

Statistical learning Seminar
Berlin airbnb: Analyse und
Angebotpreisvorhersage

Minh Anh Le - Leander Piepenbring

Wirtschaftsmathematik - HTW Berlin
Februar 2022

Inhaltsverzeichnis

I	Einleitung	4
1	Motivation	4
II	Datenanalyse	5
2	Explorative Datenanalyse	5
III	Methodik und Modellierung	9
3	Random Forest	9
4	Entscheidungsbaum	10
5	Modellierung	11
IV	Visualisierung	13
6	Deskriptive Analyse	13
V	Empirische Ergebnisse	18
VI	Fazit	20
VII	Declaration of Authorship	22
	Literaturverzeichnis	23

Abbildungsverzeichnis

1	Beispiel Datensatz aus Listings	6
2	Beispiel Datensatz aus Listings	8
3	Beispiel Entscheidungsbaum	11
4	Train-,Test-Set Übersicht	12
5	Anzahl Wohnungen nach Stadtteil	13
6	Preisverteilung nach Wohnungskategorie	14
7	Preisverteilung nach Stadtteil	16
8	Preisentwicklung über die Woche	17
9	Ergebnistabelle	19
10	Vegleich zwischen Original - und vorhergesagtem Preis	19
11	Auflistung der wichtigsten Mermalen	20

Teil I

Einleitung

1 Motivation

Seit einigen Jahren gehen die Immobilienpreise, und somit auch die Mieten steil nach oben. Besonders stark sind davon große Städte betroffen. Viele Leute möchten deshalb in der Zeit, in der sie nicht vor Ort sind, ihre Wohnung (unter)vermieten.



An dieser Stelle kommt Airbnb in Spiel. Airbnb ist ein Online-Wohnungsportal, das den Benutzern ermöglicht, ihre eigenen Wohnungen anzubieten, andere Angebote zu entdecken und Buchungen abzuschließen. Airbnb finanziert sich dabei über eine Gebühr von 3 Prozent pro Buchung und weitere kleine Service-Gebühren.

Dieses Projekt zielt darauf ab, die Immobilienpreisentwicklung in Berlin anhand eines Airbnb-Datensatzes nachzuvollziehen. Dabei werden unterschiedliche Fragestellungen betrachtet und diese anhand von deskriptiven Analyse-Methoden beantworten.

Ein weiteres Ziel dieses Projektes ist, einen Preis für ein gegebenes Angebot vorherzusagen und die Merkmale zu filtern, die den Preis am stärksten beeinflussen. Dafür wird die Random Forest Methode angewandt.

Teil II

Datenanalyse

2 Explorative Datenanalyse

2.1 Abrufen der Daten

Die Daten, die für dieses Projekt verwendet werden, kommen von InsideAirbnb. InsideAirbnb ist eine unabhängige Internetseite, auf der man detaillierte Informationen über Airbnb Angebote auf der ganzen Welt erhalten kann. InsideAirbnb nimmt seine Daten dabei direkt von der Airbnb Website. Die Daten werden für alle Städte separat zur Verfügung gestellt.

Die Daten umfassen eine Fülle an Informationen über die Angebote, wie den Preis pro Nacht, die Anzahl der Gäste, die Verfügbarkeit der Unterkunft, die Lage und vieles weitere.

Für dieses Projekt wurde der Listings Datensatz und der Calendar Datensatz verwendet. Die Daten umfassen den Zeitraum vom 22.09.2021, bis zum 21.09.2022 und wurden am 05.12.2021 für dieses Projekt heruntergeladen.

2.2 Listings Datensatz

Der Listings Datensatz umfasst insgesamt 18288 Angebote und es gibt 74 verschiedene Merkmale. Aufgrund der hohen Anzahl an Merkmalen, werden für dieses Projekt nur Merkmale berücksichtigt, von denen erwartet wird, dass sie einen signifikanten Einfluss auf die Preisbildung haben.

id	name	description	host_id	host_name	host_since	host_location	host_response_time	host_is_superhost	host_neighbourhood	neighbourhood_group_cleansed
	Berlin-Mitte Value! Quiet location! <br courtyard/ve ry central	Great location!
30 of 75 sq meters. Thi...	2217	kon	18.08.2008	Key Biscayne, Florida, United States	within an hour	t	Mitte	Mitte

neighbourhood	property_type	room_type	accommodates	price	minimum_nights	number_of_reviews	first_review	last_review	review_scores_rating
Berlin, Germany	Entire guesthouse	Entire home/apt	2	\$77.00	60	143	11.04.2016	22.07.2021	4.56

review_scores_cleanliness	reviews_per_month	calculated_host_listings_count	availability_365
4,52	2,15	5	328

