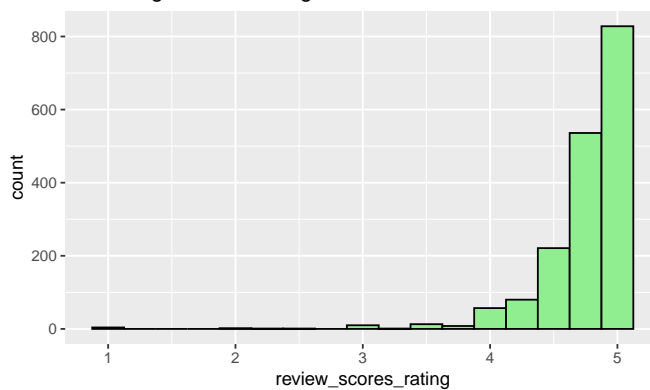


Verteilung des Preises pro person in USD

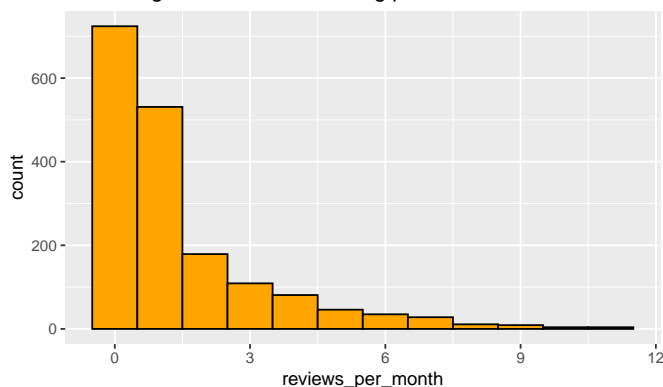


Die analysierten Werte zeigen eine ausgeglichene Verteilung ohne signifikante Ausreisser oder extreme Werte. Dies deutet darauf hin, dass der Datenbereinigungsprozess effektiv war und die Daten konsistent und zuverlässig für weitere Analysen vorliegen.

Verteilung der Bewertung



Verteilung von Anzahl Bewertung pro Monat



#### E. 5. Datenanalyse (Informationen zur Datenstruktur, Organisation und zu den für die Analyse verwendeten Methoden)

1) **Datenstruktur und Organisation:** Die vorliegenden bereinigten Datensets besteht aus Informationen zu Airbnb-Unterkünften in Zürich, Geneve und Vaud und umfasst 63 Variablen und mehrere tausend Zeilen, wobei jede Zeile eine einzelne Unterkunft repräsentiert. Die Variablen enthalten diverse Informationen, darunter:

- **ID und URLs:** Eindeutige Identifikationsnummern und Links zu den Listings.
- **Host-Informationen:** Daten über die Gastgeber, wie `host_id`, `host_name`, `host_is_superhost`, `host_identity_verified`, und `host_listings_count`.
- **Eigenschaften der Unterkunft:** Variablen wie `accommodates`, `bedrooms`, `bathrooms`, `beds`, `price_in_usd`, und `price_per_person`.
- **Bewertungen:** Bewertungsscores wie `review_scores_rating`, `review_scores_cleanliness`, `review_scores_checkin`, etc.
- **Geografische Daten:** `latitude` und `longitude`.
- **Verfügbarkeitsdaten:** `availability_30`, `availability_60`, `availability_90`, `availability_365`.

#### Zusammenfassung Preis pro Person und Bewertung Zürich

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
12.75	42.50	56.50	68.79	80.38	676.00

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.000	4.660	4.850	4.749	5.000	5.000

#### Zusammenfassung Preis pro Person und Bewertung Vaud

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
3.75	34.29	47.50	52.96	65.00	310.50

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
------	---------	--------	------	---------	------

```
# A tibble: 4 x 2
  room_type count
  <chr>      <int>
1 Entire home/apt 2119
2 Private room    670
3 Shared room     21
4 Hotel room       9
```

1.000    4.670    4.860    4.752    5.000    5.000

### Zusammenfassung Preis pro Person und Bewertung Genf

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
13.75	41.50	57.00	64.46	75.50	410.00

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.00	4.60	4.82	4.71	5.00	5.00

2) *Methoden der Datenanalyse:* Um die Forschungsfragen zu beantworten und die Hypothesen zu testen, wurden verschiedene statistische Methoden und Verfahren angewendet:

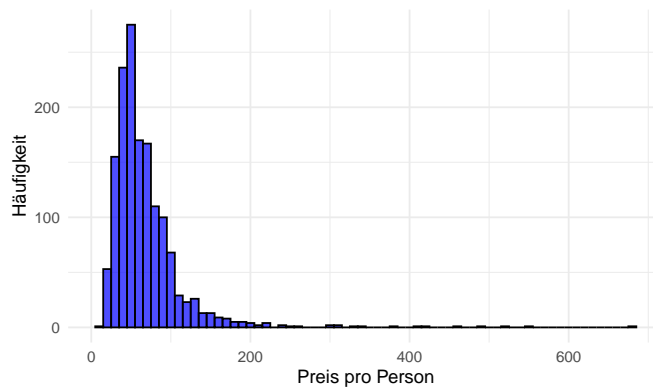
#### a) *Deskriptive Statistik:*

Die deskriptive Statistik liefert grundlegende Informationen über die Struktur und Verteilung der Daten. Ziel ist es, einen Überblick über die Daten zu bekommen und mögliche Anomalien oder interessante Muster zu erkennen.

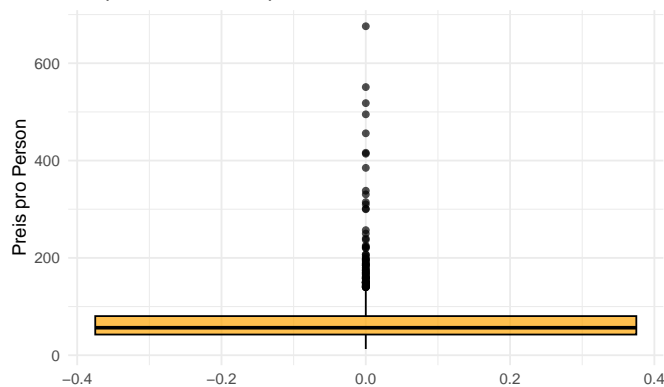
- **Ziel:** Ein grundlegendes Verständnis der Datenstruktur und Verteilung der Variablen zu erlangen.
- **Massnahmen:** Berechnung von Mittelwerten, Mediane, Standardabweichungen und Visualisierung der Datenverteilung durch Histogramme und Boxplots.

Kurze Übersicht über die Datensets Zürich, Vaud und Geneva

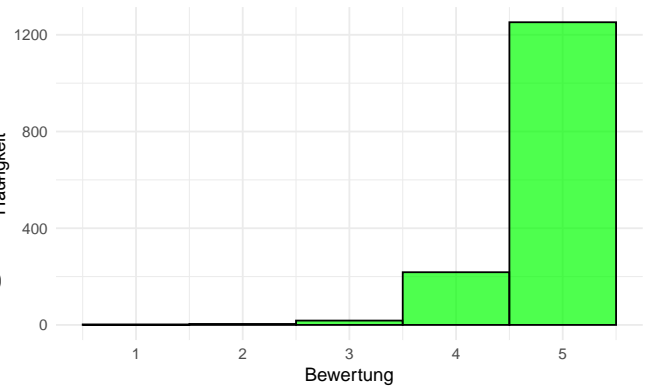
Verteilung des Preises pro Person in Zürich



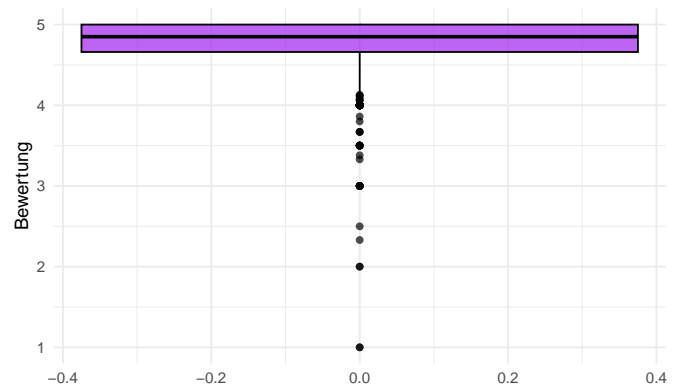
Boxplot des Preises pro Person in Zürich



