

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

Cedric Gisler & Jovan Pajic

I. 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.

II. 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 Bewertung hat keinen Einfluss auf die Höhe des Preises pro Person der Unterkunft.
- **Alternativhypothese (H1):** Die Höhe der Bewertung hat einen Einfluss auf die Höhe des Preises pro Person der Unterkunft.

2. Gastgeber der Airbnb-Unterkunft:

- **Nullhypothese (H0):** Die Informationen über den Host, wie "host_is_superhost", "host_identify_verified" und "host_listings_count" haben keinen Einfluss auf die Höhe des Preises pro Person der Unterkunft.

- **Alternativhypothese (H1):** Die Informationen über den Host, wie "host_is_superhost", "host_identify_verified" und "host_listings_count" haben einen Einfluss auf die Höhe des Preises pro Person der Unterkunft.

3. Lage des Airbnb:

- **Nullhypothese (H0):** Die Nähe zum Stadtzentrum hat keinen Einfluss auf die Höhe des Preises pro Person der Unterkunft.
- **Alternativhypothese (H1):** Die Nähe zum Stadtzentrum hat einen Einfluss auf die Höhe des Preises pro Person der Unterkunft.

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.

III. 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.

IV. 4. DATENQUALITÄT (ANALYSE IM HINBLICK AUF DATENQUALITÄTSASPEKTE)

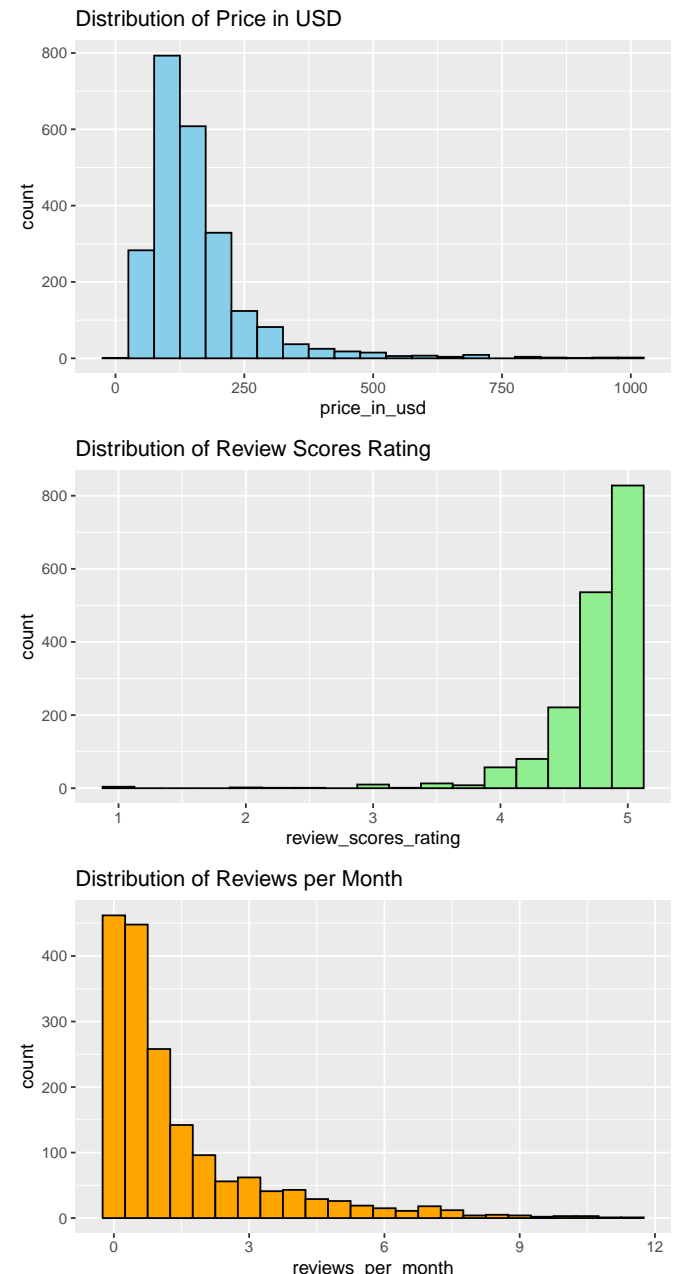
Nachdem wir die Zusammenfassung der Daten betrachtet und eine detaillierte Analyse der Dataframes durchgeführt haben, haben wir die folgenden Datenanpassungen vorgenommen:

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.