Abbildung 1: Beispiel Datensatz aus Listings

- **ID:** Das Kennzeichen des Angebotes.
- **Name:** Der Titel des Angebotes.
- **Description:** Kurze Beschreibung über das Angebot.
- **Host ID:** Das Kennzeichen des Vermieters.
- **Host name:** Der Name des Vermieters.
- **Host since:** Der Tag, am dem der Vermieter das Konto bei airbnb eröffnet hat.
- **Host location:** Der aktuelle Ort von dem Vermieter
- **Host response time:** Wie schnell Anfragen vom Host beantwortet werden
- **Host ist superhost:** Status des Vermieters, ob er Superhost ist.
(t oder f)

- **Host neighbourhood:** Stadtteil, in dem der Host wohnt
- **Neighbourhood group cleansed:** Der Stadtteil des Angebotes
- **Neighbourhood:** Die Stadt des Angebotes
- **Property type:** Angaben, über die Immobilie
- **Room type:** Angaben über Räumlichkeiten
- **Accomodate:** Die Anzahl der Gäste, die zu dem Angebot passt.
- **Price:** Der Preis pro Nacht (angegeben in USD)
- **Minimum nights:** Mindestanzahl an Nächten
- **Number of reviews:** Anzahl der Reviews
- **First review:** Datum des ersten Reviews
- **Last review:** Datum des letzten Reviews
- **Review score rating:** Gesamtbewertung des Angebots (0 - 5)
- **Review score cleanliness:** Bewertung der Sauberkeit (0 - 5)
- **Review per month:** Anzahl der Reviews pro Monat
- **Calculated host listings count:** Anzahl der Angebote des Hosts.
- **Availability 365 :** Anzahl der Tage, die das Angebot verfügbar ist pro Jahr.

2.3 Calendar Datensatz

Der Calendar Datensatz gibt Auskunft darüber, ob ein Angebot für jedes Datum zwischen dem 22.09.2021 und dem 21.09.2022 verfügbar (t) oder nicht Verfügbar (f) ist. Für jedes Angebot gibt es also 365 Datensätze, wodurch der Calendarsatz insgesamt 6674848 Datensätze umfasst. Der Datensatz ist in 7 verschiedene Merkmale unterteilt.

listing_id	date	available	price	adjusted_price	minimum_nights	maximum_nights
162111	22.09.2021	f	42.0	42.0	5.0	21.0

Abbildung 2: Beispiel Datensatz aus Listings

- **Listing ID:** Eindeutiges Kennzeichen des Angebots
- **Date:** Datum TT.MM.JJJJ
- **Availability:** Verfügbarkeit des Angebots zu jedem Tag des Jahres (t oder f)
- **Price:** Preis pro Nacht
- **Adjusted Price:** Angepasster Preis
- **Minimum Nights:** Mindestanzahl an Nächten
- **Maximum Nights:** Maximalanzahl an Nächten

Teil III

Methodik und Modellierung

Um den Preis pro Nacht vorherzusagen und die Merkmale zu filtern, die den Preis am meisten beeinflussen wurden Modelle des Machine Learning verwendet. Diese Algorithmen verwenden die Daten und verschiedene Annahmen, um die Ausgabevariable, den Preis pro Nacht eines Airbnb Angebots vorherzusagen. Der Datensatz wurde in ein 70/30 Train-Set und Test-Set aufgeteilt. Diese Stichproben wurden in Random Forest-Modelle implementiert, um ihre Leistungsmetriken in das Bestimmtheitsmaß R^2 , die mittlere quadratische Abweichung (MSE), den mittleren absoluten Fehler (MAE) und die Wurzel der mittleren Fehlerquadratsumme (RMSE) zu bewerten.

3 Random Forest

3.1 Definition

Was ist Random Forest?

Der Random Forest Algorithmus ist ein überwachter Lernalgorithmus, der für Klassifizierungs- oder Regressionsaufgaben verwendet wird. Er wird sehr gerne verwendet, weil nur wenige Annahmen über die Verteilung der Daten getroffen werden müssen und die Ergebnisse trotzdem sehr zufriedenstellend sind. Diese Methode wird wegen ihrer guten Vorhersagekraft, hohen Interpretierbarkeit und Gewichtung der wichtigsten Merkmale verwendet.

3.2 Funktionalität

Wie funktioniert Random Forest?