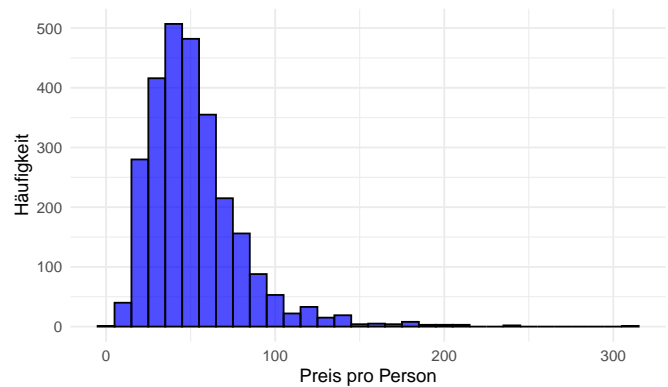
Verteilung der Bewertung in Zürich



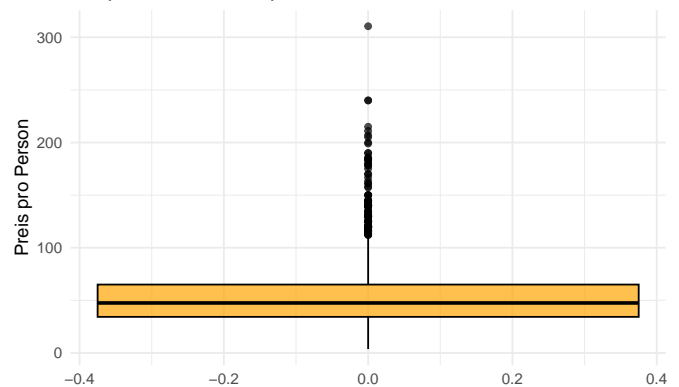
Boxplot der Bewertung in Zürich



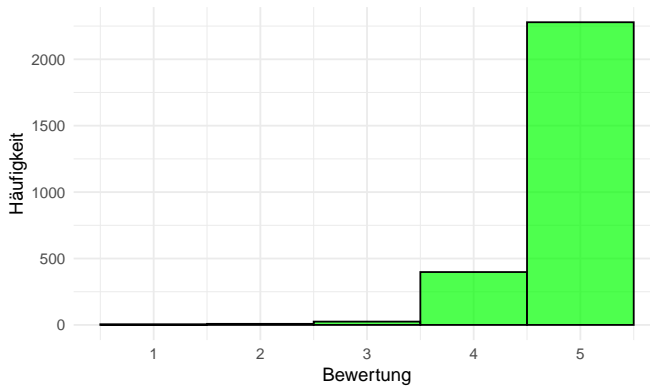
Verteilung des Preises pro Person in Vaud



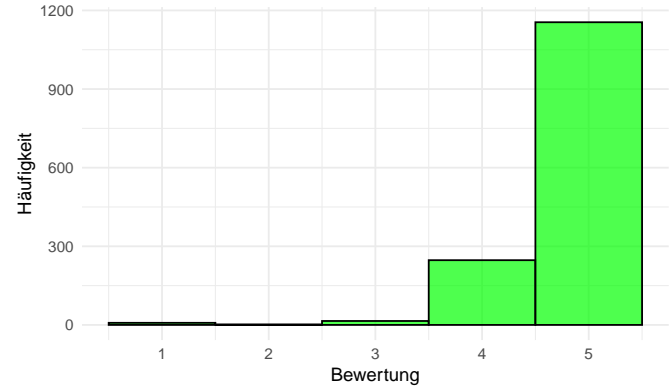
Boxplot des Preises pro Person in Vaud



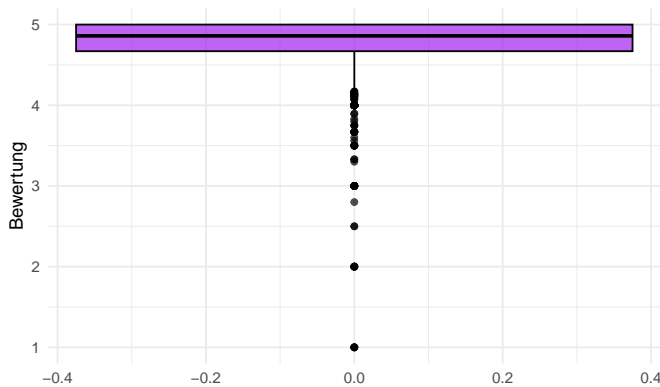
Verteilung der Bewertung in Vaud



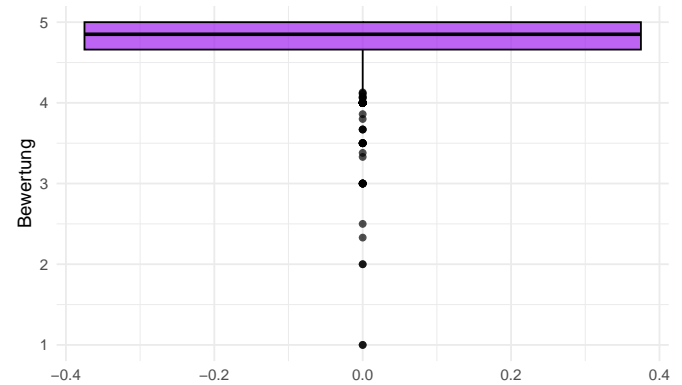
Verteilung der Bewertung in Geneva



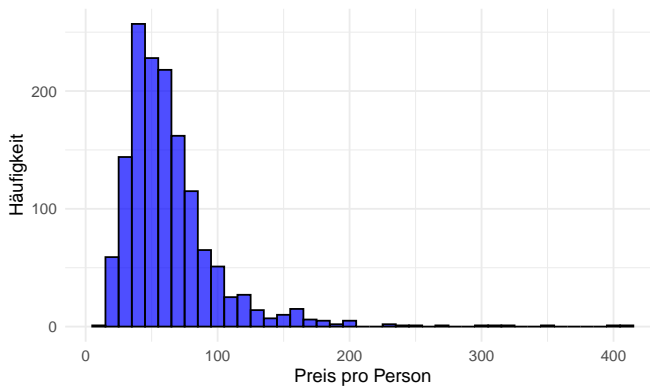
Boxplot der Bewertung in Vaud



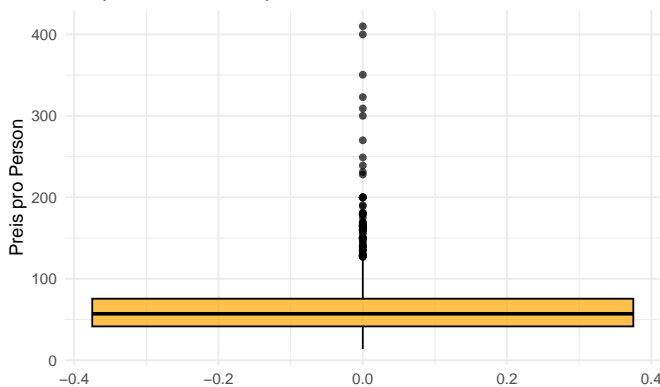
Boxplot der Bewertung in Geneva



Verteilung des Preises pro Person in Geneva



Boxplot des Preises pro Person in Geneva



### b) Korrelationsanalyse:

Die Korrelationsanalyse untersucht die Stärke und Richtung der Beziehung zwischen verschiedenen numerischen Variablen. Pearson-Korrelationskoeffizienten wurden berechnet, um zu quantifizieren, wie stark zwei Variablen miteinander variieren.

- **Ziel:** Identifikation von Zusammenhängen zwischen den Variablen, insbesondere zwischen den Bewertungen, Host-Informationen, geografischen Daten und dem Preis pro Person.
- **Massnahmen:** Berechnung von Korrelationskoeffizienten und Erstellung von Korrelationsmatrizen.

Zürich Korrelation:

Variable, Korrelation

calculated\_host\_listings\_count\_private\_rooms, 0.25

Vaud Korrelation:

Variable, Korrelation

calculated\_host\_listings\_count\_private\_rooms, 0.16

Genf Korrelation:

Variable, Korrelation

calculated\_host\_listings\_count\_private\_rooms, 0.26

### c) Lineare Regression:

Die lineare Regression wurde verwendet, um die Beziehung zwischen einer abhängigen Variable (Preis pro Person) und einer oder mehreren unabhängigen Variablen zu modellieren. Dies hilft, die Auswirkungen der unabhängigen Variablen auf die abhängige Variable zu quantifizieren und Vorhersagen zu treffen.