Um die Daten nach der Bereinigung zu überprüfen und einen Überblick zu erhalten, führen wir eine standardmässige Datenanalysen durch am Beispiel **df_zuerich_cleansed**.

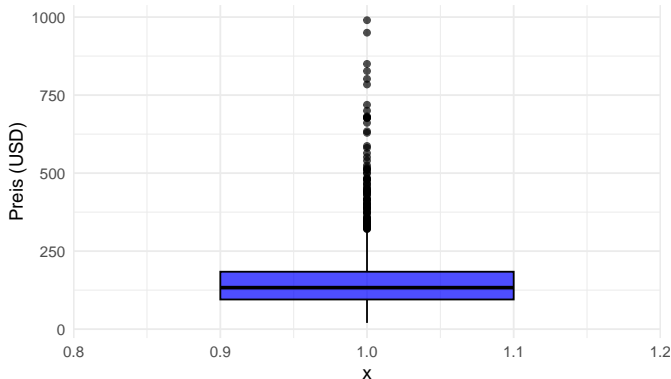


Da der Preis stark variiert, analysieren wir die Möglichkeit, den Preis pro Person zu berechnen. Dabei fällt auf, dass der Preis bei “Entire home/apt” für die maximale Anzahl der Gäste (**accommodates**) berechnet wird, was sinnvoll ist, da die gesamte Unterkunft gebucht wird. Bei “Private room” und “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.

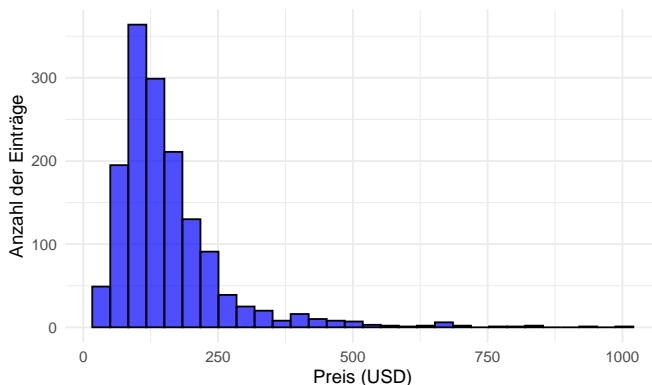
V. 5. DATENANALYSE (INFORMATIONEN ZUR DATENSTRUKTUR, ORGANISATION UND ZU DEN FÜR DIE ANALYSE VERWENDETEN METHODEN)

Die bereinigten Datensets der verschiedenen Orten sind alle gleich aufgebaut. Da wir den Preis der einzelnen Airbnb Appartment anschauen möchten ist dies unser wichtigster Wert:

Verteilung der Preise in USD



Verteilung der Preise in USD



Nun stellt sich die Frage welche anderen Eigenschaften die grösste Auswirkung auf den Preis haben. Dazu gilt es herauszufinden wie die Korrelationen zwischen dem Preis pro USD und den anderen Attributen sind:

Die Korrelationen der verschiedenen Variablen mit dem Preis (**price_in_usd**) im Datensatz können wichtige Einsichten bieten, welche Faktoren den Preis beeinflussen. Hier ist eine Analyse der signifikanten positiven und negativen Korrelationen:

A. Positive Korrelationen:

1. **accommodates** (0.449): Es besteht eine moderate positive Korrelation zwischen der Anzahl der Gäste, die eine Unterkunft aufnehmen kann, und dem Preis. Dies deutet darauf hin, dass grössere Unterkünfte, die mehr Gäste beherbergen können, in der Regel teurer sind.
2. **beds** (0.329): Eine ähnliche positive Korrelation gibt es zwischen der Anzahl der Betten und dem Preis, was darauf hindeutet, dass mehr Betten oft höhere Preise bedeuten, was auch mit der Grösse der Unterkunft zusammenhängen kann.
3. **availability_365** (0.157): Längere Verfügbarkeit im Laufe eines Jahres korreliert leicht positiv mit höheren Preisen. Dies könnte bedeuten, dass Unterkünfte, die

seltener verfügbar sind, zu höheren Preisen angeboten werden.