Der Random-Forest-Algorithmus stellt Regeln bereit, wie viele verschiedene Entscheidungsbäume generiert werden, die dann mithilfe einer speziellen Ensemble-Methode kombiniert werden, um ein Ergebnis zu erzeugen. Beim Erstellen dieser Entscheidungsbäume wird jedes Mal, wenn eine Teilung in einem Baum betrachtet wird, eine Zufallsstichprobe von m Prädiktoren als Teil-

lungskandidaten aus dem vollständigen Satz von p Prädiktoren ausgewählt. Die Aufteilung darf nur einen dieser m Prädiktoren verwenden. Normalerweise wird $m \approx \sqrt{p}$ verwendet, das bedeutet, dass die Anzahl der bei jeder Teilung berücksichtigten Prädiktoren ungefähr gleich der Quadratwurzel der Gesamtzahl der Prädiktoren ist. Das Ziel des Projektes war es, den Random Forest Algorithmus zu verwenden, um den Preis eines Angebots pro Nacht in Berlin basierend auf den 51 Merkmalen mit der größten Varianz im Trainingsset vorherzusagen. Der gesamte Datensatz wird zufällig in einen Trainings- und einen Testsatz aufgeteilt und Random Forest auf den Trainingssatz für drei verschiedene Werte der Anzahl der Teilungsvariablen m angewendet. Die Ergebnisse sind in Abbildung 8 dargestellt. Die bestmögliche Wahl für m und p sind $m = 40$ und $p = 200$.

4 Entscheidungsbaum

4.1 Definition

Was ist ein Entscheidungsbaum?

Ein Entscheidungsbaum ist eine grafische Darstellung der möglichen Ergebnisse oder Auswirkungen einer Reihe verwandter Entscheidungen. Dieser Ansatz ermöglicht die Klassifizierung oder Regression einer großen Anzahl unterschiedlicher Datensätze und kann zur Entwicklung mathematischer Algorithmen zur Vorhersage der besten Wahl verwendet werden.

4.2 Funktionalität

Wie funktioniert Entscheidungsbaum?

Entscheidungsbäume beginnen typischerweise mit einem einzelnen Knoten, aus dem Zweige hervorgehen, die die entsprechenden Ergebnisse verwandter Entscheidungen darstellen. Jedes Ergebnis ist mit anderen Knoten verbunden, die wiederum andere potenzielle Auswirkungen haben. Die Knoten haben zwei Möglichkeiten:

- Bedingungen [Entscheidungsknoten]

- Ergebnis [Endknoten]

Die Kanten stellen die Wahrheit oder Falschheit einer Aussage dar, und eine Entscheidung wird basierend auf dem folgenden Beispiel getroffen, das einen Entscheidungsbaum zeigt, der die größte von drei Zahlen auswertet.

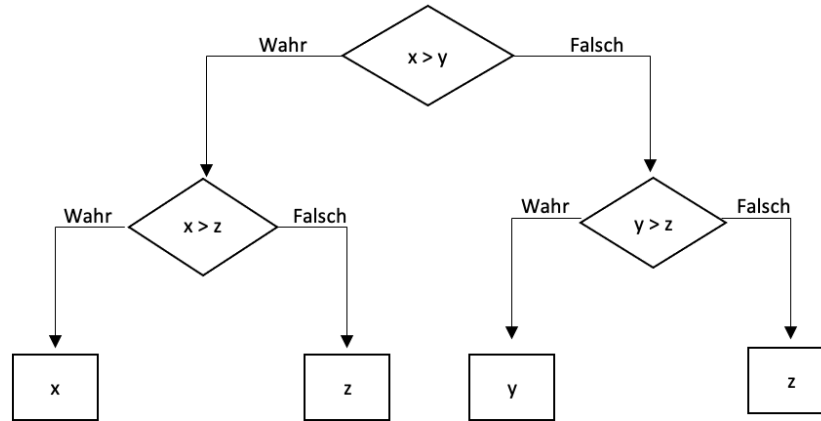


Abbildung 3: Beispiel Entscheidungsbaum

5 Modellierung

Bevor der Random Forest Algorithmus angewendet werden kann, müssen die Merkmale mit Datentyp **String** oder **Object** zu **float**, **integer**, **boolean** oder **array** umgewandelt werden. Danach werden die Daten in zwei verschiedene Sets aufgeteilt: Training Set und Test Set, enthält jeweils einen Vektor Y und eine Matrix X:

- der Vektor Y, der alle realen Preise des Datensatzes enthält
- die Matrix X, die alle Merkmale enthält, die für den Preis relevant sind

Um einen Overfit dieses Modells zu vermeiden, wird ein Tool von sklearn verwendet, um die Beobachtungen in zufälligen Training-Set und Test-Set aufzuteilen:

```
Training set: X_train:(4958, 1440) Y_train:(4958,)
Test set: X_test:(2125, 1440) Y_test:(2125,)
-
Full dataset: X:(7083, 1440) Y:(7083,)
```

Abbildung 4: Train-,Test-Set Übersicht

Danach wird Random Forest auf den Trainingssatz für drei verschiedene Werte der Anzahl der Teilungsvariablen $m = \text{max_depth}$ und den Wert der Gesamtsbäume $p = \text{n_estimators}$ angewendet.

m entspricht der Anzahl der Merkmale, die durchlaufen werden und p der Anzahl der Entscheidungsbäume, die generiert werden sollen.