- **Ziel:** Untersuchung des Einflusses einzelner Variablen (z.B. Bewertungen, Host-Attribute, Entfernung zum

Stadtzentrum) auf den Preis pro Person.

- **Massnahmen:** Aufbau mehrerer linearer Regressionsmodelle, um den Einfluss spezifischer unabhängiger Variablen auf die abhängige Variable (Preis pro Person) zu analysieren.

Residuals:

Min	1Q	Median	3Q	Max
-57.91	-26.48	-11.50	12.19	605.12

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	4453.59	3636.89	1.225	0.221
longitude	66.41	55.90	1.188	0.235
latitude	-104.51	74.19	-1.409	0.159

Call:

```
lm(formula = formula, data = data)
```

Residual standard error: 49.08 on 1491 degrees of freedom  
Multiple R-squared: 0.002804, Adjusted R-squared: 0.001804  
F-statistic: 2.096 on 2 and 1491 DF, p-value: 0.159

Residuals:

Min	1Q	Median	3Q	Max
-53.83	-26.27	-11.45	11.82	605.20

Lineare Regression für Vaud - Bewertung :  
Zusammenfassung des Modells:

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	24.0091	22.2715	1.078	0.2812
review_scores_rating	0.3043	6.0424	0.050	0.9598
review_scores_cleanliness	12.9328	5.1802	2.497	0.0126 *
review_scores_communication	-3.7720	5.7870	-0.652	0.5141

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	44.535	8.137	5.473	0.000146
review_scores_rating	-1.091	2.574	-0.424	0.6766
review_scores_cleanliness	8.596	2.070	4.152	0.000146
review_scores_communication	-5.619	2.367	-2.373	0.0206

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Call:

```
lm(formula = formula, data = data)
```

Residual standard error: 28.29 on 2711 degrees of freedom  
Multiple R-squared: 0.008326, Adjusted R-squared: 0.007326  
F-statistic: 7.587 on 3 and 2711 DF, p-value: 4.7e-05

Residuals:

Min	1Q	Median	3Q	Max
-66.66	-25.83	-11.18	11.13	590.87

Lineare Regression für Vaud - Gastgeber :  
Zusammenfassung des Modells:

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	85.38588	7.5650	11.286	0.000146
host_is_superhostTRUE	4.61051	2.74373	1.680	0.0931 .
host_identity_verifiedTRUE	-14.39683	7.6863	-1.874	0.0611 .
host_listings_count	-0.12963	0.02044	-6.340	3.103e-06 ***

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	57.858072	2.932052	19.745	0.000146
host_is_superhostTRUE	0.154029	1.180702	0.129	0.8984
host_identity_verifiedTRUE	-4.994676	2.969186	-1.680	0.0931
host_listings_count	-0.006069	0.007536	-0.805	0.4211

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Call:

```
lm(formula = formula, data = data)
```

Residual standard error: 28.39 on 2711 degrees of freedom  
Multiple R-squared: 0.001344, Adjusted R-squared: 0.000344

```

F-statistic: 1.216 on 3 and 2711 DF,  p-value: 0.3022
(Intercept) 57.54004 5.55144 10.28
Linear Regression für Vaud - Lage : host_is_superhostTRUE 2.81315 2.22329 1.26
Zusammenfassung des Modells: host_identity_verifiedTRUE 6.05701 5.61251 1.17
host_listings_count 0.01829 0.02936 0.62
Call:
lm(formula = formula, data = data)
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residuals:
    Min       1Q   Median       3Q      Max
-53.440 -18.109  -5.071  10.540  254.935
Residual standard error: 37.33 on 1423 degrees of freedom
Multiple R-squared:  0.002167, Adjusted R-squared:  0.001999
F-statistic: 1.03 on 3 and 1423 DF,  p-value: 0.3816

Coefficients:
(Intercept) 404.690 198.159 2.042 0.0412 *
longitude -16.129 2.283 -7.064 2.056e-06 ***
latitude -5.223 4.131 -1.264 0.2106
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 28.14 on 2712 degrees of freedom
Multiple R-squared:  0.01877, Adjusted R-squared:  0.01805
F-statistic: 25.94 on 2 and 2712 DF,  p-value: 0.0001291

Linear Regression für Geneva - Lage :
Zusammenfassung des Modells:
Call:
lm(formula = formula, data = data)
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 37.35 on 1424 degrees of freedom
Multiple R-squared:  0.0007315, Adjusted R-squared:  0.0005633
F-statistic: 0.5212 on 2 and 1424 DF,  p-value: 0.4749

Coefficients:
(Intercept) -2192.77 2507.39 -0.875 0.382
longitude 20.18 39.64 0.509 0.611
latitude 46.17 54.36 0.849 0.396

```

**d) Visualisierung:**

Visualisierungen sind ein wesentliches Werkzeug, um Datenmuster und Zusammenhänge intuitiv darzustellen und zu verstehen. Verschiedene Arten von Plots wurden verwendet, um die Ergebnisse der deskriptiven Statistik, der Korrelationsanalyse und der Regressionsmodelle zu visualisieren.

- **Ziel:** Verständliche Darstellung der Ergebnisse und Identifikation von Mustern in den Daten.
- **Massnahmen:** Erstellung von Scatterplots, Boxplots, Barplots und Heatmaps zur Visualisierung der Beziehungen zwischen den Variablen.

**e) Predictive Analysis:**

Zusätzlich zu den deskriptiven und inferenziellen Analysen wurde eine prädiktive Analyse durchgeführt, um den Preis pro Person basierend auf den verfügbaren Merkmalen vorherzusagen. Hierfür wurde ein Regressionsmodell mit mehreren Prädiktoren erstellt und evaluiert.

**f) Datenvorbereitung:**

- **Feature Engineering:** Erstellung neuer Variablen, wie z.B. `distance_to_center`, und Transformation von kategorialen Variablen in numerische Werte.
- **Train-Test-Split:** Aufteilung der Daten in Trainings- und Testdatensätze, um die Modellleistung zu evaluieren.

**g) Modelltraining und -evaluation:**

- **Modelltraining:** Ein multiple lineares Regressionsmodell wurde auf den Trainingsdaten trainiert.

- **Modellbewertung:** Das Modell wurde auf den Testdaten evaluiert, um die Vorhersagegenauigkeit zu beurteilen.

## F. 6. Ergebnisse (statistische Ergebnisse, Zahlen, Diagramme)

### 1) Deskriptive Statistik:

### 2) Korrelationsanalyse:

#### a) Zürich:

#### • Positive Korrelationen:

- **price\_in\_usd (0.4542):** Höherer Gesamtpreis → höherer Preis pro Person.
- **calculated\_host\_listings\_count\_private\_rooms (0.2536):** Mehr private Zimmer → höherer Preis pro Person.
- **maximum\_maximum\_nights (0.1956) & maximum\_nights\_avg\_ntm (0.1956):** Längere maximale Aufenthaltsdauer → höherer Preis pro Person.
- **availability\_365 (0.1922):** Ganzjährige Verfügbarkeit → höherer Preis pro Person.

#### • Negative Korrelationen:

- **host\_listings\_count (-0.1772), calculated\_host\_listings\_count (-0.1909), host\_total\_listings\_count (-0.2006), calculated\_host\_listings\_count\_entire\_home (-0.2063):** Mehr Listings → niedrigerer Preis pro Person.
- **accommodates (-0.2531):** Mehr Plätze/Betten → niedrigerer Preis pro Person.

#### b) Vaud:

#### • Positive Korrelationen:

- **price\_in\_usd (0.2369):** Höherer Gesamtpreis → höherer Preis pro Person.
- **calculated\_host\_listings\_count\_private\_rooms (0.1654):** Mehr private Zimmer → höherer Preis pro Person.
- **availability\_90 (0.1133), availability\_365 (0.0997), availability\_30 (0.0891):** Höhere Verfügbarkeit → höherer Preis pro Person.

#### • Negative Korrelationen:

- **last\_review (-0.0820):** Länger zurückliegende letzte Bewertung → niedrigerer Preis pro Person.
- **reviews\_per\_month (-0.1043):** Mehr Bewertungen pro Monat → niedrigerer Preis pro Person.
- **longitude (-0.1349):** Östlichere Längengrade → niedrigerer Preis pro Person.
- **beds (-0.2090) & accommodates (-0.3130):** Mehr Betten/Plätze → niedrigerer Preis pro Person.

#### c) Geneva:

#### • Positive Korrelationen:

- **price\_in\_usd (0.3902):** Höherer Gesamtpreis → höherer Preis pro Person.
- **calculated\_host\_listings\_count\_private\_rooms (0.2662):** Mehr private Zimmer → höherer Preis pro Person.