4. **last_scraped** (0.308), **calendar_last_scraped** (0.308): Diese Korrelationen deuten darauf hin, dass die Zeitpunkte der Datenerfassung mit den Preisänderungen

B. Negative Korrelationen:

1. **latitude** (-0.097): Es gibt eine leichte negative Korrelation zwischen der geographischen Breite und dem Preis. Je weiter nördlich die Unterkunft liegt, desto geringer könnte der Preis sein, was auf regionale Preisunterschiede in der Stadt oder der Umgebung hinweisen könnte.
2. **minimum_nights** (-0.053), **first_review** (-0.053): Längere Mindestaufenthalte und frühere erste Bewertungen korrelieren leicht negativ mit dem Preis. Dies könnte darauf hinweisen, dass preisgünstigere Unterkünfte möglicherweise längere Aufenthalte erfordern oder schon länger auf dem Markt sind.
3. **calculated_host_listings_count_private_rooms** (-0.103): Eine höhere Anzahl von Inseraten, die private Zimmer betreffen, korreliert leicht negativ mit dem Preis, was darauf hinweisen könnte, dass Gastgeber mit mehreren Einträgen möglicherweise günstigere Preise anbieten, um wettbewerbsfähig zu bleiben.

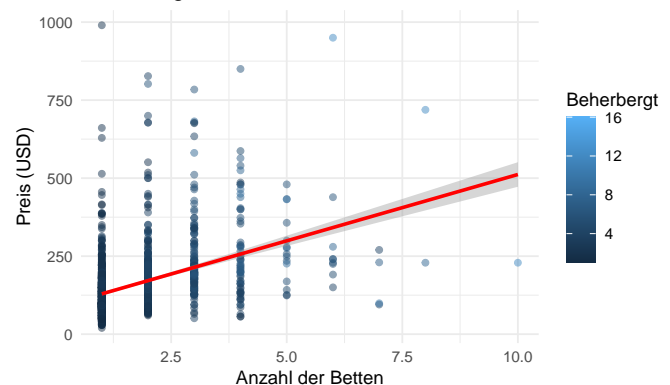
C. Interpretation

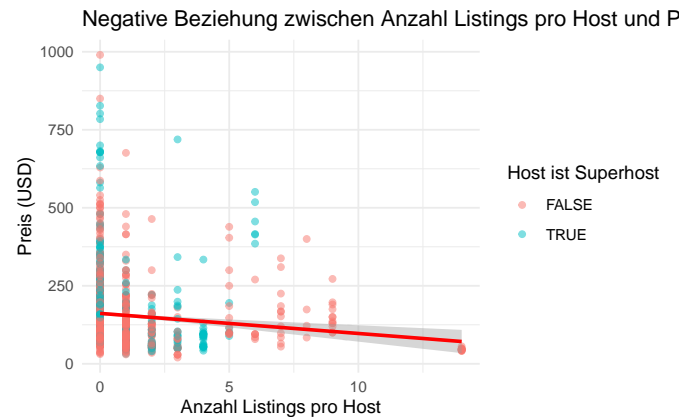
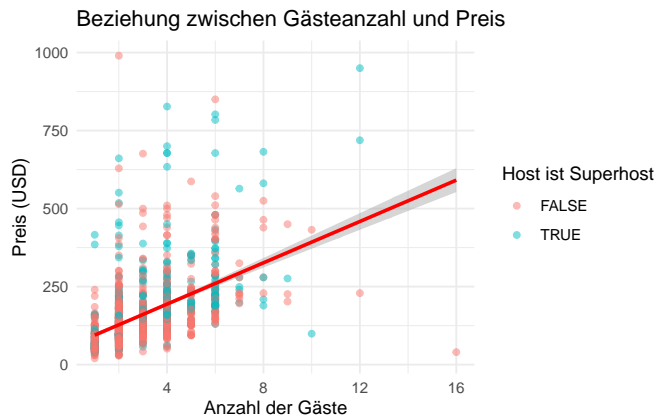
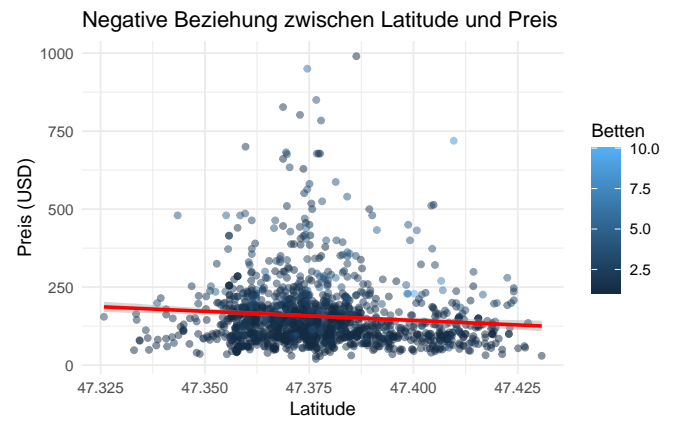
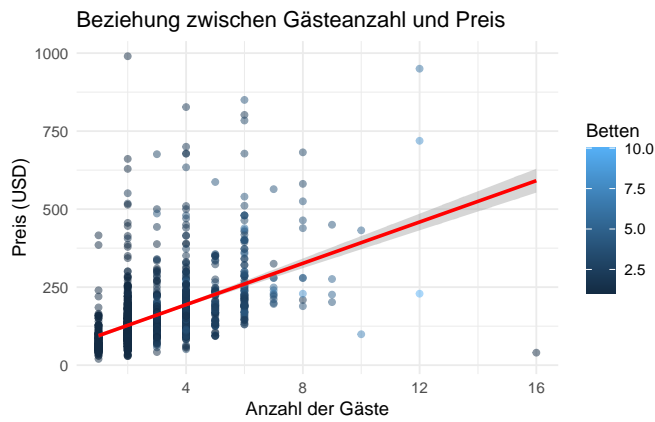
- **Hohe positive Korrelationen** zeigen an, dass mit zunehmender Kapazität und Verfügbarkeit der Unterkünfte der Preis steigt. Dies reflektiert die Marktlogik, dass grössere und häufiger verfügbare Unterkünfte als wertvoller angesehen werden.
- **Negative Korrelationen** deuten darauf hin, dass bestimmte Faktoren wie die geografische Lage (nördlicher) oder die Politik längerer Mindestaufenthalte die Preise senken können. Dies könnte für Gäste attraktiv sein, die längerfristige Aufenthalte suchen oder flexibel in der Wahl der Region sind.