Als Beispiele werden $n = 20$, $n = 35$ und $n = 40$ genommen. Für die Anzahl der Bäume wird $p = 200$ verwendet.

Ergebnis in Abbildung 9.

Teil IV

Visualisierung

6 Deskriptive Analyse

Um einen Überblick über den Berliner Airbnb-Wohnungsmarkt zu bekommen werden im folgenden einige datenspezifische Fragen anhand deskriptiver Analysemethoden beantwortet.

Wo gibt es die meisten Angebote in Berlin?

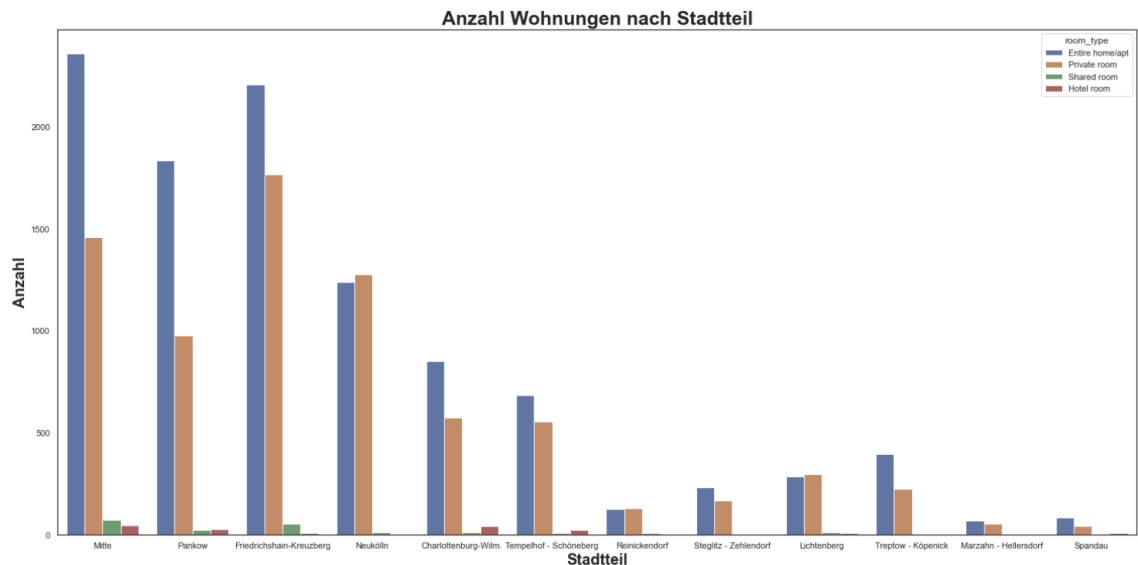


Abbildung 5: Anzahl Wohnungen nach Stadtteil

In dieser Grafik werden die Anzahl der Wohnungen nach Stadtteil und Wohnungsart dargestellt. Es lässt sich leicht erkennen, dass die Stadtteile mit den meisten Angeboten Mitte, Kreuzberg, Pankow und Neukölln sind. Diese Verteilung ist wahrscheinlich darauf zurückzuführen, dass diese Stadtteile sehr zentral liegen und sich dadurch gut für einen Urlaub in Berlin eignen.

Außerdem können wir der Grafik entnehmen, dass es vier verschiedene Wohnungsarten gibt (private Zimmer, geteilte Zimmer, Wohnungen/Häuser, Ho-

telzimmer) und es in Berlin hauptsächlich Angebote für private Zimmer und Wohnungen/Häuser gibt. Das Angebot für geteilte - und Hotelzimmer ist sehr gering. In manchen Stadtteilen, wie zum Beispiel Marzahn-Hellersdorf oder Treptow-Köpenick gibt es gar keine Angebote für diese Wohnungsarten.

Wie verteilt sich der Preis nach Wohnungsart?

Nun soll veranschaulicht werden, wie sich der Preis pro Nacht innerhalb der einzelnen Wohnungsarten verteilt.