- **availability\_30 (0.2246), availability\_60 (0.2184), availability\_90 (0.1991):** Höhere Verfügbarkeit → höherer Preis pro Person.

#### • Negative Korrelationen:

- **maximum\_minimum\_nights (-0.0907):** Höhere minimale Aufenthaltsdauer → niedrigerer Preis pro Person.
- **number\_of\_reviews\_130d (-0.0957):** Mehr Bewertungen in den letzten 30 Tagen → niedrigerer Preis pro Person.
- **minimum\_maximum\_nights (-0.0966):** Höhere minimale maximale Aufenthaltsdauer → niedrigerer Preis pro Person.
- **reviews\_per\_month (-0.0976):** Mehr Bewertungen pro Monat → niedrigerer Preis pro Person.
- **accommodates (-0.2185):** Mehr Plätze/Betten → niedrigerer Preis pro Person.

#### d) Gesamtschlussfolgerungen:

Die Analyse der Korrelationen von `price_per_person` mit verschiedenen Variablen in den Städten Zürich, Vaud und Geneva liefert wertvolle Einblicke in die Faktoren, die den Preis pro Person beeinflussen. Hier sind die ausführlicheren Schlussfolgerungen:

**Höherer Gesamtpreis (`price_in_usd`):** In allen drei Städten zeigt sich eine positive Korrelation zwischen dem Gesamtpreis und dem Preis pro Person. Dies bedeutet, dass teurere Unterkünfte auch pro Person mehr kosten. Dies ist intuitiv, da eine höhere Grundmiete oft höhere Pro-Kopf-Kosten nach sich zieht.

#### • Mehr private Zimmer (`calculated_host_listings_count_private_rooms`):

Ebenfalls in allen drei Städten korreliert die Anzahl der privaten Zimmer im Angebot des Hosts positiv mit dem Preis pro Person. Dies könnte darauf hinweisen, dass Hosts, die mehr private Zimmer anbieten, tendenziell höhere Preise verlangen, möglicherweise aufgrund der höheren Privatsphäre und Exklusivität.

**Kurzfristige und langfristige Verfügbarkeit:** In Geneva und Vaud zeigt die Verfügbarkeit der Unterkunft über verschiedene Zeiträume (30, 60, 90 Tage und das ganze Jahr) eine positive Korrelation mit dem Preis pro Person. Dies deutet darauf hin, dass Unterkünfte, die sowohl kurzfristig als auch langfristig verfügbar sind, tendenziell höher bepreist werden. Es könnte sein, dass diese Unterkünfte eine höhere Nachfrage bedienen oder flexiblere Buchungsoptionen bieten, die höher bewertet werden.

**Häufigkeit der Bewertungen (`reviews_per_month`, `number_of_reviews_130d`):** In Vaud und Geneva gibt es eine negative Korrelation zwischen der Anzahl der Bewertungen pro Monat und dem Preis pro Person. Dies könnte darauf hindeuten, dass Unterkünfte, die häufiger bewertet werden, tendenziell günstiger sind. Diese Unterkünfte könnten eine breitere Zielgruppe ansprechen, einschließlich preisbewusster Reisender, was zu häufigeren Buchungen und somit zu mehr Bewertungen führt.

**Zeit seit der letzten Bewertung (`last_review`):** In



Vaud zeigt sich eine leichte negative Korrelation zwischen der Zeit seit der letzten Bewertung und dem Preis pro Person. Dies könnte darauf hinweisen, dass aktuellere Bewertungen mit höheren Preisen pro Person verbunden sind, möglicherweise weil aktuelle Bewertungen das Vertrauen und die Attraktivität der Unterkunft erhöhen.

- **Anzahl der Listings (host\_listings\_count, calculated\_host\_listings\_count, host\_total\_listings\_count, calculated\_host\_listings\_count\_entire\_homes):**

In Zürich zeigen mehrere Variablen, die die Anzahl der Listings des Hosts messen, eine negative Korrelation mit dem Preis pro Person. Dies deutet darauf hin, dass Hosts mit vielen Listings ihre Preise wettbewerbsfähiger gestalten müssen, um die Nachfrage zu halten. Größere Host-Operationen könnten Skaleneffekte nutzen, um niedrigere Preise anzubieten.

- **Anzahl der Betten und verfügbare Plätze (beds, accommodates):** In allen drei Städten zeigen sich negative Korrelationen zwischen der Anzahl der Betten bzw. der verfügbaren Plätze und dem Preis pro Person. Dies weist darauf hin, dass größere Unterkünfte tendenziell niedrigere Pro-Kopf-Kosten haben. Dies könnte auf Skaleneffekte oder die Notwendigkeit, größere Gruppen anzuziehen, zurückzuführen sein.
- **Geografische Koordinaten (longitude, latitude):** In Vaud zeigt der Längengrad eine signifikante negative Korrelation mit dem Preis pro Person, was auf eine geografische Preisstruktur hindeuten könnte. Dies könnte bedeuten, dass östlichere Lagen tendenziell günstiger sind. In Zürich und Geneva sind die geografischen Koordinaten nicht signifikant.

### 3) Lineare Regression:

#### a) Bewertung der Airbnb-Unterkünfte:

##### Hypothesen:

- Nullhypothese (H0): Die Höhe der Gesamtbewertung (review\_scores\_rating) hat keinen signifikanten Einfluss auf den Preis pro Person der Unterkunft.
- Alternativhypothese (H1): Die Höhe der Gesamtbewertung (review\_scores\_rating) hat einen signifikanten Einfluss auf den Preis pro Person der Unterkunft.

##### Ergebnisse:

##### Zürich:

- review\_scores\_rating: Nicht signifikant (p = 0.9598)
- review\_scores\_cleanliness: Signifikant (p = 0.0126) mit positivem Einfluss (Estimate = 12.9328)
- review\_scores\_communication: Nicht signifikant (p = 0.5146)
- **Interpretation:** Die Sauberkeitsbewertung beeinflusst den Preis pro Person positiv und signifikant, während die Gesamtbewertung und die Kommunikationsbewertung keinen signifikanten Einfluss haben.

##### Vaud:

- review\_scores\_rating: Nicht signifikant (p = 0.6718)

- review\_scores\_cleanliness: Signifikant (p = 3.40e-05) mit positivem Einfluss (Estimate = 8.596)
- review\_scores\_communication: Signifikant (p = 0.0177) mit negativem Einfluss (Estimate = -5.619)
- **Interpretation:** In Vaud hat die Sauberkeitsbewertung einen signifikanten positiven Einfluss, während die Kommunikationsbewertung einen signifikanten negativen Einfluss auf den Preis pro Person hat. Die Gesamtbewertung ist nicht signifikant.

##### Geneva:

- review\_scores\_rating: Nicht signifikant (p = 0.1512)
- review\_scores\_cleanliness: Signifikant (p = 0.0002) mit positivem Einfluss (Estimate = 14.473)
- review\_scores\_communication: Grenzwertig signifikant (p = 0.0809) mit negativem Einfluss (Estimate = -6.894)
- **Interpretation:** In Geneva hat die Sauberkeitsbewertung einen signifikanten positiven Einfluss auf den Preis pro Person, während die Gesamtbewertung keinen signifikanten Einfluss hat. Die Kommunikationsbewertung ist grenzwertig signifikant negativ.

#### b) Gastgeber der Airbnb-Unterkunft:

##### Hypothesen:

- Nullhypothese (H0): Die Attribute des Hosts wie "host\_is\_superhost", "host\_identity\_verified" und "host\_listings\_count" haben keinen signifikanten Einfluss auf den Preis pro Person der Unterkunft.
- Alternativhypothese (H1): Die Attribute des Hosts wie "host\_is\_superhost", "host\_identity\_verified" und "host\_listings\_count" haben einen signifikanten Einfluss auf den Preis pro Person der Unterkunft.

##### Ergebnisse:

##### Zürich:

- host\_is\_superhost: Grenzwertig signifikant (p = 0.0931)
- host\_identity\_verified: Grenzwertig signifikant (p = 0.0611) mit negativem Einfluss (Estimate = -14.39683)
- host\_listings\_count: Signifikant (p = 3.03e-10) mit negativem Einfluss (Estimate = -0.12963)
- **Interpretation:** In Zürich hat die Anzahl der Listings eines Hosts einen signifikant negativen Einfluss auf den Preis pro Person. Der Status "Superhost" und die Verifizierung des Hosts sind grenzwertig signifikant.