Diese Korrelationen sollten jedoch vorsichtig interpretiert werden, da Korrelation nicht gleich Kausalität ist. Andere verborgene Variablen könnten ebenfalls eine Rolle spielen, und die Effekte könnten durch spezifische Marktbedingungen oder andere nicht berücksichtigte Faktoren beeinflusst werden.

Um diese Korrelationen auch noch grafisch aufzuzeigen mit einer linearen Regression.

Beziehung zwischen Bettenanzahl und Preis





Die Analyse zeigt auf die irgendwie logische Korrelation zwischen den Anzahl Gästen respektive Betten und den Preisen eines AirBNBs... Also je grösser das AirBNB desto mehr kostet es...

Sieht schon ein wenig besser aus - vorallem die Latitude. Wobei auch hier ist es irgendwie klar. Je weiter weg das AirBNB von zentrum weg ist desto günstiger ist die Wohnung...

Gibt es sonst noch irgendwelche Möglichkeiten den Preis vorherauszusagen?

RMSE: 41.90135

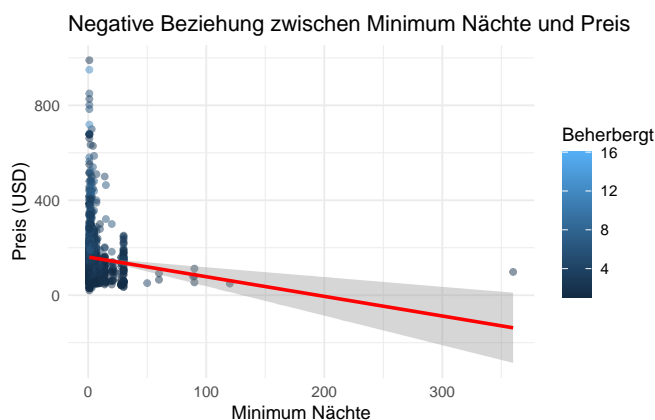
MAE: 27.91819

R^2 : -0.01782196

Gibt es noch andere Korrelationen welche nicht so klar ersichtlich sind?

	IncNodePurity
availability_365	833646.7
review_scores_rating	488147.3
number_of_reviews	718716.6

Vielleicht sind die negativen Beziehungen spannender:



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.

- **R²: -0.01782196** – Ein negativer R²-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.

VI. 6. ERGEBNISSE (STATISTISCHE ERGEBNISSE, ZAHLEN, DIAGRAMME)

VII. 7. SCHLUSSFOLGERUNG (BEANTWORTUNG DER FRAGE)

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/liWCNJcSutYqpULSQHINyGInUvHg2BoUGoNRIGa6Szc4/edit#gid=1322284596>. [Accessed: 30-May-2024]