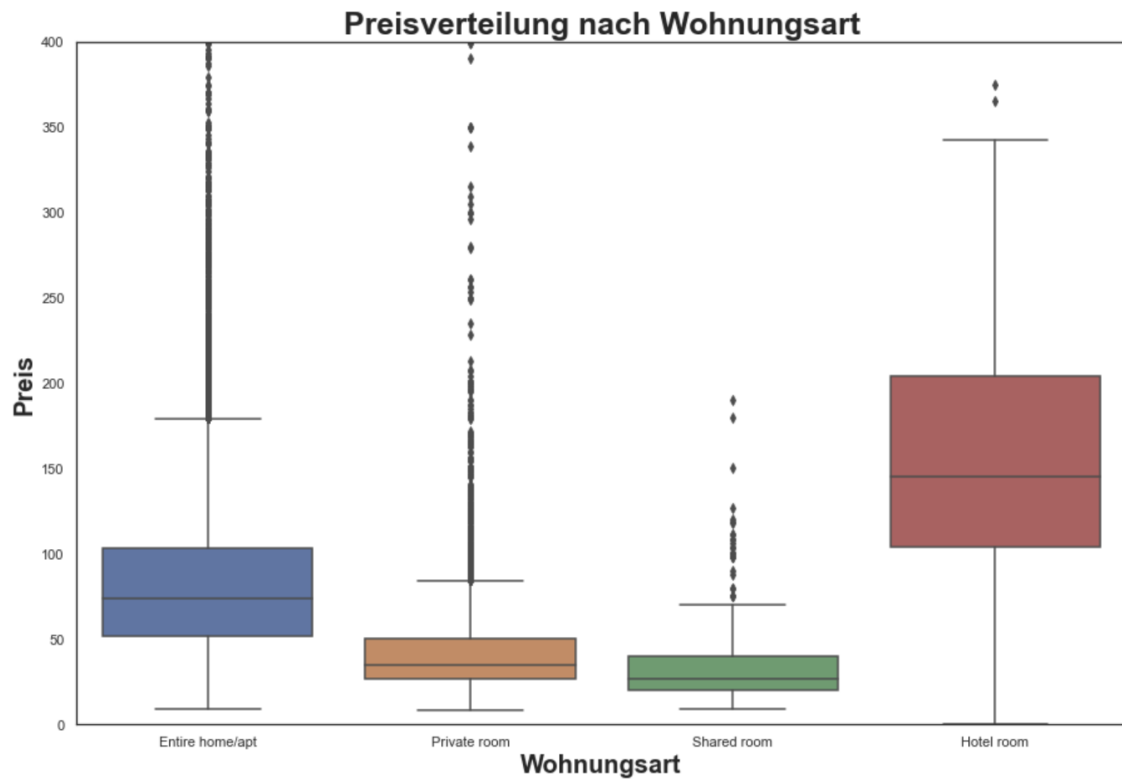


Abbildung 6: Preisverteilung nach Wohnungskategorie

Zunächst soll die Kategorie "Wohnungen/Häuser" betrachtet werden. Der Median für diese Wohnungsart liegt bei 74\$ pro Nacht. Es liegen also die Hälfte der Angebote über und die andere Hälfte unter dem Wert von 74\$ pro Nacht. Ein Viertel der Angebote hat einen geringeren Preis als 52\$ und ein Viertel einen höheren Preis als 103\$ pro Nacht. Es liegt also die Hälfte der Angebote zwischen 52\$ und 103\$ pro Nacht. Es fällt allerdings auf, dass es nach oben sehr viele Ausreißer gibt. Diese reichen bis zu einem Preis von 1000\$ pro Nacht.

Die nächste zu betrachtende Kategorie ist "Private Zimmer". In dieser Wohnungsart liegt der Median bei 35\$ pro Nacht. Die Hälfte der Angebote liegt zwischen 27\$ und 50\$ pro Nacht. Auch in dieser Kategorie fällt, auf, dass es nach oben sehr viele Ausreißer gibt.

Nun zur Kategorie "Geteilte Zimmer". Der Median liegt bei 27\$ und der Interquartilsabstand befindet sich im Bereich von 20 - 40\$. Der Großteil der Ausreißer bei dieser Wohnungsart erstreckt sich bis zu einem Preis von 200\$ pro Nacht. Bei dieser Wohnungsart handelt es sich also um die günstigste Kategorie.

Die letzte Kategorie bildet die Wohnungsart "Hotel Zimmer". Der Median liegt hier bei 147\$ und ist damit deutlich höher als bei den vorherigen Kategorien. Der Interquartilsabstand erstreckt sich von 105\$ bis 205\$. Im Vergleich zu Kategorie 1 und 2 weist die Wohnungsart "Hotel Zimmer" deutlich weniger Ausreißer nach oben auf. Dies kann auf die seriöse Preisgestaltung von Hotels zurückgeführt werden.

Wie verteilt sich der Preis innerhalb der Stadtteile?

Da es für die Kategorien Geteilte - und Hotelzimmer nur sehr wenige Angebote gibt, werden in der folgenden Grafik lediglich die Kategorien Häuser/Wohnungen und Private Zimmer betrachtet.