##### Vaud:

- host\_is\_superhost: Nicht signifikant (p = 0.8962)
- host\_identity\_verified: Grenzwertig signifikant (p = 0.0927) mit negativem Einfluss (Estimate = -4.994676)
- host\_listings\_count: Nicht signifikant (p = 0.4207)
- **Interpretation:** In Vaud sind keine der Host-Attribute signifikant, jedoch ist die Verifizierung des Hosts grenzwertig signifikant negativ.

##### Geneva:

- `host_is_superhost`: Nicht signifikant ( $p = 0.206$ )
- `host_identity_verified`: Nicht signifikant ( $p = 0.281$ )
- `host_listings_count`: Nicht signifikant ( $p = 0.533$ )
- **Interpretation:** In Geneva haben die Host-Attribute keinen signifikanten Einfluss auf den Preis pro Person.

c) *Lage des Airbnb:*

#### Hypothesen:

- Nullhypothese (H0): Die Entfernung zum Stadtzentrum (berechnet durch geografische Koordinaten) hat keinen signifikanten Einfluss auf den Preis pro Person der Unterkunft.
- Alternativhypothese (H1): Die Entfernung zum Stadtzentrum (berechnet durch geografische Koordinaten) hat einen signifikanten Einfluss auf den Preis pro Person der Unterkunft.

#### Ergebnisse:

##### Zürich:

- `longitude`: Nicht signifikant ( $p = 0.235$ )
- `latitude`: Nicht signifikant ( $p = 0.159$ )
- **Interpretation:** Die geografischen Koordinaten haben keinen signifikanten Einfluss auf den Preis pro Person in Zürich.

##### Vaud:

- `longitude`: Signifikant ( $p = 2.05e-12$ ) mit negativem Einfluss (Estimate = -16.129)
- `latitude`: Nicht signifikant ( $p = 0.2062$ )
- **Interpretation:** In Vaud hat der Längengrad einen signifikanten negativen Einfluss auf den Preis pro Person, während der Breitengrad keinen signifikanten Einfluss hat.

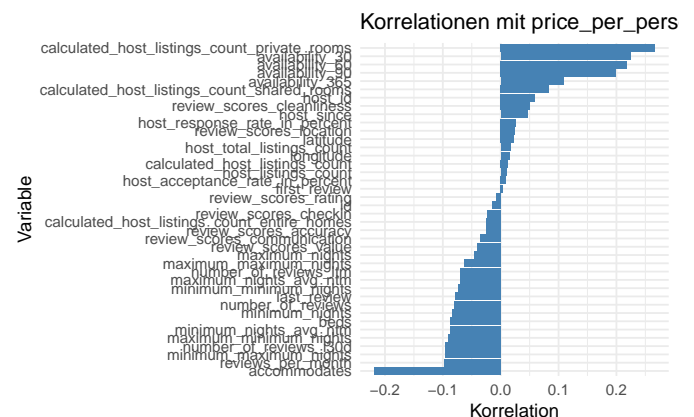
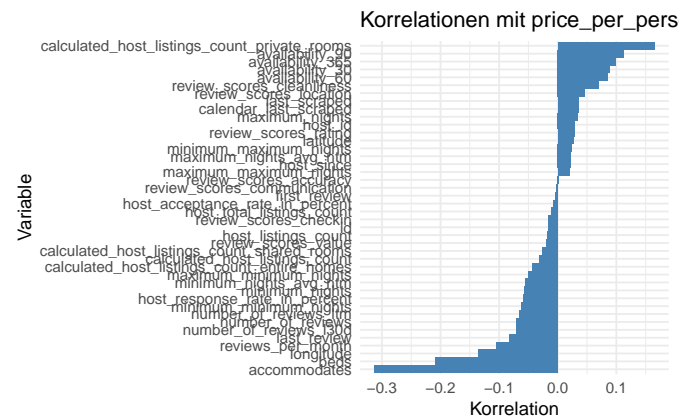
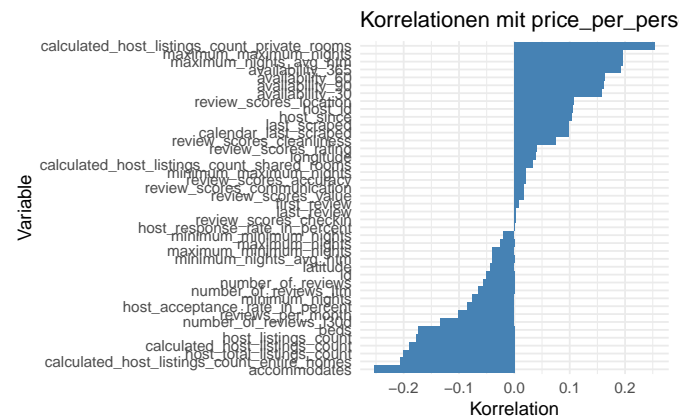
##### Geneva:

- `longitude`: Nicht signifikant ( $p = 0.611$ )
- `latitude`: Nicht signifikant ( $p = 0.396$ )
- **Interpretation:** In Geneva haben die geografischen Koordinaten keinen signifikanten Einfluss auf den Preis pro Person.

#### 4) Zusammenfassung::

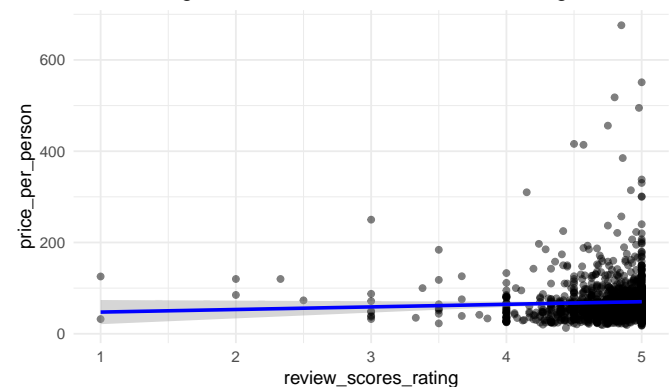
- **Bewertung:** Die Sauberkeitsbewertung hat in allen Städten einen signifikanten positiven Einfluss auf den Preis pro Person. Die Gesamtbewertung hat keinen signifikanten Einfluss. Die Kommunikationsbewertung hat in Vaud einen signifikanten negativen Einfluss.
- **Gastgeber:** In Zürich hat die Anzahl der Listings eines Hosts einen signifikant negativen Einfluss. In Vaud und Geneva sind die Host-Attribute nicht signifikant.
- **Lage:** In Vaud hat der Längengrad einen signifikanten negativen Einfluss auf den Preis pro Person. In Zürich und Geneva haben die geografischen Koordinaten keinen signifikanten Einfluss.

5) **Visualisierung:** Die folgenden Plots zeigen die Korrelationen des Preis pro Person (`price_per_person`) in den verschiedenen Städten. Anhand dieser Diagramme lässt sich leicht erkennen, dass einige Korrelationen viel stärker positiv oder negativ ausfallen. Diese starken Korrelationen sollten genauer untersucht werden, um ihre Ursachen und Auswirkungen besser zu verstehen.

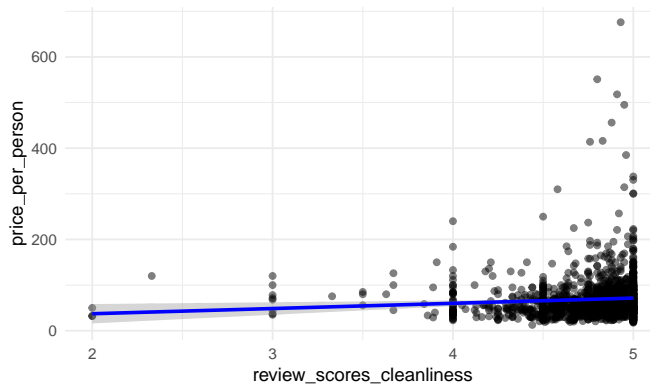


#### a) Lineare Regression für Zürich:

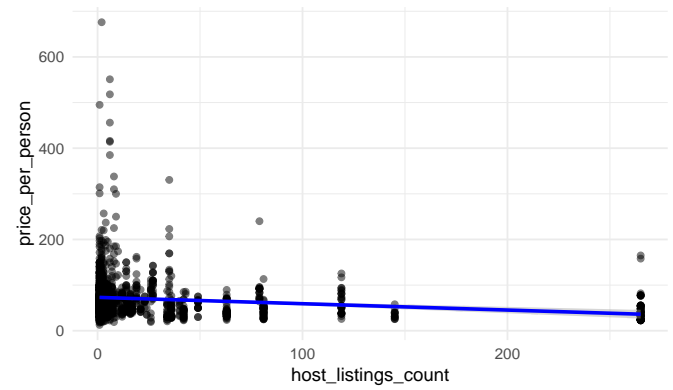
##### Lineare Regression für Zürich – Gesamtbewertung



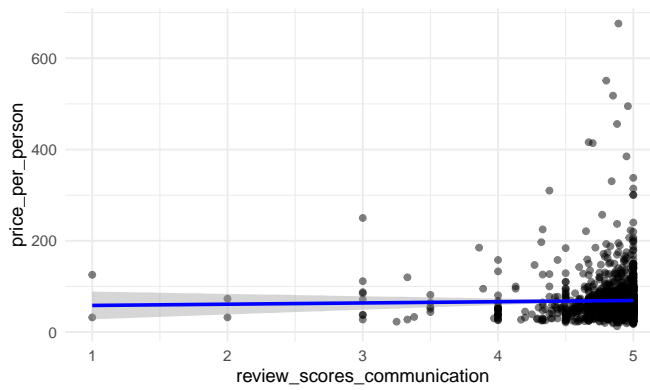
Lineare Regression für Zürich – Sauberkeit



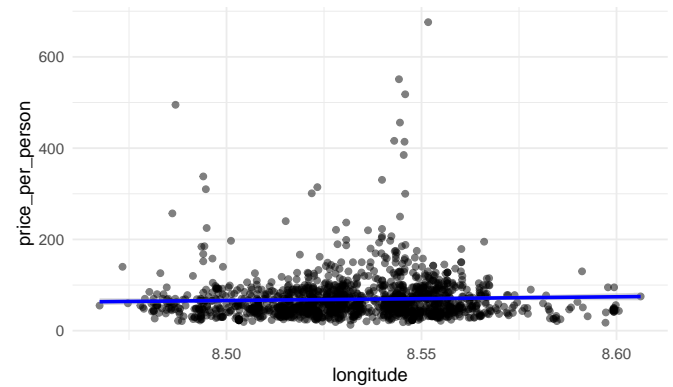
Lineare Regression für Zürich – Anzahl der Listings



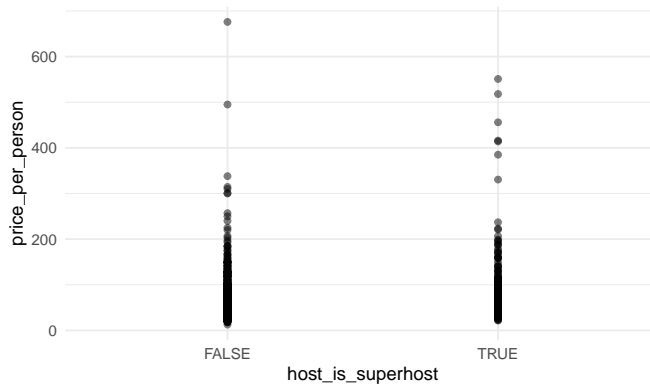
Lineare Regression für Zürich – Kommunikation



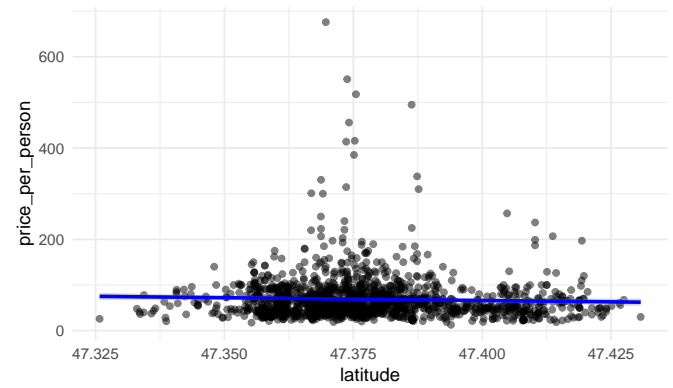
Lineare Regression für Zürich – Längengrad



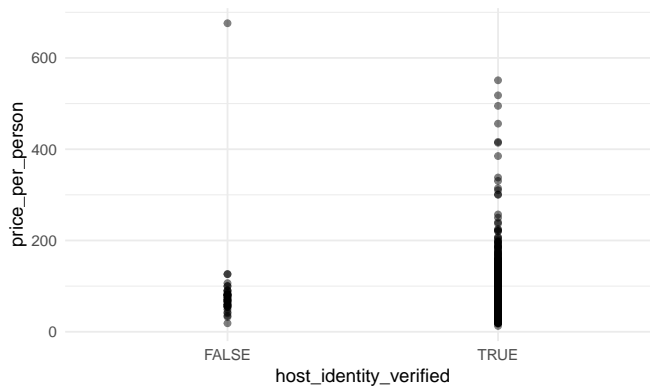
Lineare Regression für Zürich – Superhost



Lineare Regression für Zürich – Breitengrad

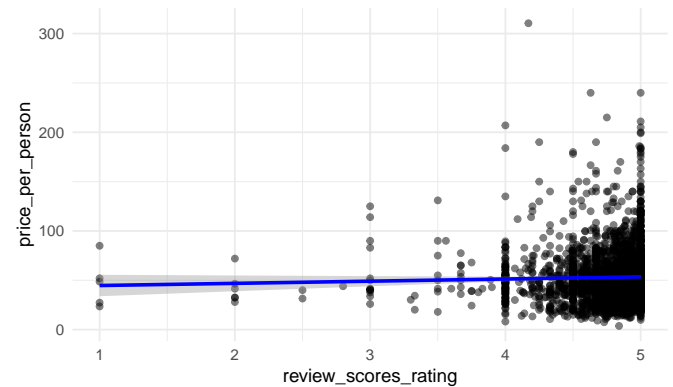


Lineare Regression für Zürich – Identität verifiziert

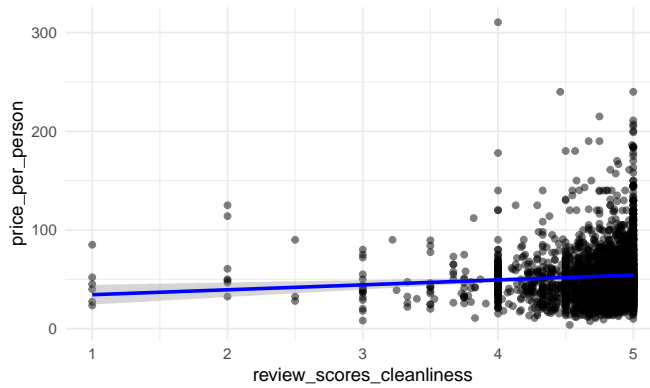


### b) Lineare Regression für Vaud:

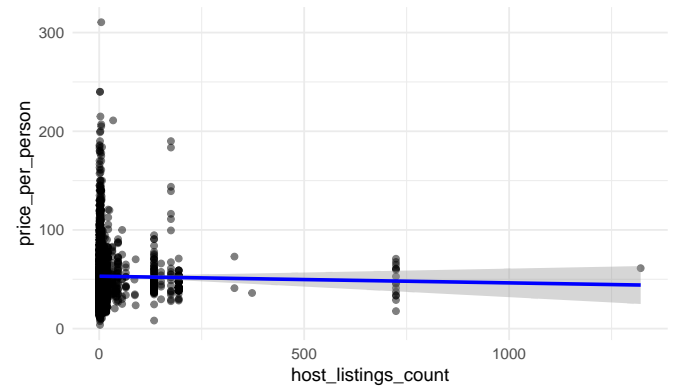
Lineare Regression für Vaud – Gesamtbewertung



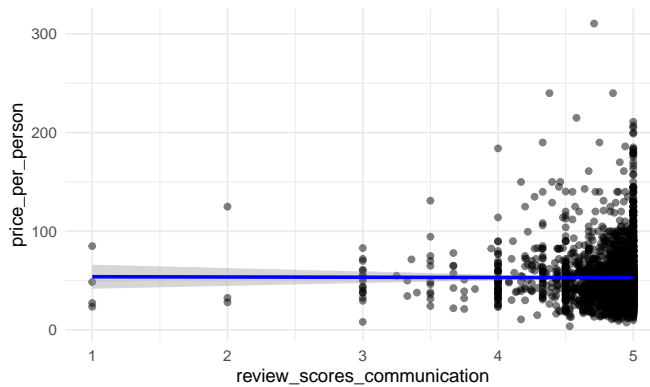
Lineare Regression für Vaud – Sauberkeit