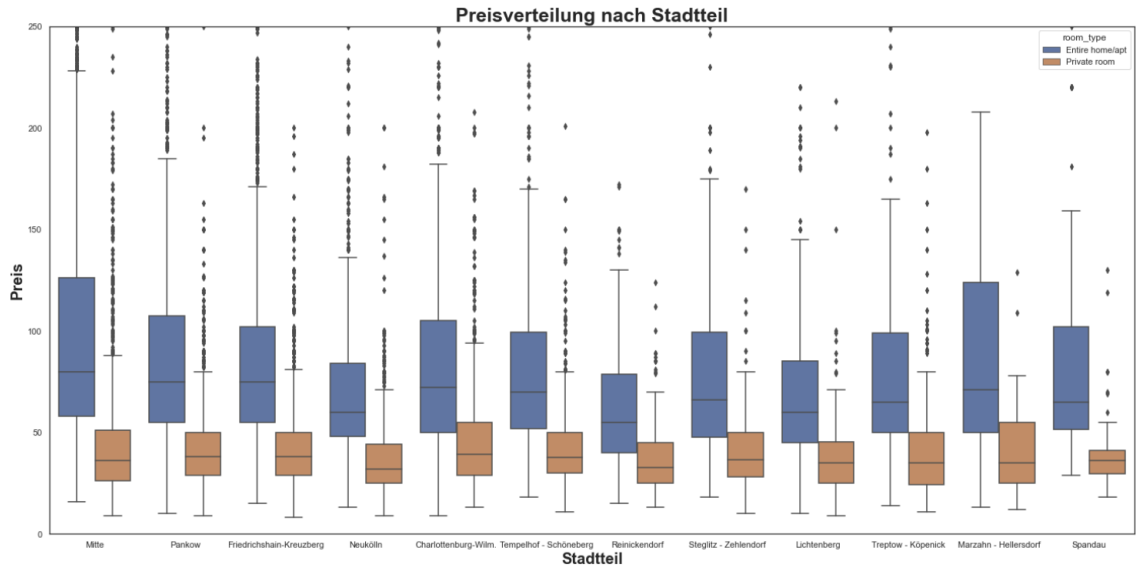


Abbildung 7: Preisverteilung nach Stadtteil

Auf den ersten Blick sieht man, dass die Mediane der verschiedenen Stadtteile sehr nahe beieinander liegen. Auf den zweiten Blick lassen sich aber durchaus Unterschiede in der Länge der Whisker und der Anzahl der Ausreißer erkennen. Als Beispiel wird im Folgenden der Unterschied zwischen Spandau und Mitte betrachtet.

In der Kategorie Häuser/Wohnungen liegt der Median von Mitte leicht über dem von Spandau. Die Länge des oberen Whiskers beträgt bei Mitte 100, während sie bei Spandau nur 57 beträgt und damit fast halb so lang wie bei Mitte. Außerdem gibt es in Mitte eine sehr große Häufung von Ausreißern, während es bei Spandau nur sehr wenige sind. Ein ähnliches Resultat lässt sich auch bei den Privaten Zimmern feststellen.

In Mitte gibt es allerdings auch günstigere Angebote als in Spandau, was man an der Länge des unteren Whiskers erkennen kann. Dies liegt daran, dass es in Mitte eine beträchtliche Anzahl an Angeboten gibt. (Siehe Abbildung 5)

Man kann also festhalten, dass die Bezirke Mitte, Pankow, Friedrichshain-Kreuzberg und Charlottenburg-Wilm. eher der höherpreisigen Kategorie anzuordnen sind und Spandau bzw. Reinickendorf eher im niedrigeren Preissegment zu finden sind.

Wie entwickelt sich der Preis im Verlauf der Woche?