Lineare Regression für Vaud – Anzahl der Listings

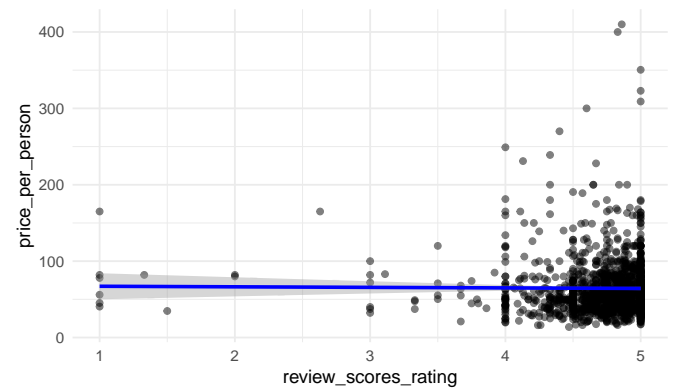


Lineare Regression für Vaud – Kommunikation

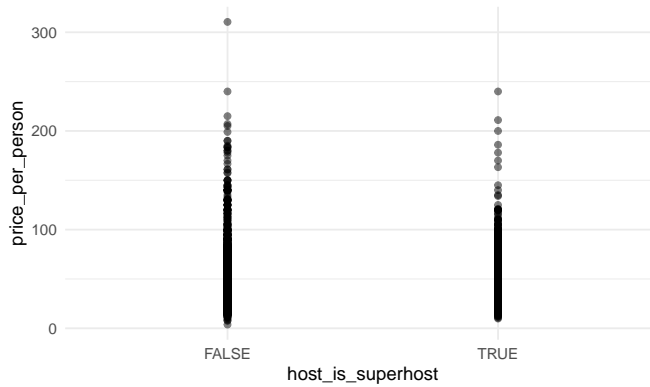


c) Lineare Regression für Geneva:

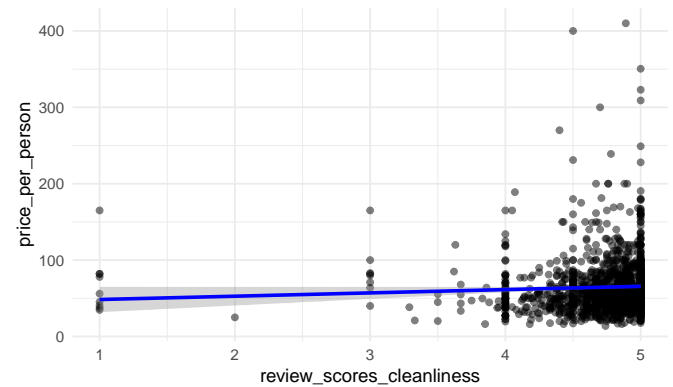
Lineare Regression für Geneva – Gesamtbewertung



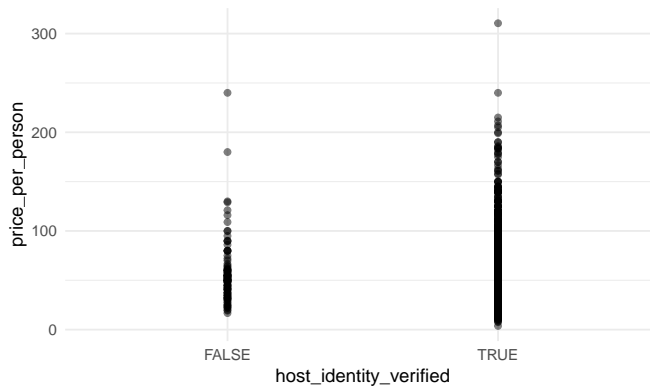
Lineare Regression für Vaud – Superhost



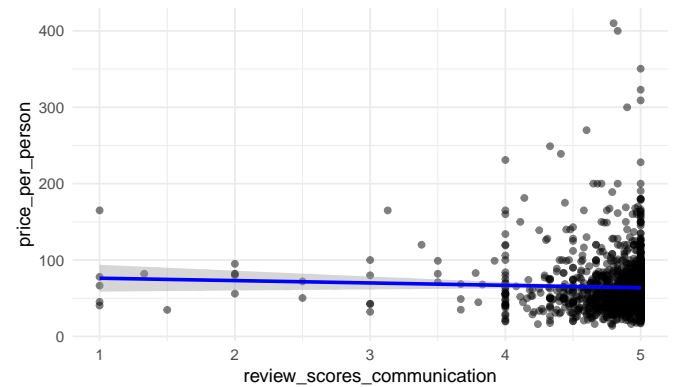
Lineare Regression für Geneva – Sauberkeit



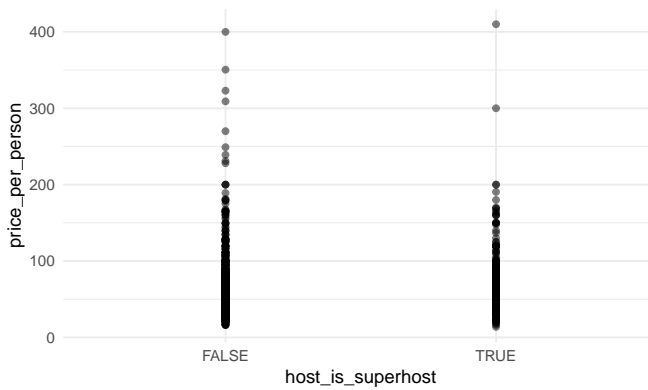
Lineare Regression für Vaud – Identität verifiziert



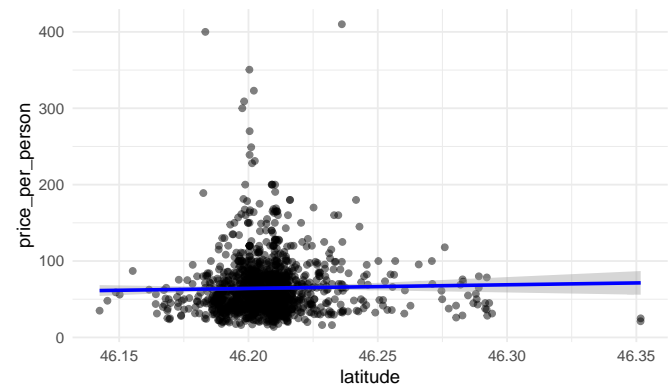
Lineare Regression für Geneva – Kommunikation



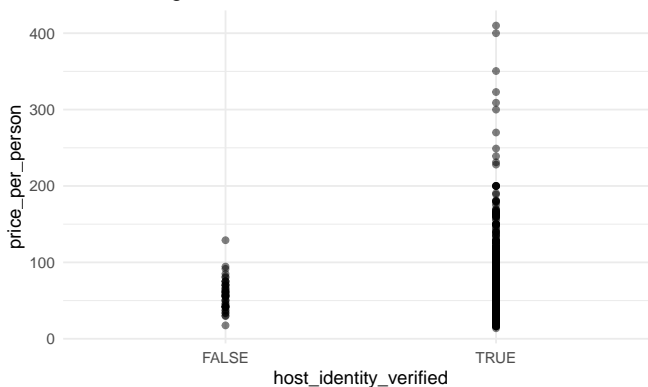
Lineare Regression für Geneva – Superhost



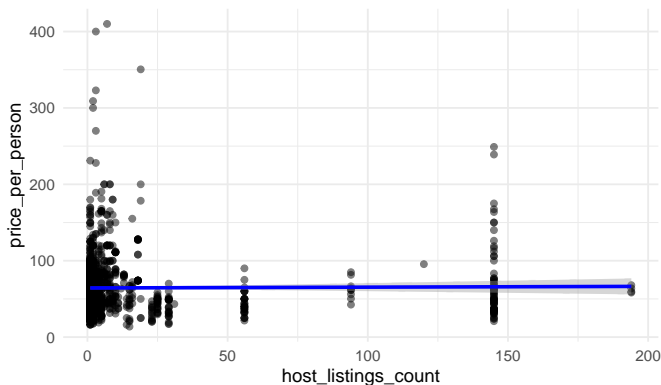
Lineare Regression für Geneva – Breitengrad



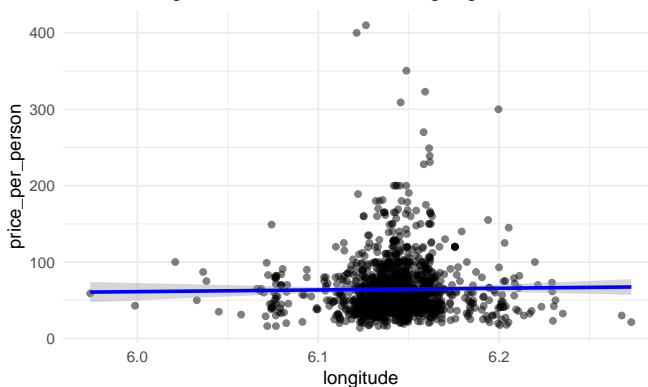
Lineare Regression für Geneva – Identität verifiziert



Lineare Regression für Geneva – Anzahl der Listings



Lineare Regression für Geneva – Längengrad



#### d) Schlussfolgerung:

Die verschiedenen Grafiken und deren lineare Regressionen liefern interessante Einblicke in die Beziehung zwischen den untersuchten Variablen und dem Preis pro Person. Allerdings zeigen die Ergebnisse, dass viele der unabhängigen Variablen nur einen geringen Einfluss auf den Preis pro Person haben.

Zusammenfassend lässt sich sagen, dass die meisten untersuchten Variablen keinen starken Einfluss auf den Preis pro Person haben. Dies könnte darauf hindeuten, dass andere, nicht berücksichtigte Faktoren eine größere Rolle spielen oder dass die Märkte in diesen Städten relativ homogen sind. Weitere Untersuchungen könnten sich auf zusätzliche Variablen oder nicht-lineare Modelle konzentrieren, um potenzielle Einflussfaktoren besser zu identifizieren.

#### 6) Predictive Analysis:

```
Error: <text>:1:15: unerwartetes Symbol
1: set.seed(123) trainIndex
      ^
```

Die vorliegende Predictive Analyse zielt darauf ab, den Trend des Preises pro Person einer Unterkunft aufzuzeigen und zu untersuchen, ob mithilfe der Bewertung und den wichtigsten Eigenschaften einer Unterkunft eine Prognose über den Preis gemacht werden kann. Dabei wurden Verfügbarkeit (availability\_365), Bewertung (review\_scores\_rating) und Anzahl der Bewertungen (number\_of\_reviews) als Prädiktoren herangezogen.

Die Modellbewertung zeigt folgende Ergebnisse:

- **RMSE: 41.90135** – Der durchschnittliche quadratische Fehler beträgt 41.90 USD, was auf erhebliche Abweichungen zwischen den vorhergesagten und tatsächlichen Preisen hinweist.
- **MAE: 27.91819** – Der mittlere absolute Fehler beträgt 27.92 USD, was ebenfalls auf erhebliche Abweichungen hinweist.
- **R<sup>2</sup>: -0.01782196** – Ein negativer R<sup>2</sup>-Wert deutet darauf hin, dass das Modell die Variation des Preises pro Person nicht gut erklärt und kaum besser als zufällige Vorhersagen ist.

Die Merkmalswichtigkeit zeigt, dass die Verfügbarkeit über 365 Tage (IncNodePurity = 833646.7) der wichtigste Prädiktor ist, gefolgt von der Anzahl der Bewertungen (IncNodePurity = 718716.6) und der Bewertung der Unterkunft (IncNodePurity = 488147.3). Dies zeigt, dass die Verfügbarkeit der Unterkunft,

die Anzahl der Bewertungen und die Bewertungen selbst wesentliche Faktoren bei der Preisgestaltung sind.

#### G. 7. Schlussfolgerung (Beantwortung der Frage)

In diesem Kapitel fassen wir die wichtigsten Ergebnisse unserer Analyse zusammen und beantworten die eingangs gestellte Forschungsfrage: “Welche Eigenschaften einer Airbnb-Unterkunft ziehen Gäste an und ermöglichen es, einen höheren Preis pro Apartment zu erzielen?”

1) *Bewertung der Airbnb-Unterkünfte:* Unsere Analyse zeigte, dass die Sauberkeitsbewertung (review\_scores\_cleanliness) einen signifikanten positiven Einfluss auf den Preis pro Person in allen drei untersuchten Städten (Zürich, Vaud und Genf) hat. Dies deutet darauf hin, dass Gäste bereit sind, mehr für Unterkünfte zu zahlen, die als besonders sauber bewertet wurden. Die Gesamtbewertung (review\_scores\_rating) und die Kommunikationsbewertung (review\_scores\_communication) hatten hingegen keinen einheitlich signifikanten Einfluss auf den Preis.

2) *Gastgeber der Airbnb-Unterkünfte:* Die Untersuchung der Gastgebermerkmale ergab, dass in Zürich die Anzahl der Listings eines Hosts (host\_listings\_count) einen signifikant negativen Einfluss auf den Preis pro Person hat. Dies könnte darauf hindeuten, dass Gastgeber mit vielen Angeboten möglicherweise niedrigere Preise anbieten, um die Auslastung zu maximieren. Der Status als Superhost (host\_is\_superhost) und die Verifizierung des Hosts (host\_identity\_verified) hatten hingegen keinen konsistenten signifikanten Einfluss auf den Preis.

3) *Lage der Airbnb-Unterkünfte:* Die geografische Analyse zeigte, dass in Vaud der Längengrad (longitude) einen signifikant negativen Einfluss auf den Preis pro Person hat. Dies könnte auf spezifische regionale Preisunterschiede hinweisen. In Zürich und Genf hatten die geografischen Koordinaten jedoch keinen signifikanten Einfluss auf den Preis.

#### 4) Predictive Analysis: NOCH ERLEDIGEN

Unsere prädiktive Analyse zeigte, dass die Verfügbarkeit der Unterkunft über 365 Tage (availability\_365), die Anzahl der Bewertungen (number\_of\_reviews) und die Sauberkeitsbewertung (review\_scores\_cleanliness) wesentliche Prädiktoren für den Preis pro Person sind. Das entwickelte Vorhersagemodell zeigte jedoch eine begrenzte Vorhersagegenauigkeit, was auf die Komplexität und die vielen Einflussfaktoren des Preises hinweist.

#### NOCH ERLEDIGEN

5) *Empfehlungen für Gastgeber:* Auf Basis unserer Ergebnisse können folgende Empfehlungen für Gastgeber abgeleitet werden, um den Preis pro Person ihrer Unterkünfte zu maximieren:

1. **Sauberkeit verbessern:** Da die Sauberkeitsbewertung einen signifikanten Einfluss auf den Preis hat, sollten Gastgeber sicherstellen, dass ihre Unterkünfte stets sauber und gepflegt sind, um positive Bewertungen in diesem Bereich zu erhalten.
2. **Anzahl der Listings optimieren:** Gastgeber mit vielen Listings sollten ihre Preisstrategie überprüfen und möglicherweise höhere Preise für besonders exklusive oder gut bewertete Unterkünfte festlegen.

3. **Geografische Preisstrategien anwenden:** Insbesondere in Regionen wie Vaud sollten Gastgeber die lokale Preisstruktur berücksichtigen und ihre Preise entsprechend anpassen.

6) *Beantwortung der Forschungsfrage:* Zusammenfassend lässt sich sagen, dass die Sauberkeitsbewertung einer Unterkunft, die Verfügbarkeit der Unterkunft und die Anzahl der Listings eines Hosts die wichtigsten Faktoren sind, die den Preis pro Person einer Airbnb-Unterkunft beeinflussen. Gastgeber sollten daher auf diese Merkmale achten, um ihre Einnahmen zu maximieren.

7) *Weiterführende Forschung:* Für zukünftige Untersuchungen wäre es sinnvoll, zusätzliche Daten über längere Zeiträume zu sammeln, um den Einfluss saisonaler Ereignisse und Trends besser zu verstehen. Zudem könnten nicht-lineare Modelle oder maschinelle Lernverfahren eingesetzt werden, um die Komplexität der Einflussfaktoren besser abzubilden und genauere Vorhersagen zu ermöglichen.

#### 8. Referenzen

- [1] Inside Airbnb, “Inside airbnb: Adding data to the debate,” 27-Dec-2023. [Online]. Available: <https://insideairbnb.com/get-the-data/>. [Accessed: 30-May-2024]
- [2] Inside Airbnb, “Inside airbnb data dictionary: Data dictionary for listings.csv detailed file.” Aug-2022 [Online]. Available: <https://docs.google.com/spreadsheets/d/1iWCNJcSutYqpULSQHINyGInUvHg2BoUGoNRIGa6Szcz4/edit#gid=1322284596>. [Accessed: 30-May-2024]