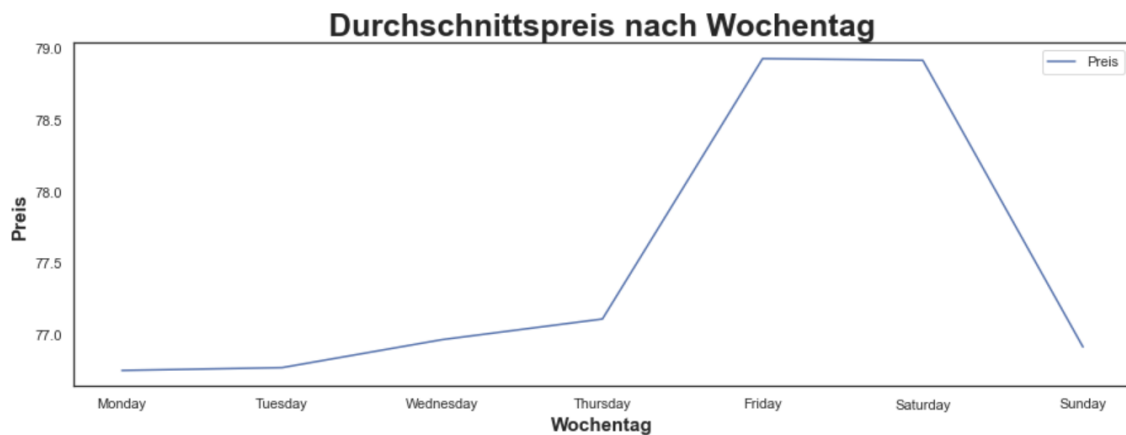


Abbildung 8: Preisentwicklung über die Woche

Wie man unschwer der Grafik entnehmen kann, gibt es einen leichten Anstieg im Durchschnittspreis am Wochenende. An den Tagen Montag und Dienstag liegt der Durchschnittspreis bei 76,75\$ und steigt im Verlauf der Woche bis 78,9\$ an.

Dies liegt offensichtlich daran, dass die Nachfrage an verfügbaren Angeboten am Wochenende ansteigt.

Teil V

Empirische Ergebnisse

Vier Hauptmetriken, anhand derer die Modelle bewertet wurden, waren das Bestimmtheitsmaß R^2 , die mittlere quadratische Abweichung (MSE), der mittlere absolute Fehler (MAE) und der Wurzel der mittleren Fehlerquadratsumme Fehler (RMSE).

Der R^2 Wert ist ein statistisches Maß für den Anteil der Variabilität in der vorhergesagten Variablen, die durch das Regressionsmodell erklärt wird.

$$R^2 = 1 - \frac{\sum (y_i - \hat{y})^2}{\sum (y_i - \bar{y})^2}$$

Der RMSE Fehler misst die Standardabweichung der Residuen um die angepasste Regressionslinie.

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_i - \hat{y})^2}$$

Der mittlere absolute Fehler (MAE) wird verwendet, um die Genauigkeit von Vorhersagen bestimmen.

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|$$

wobei folgende Variablen Verwendung finden:

- n : Anzahl der Vorhersagewerte
- \hat{y}_i : Vorhersagewerte
- y_i : Beobachtungswerte

Diese Werte wurden für jedes Modell für die 70/30 Trainings- und Testdatensplits berechnet.

Ergebnis	m=40, p=200	m=35, p=200	m=20, p=200
R^2	0,7963	0,7667	0,5655
MAE	20,9689	23,7181	31,1334
MSE	1727,3639	1978,6933	2312,77
RMSE	42	44	48

Abbildung 9: Ergebnistabelle

Wie man der Tabelle entnehmen kann, hat man das beste Ergebnis bei $n = 40$ und $p = 200$. Der Wert von R^2 liegt bei 0,7963 und somit höher als bei den anderen Versuchen. Die MAE, MSE und RMSE Fehler weisen alle einen geringeren Wert auf.

Nun werden Originalpreis und vorhergesagter Preis gegenübergestellt.

	actual_y	predicted_y
0	59.0	56.330111
1	18.0	44.572077
2	70.0	57.435863
3	28.0	48.091052
4	30.0	45.558993
5	89.0	73.183768
6	59.0	64.474149
7	89.0	77.964506
8	237.0	172.884995
9	54.0	63.915526

Abbildung 10: Vergleich zwischen Original - und vorhergesagtem Preis

Die obenstehende Abbildung zeigt eine Stichprobe des Original Preises pro Nacht (actual_y) und dem vorhergesagtem Preis (predicted_y). Für dieses Ergebnis wurde $\text{max_depth} = 40$ und $\text{n_estimators} = 200$ gewählt. Wie

man sehen kann, gibt es Datensätze, bei denen die Vorhersage gut funktioniert hat (0, 2, 6, 7, 9), aber auch einige bei denen die Abweichung größer ausfällt.

Jetzt sollen die Merkmale ermittelt werden, die den größten Einfluss auf die Preisbildung haben.

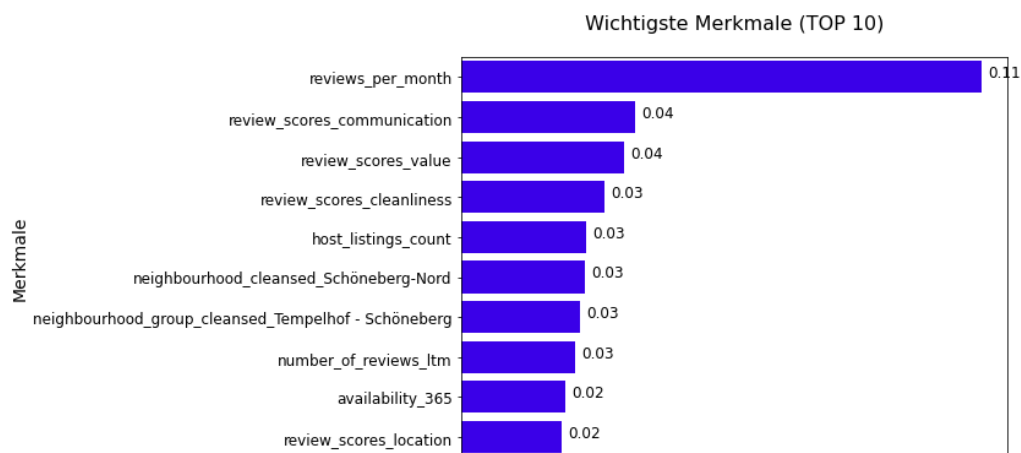


Abbildung 11: Auflistung der wichtigsten Merkmale

Wie man der Grafik entnehmen kann, hat das Merkmal "review per month" den größten Einfluss auf den Preis.

Darauf folgen "review scores communication" (Zufriedenheit mit Kommunikation mit Host), "review scores value" (Gesamtbewertung) und "review scores cleanliness" (Zufriedenheit mit Sauberkeitszustand des Angebots).

Es fällt auf, dass alle vier, der wichtigsten Merkmale mit der Kundenzufriedenheit zusammen hängen und Merkmale wie die Anzahl der Personen pro Wohnung oder der Ort der Wohnung nicht so einen großen Einfluss auf den Preis haben, wie man vorher vermutet hätte.

Teil VI

Fazit

Ziel dieses Projektes war es die Immobilienpreisentwicklung in Berlin nachzuvollziehen, einen Preis für ein gegebenes Angebot vorherzusagen und die Merkmale zu ermitteln, die den größten Einfluss auf den Preis haben.

Die Analysen haben ergeben, dass der Großteil der Airbnb Angebote zentral liegt und es sich vor Allem um Häuser/Wohnungen und private Zimmer handelt.

Der Median ist stadtteilübergreifend sehr konstant, allerdings gibt es bei den zentral gelegenen Bezirken höhere Ausprägungen nach oben, als bei den am Rand gelegenen.

Der Durchschnittspreis steigt an, je näher das Wochenende kommt und fällt zum Wochenstart wieder auf den Tiefpunkt.

Der Random Forest Algorithmus hat sich als eine geeignete Methode zur Vorhersage der Preise erwiesen.

Als Merkmale, die den Preis am stärksten beeinflussen, haben sich "Reviews pro Monat", "Bewertung der Kommunikation", "Bewertung der Sauberkeit" und die "Gesamtbewertung" ergeben.

Alle vier Merkmale hängen, wie vorher nicht zu erwarten war, sehr stark von der Mieterzufriedenheit ab. Wenn man also seine Wohnung vermieten möchte, sollte man besonders darauf achten, positive Reviews von den Mietern zu erhalten. Das erreicht man am besten, indem man schnell und freundlich auf Anfragen der Mieter antwortet und die Wohnung stets in einem guten Zustand ist.

Teil VII

Declaration of Authorship

We hereby confirm that we have authored this Seminar paper independently and without use of others than the indicated sources. All passages (and codes) which are literally or in general matter taken out of publications or other sources are marked as such.

Berlin, 27.02.2022, Leander Piepenbring, Minh Anh Le

Literaturverzeichnis

[1] **An Introduction to Statistical Learning**, <https://www.statlearning.com>,
letzte Stand: 27.02.2022.

[2] **Stackabuse**, <https://stackabuse.com/random-forest-algorithm-with-python-and-scikit-learn/>, letzte Stand: 27.02.2022.

[3] **Github**, <https://github.com/Franky007Bond/Airbnb-Munich-Data-Analysis/blob/master/M>
27.02.2022.

[4] **Towards Data Science**, <https://towardsdatascience.com/predicting-airbnb-prices-in-lisbon-trees-and-random-forests-336d19cdf5a2>, letzte Stand:
27.02.2022

[5] **Machine Learning Mastery**, <https://machinelearningmastery.com/random-forest-ensemble-in-python/>, letzte Stand: 27.02.2022

[6] **Kaggle**, <https://www.kaggle.com/adityadeshpande23/random-forest-regressor-for-airbnb-rooms-price>, letzte Stand: 27.02.2022

[7] **Medium**, <https://medium.com/analytics-vidhya/model-to-predict-booking-prices-on-airbnb-data-science-interview-at-airbnb-c94675aadc1>, letzte Stand:
27.02.2022

[8] **Medium**, <https://medium.com/analytics-vidhya/how-to-analyze-airbnb-performance-data-in-the-right-way-b83f3dad1458>, letzte Stand: 27.02.2022