

01 BACKGROUND

02 APPROACH

03 OUTCOMES

04 NEXT STEPS





01 BACKGROUND

Why are we here today?

02 APPROACH

03 OUTCOMES





AIRBNB IS THE LEADING PLATFORM FOR RENTING OUT PRIVATE PROPERTY

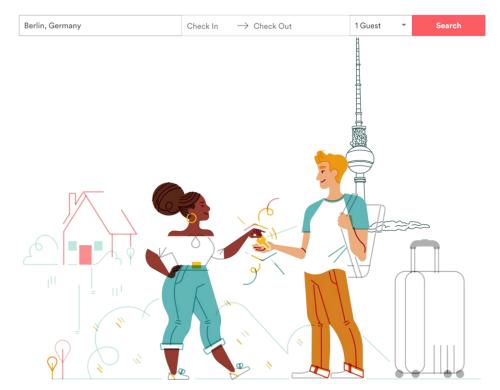
ABOUT AIRBNB

Online platform offering vacation rentals for travelers - primarily homestays - as an alternative to traditional hotel or hostel stays.

Do not own or host themselves but instead act as an intermediary broker

Have ventured into complementary businesses, such as offering **experiences** at major cities

Search over 2 million homes in over 190 countries





CURRENTLY USERS CAN ONLY GET A PRICING ESTIMATE DURING LISTING CREATION

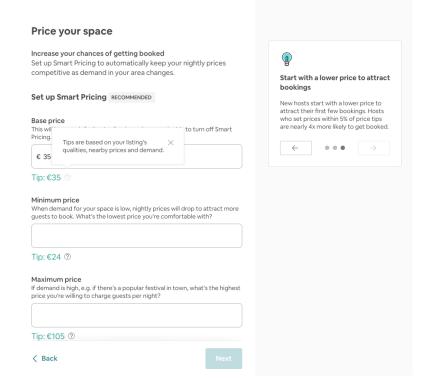
OPPORTUNITY



Airbnb does not publicly publish **booking data** as this presents one of its core assets

Users receive a pricing indication as part of the listing creation process

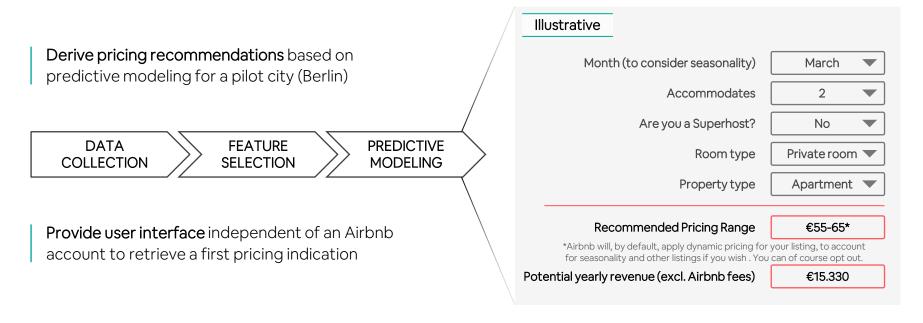
Hence it is difficult for individuals to **estimate a** good listing price before account creation





THIS PROJECT AIMS AT CREATING A TOOL PROVIDING AN UPFRONT PRICING ESTIMATE

CORFIDEA





THIS TARGET HAS NOW BEEN IMPLEMENTED IN THE FORM OF AN INTERACTIVE WEB APP

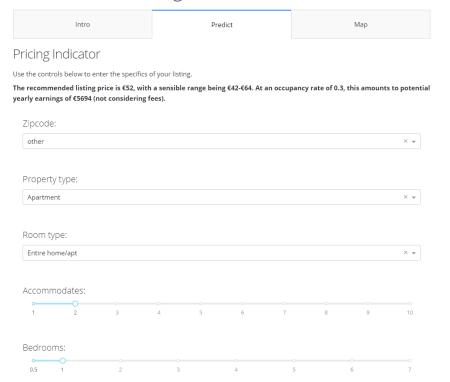
OUTCOME

Pricing recommendation and potential earnings estimate based on up to 24 listing features

Interactive input fields with sensible default values based on average listings

Dynamically changing listing recommendation as soon as input values are adapted

Berlin Airbnb Pricing Recommendation





01 BACKGROUND

02 APPROACH

How was the problem tackled?

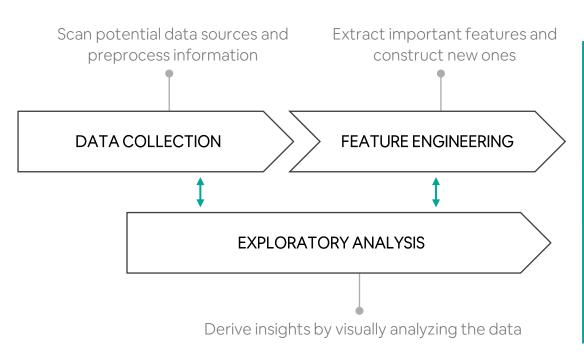
03 OUTCOMES





A STRUCTURED APPROACH WAS FOLLOWED FOR CREATING THE UNDERLYING MODEL

APPROACH



FEATURE SELECTION

Define feature set for predictive modeling

PREDICTIVE MODELING

Train machine learning models, evaluate and make predictions



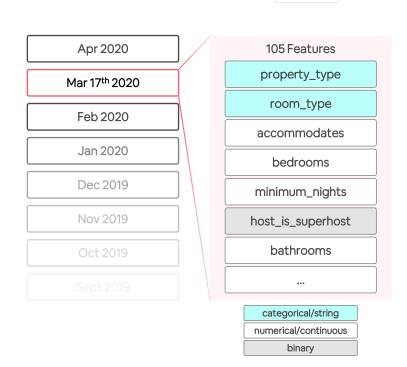
DATA AVAILABLE AT INSIDEAIRBNB.COM LAID THE FOUNDATION FOR THE ANALYSIS

DATA COLLECTION

>25.000 listings with 105 features each were retrieved for Berlin on March 17th 2020

Dataset was cleaned extensively (e.g. inactive listings, errors), reducing eventually used listings to slightly above 10.000

Dynamically changing listing recommendation as soon as input values are adapted





THE DATA HAS SUBSTANTIAL LIMITATIONS, BUT STILL OFFERS A RICH SET OF FEATURES

DATA COLLECTION

Structured provision of publicly available data scraped from all listings at a given date

Key target and predictors available in the form of price and main listing parameters (guests, etc.)

Monthly scraping routine theoretically enables automatized actualization of prediction

No actual booking data provided as this information is not publicly available

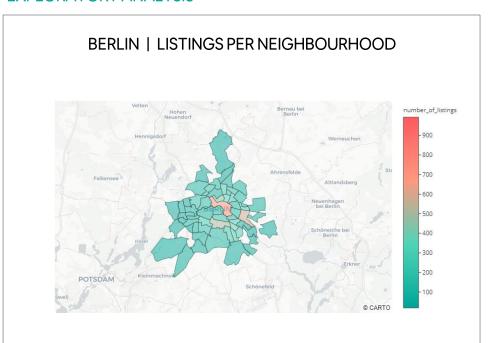
No occupation rates provided as a consequence, making only vague assumptions possible

Price is only listed price and might not in all cases reflect the final price at booking



A BROAD OVERVIEW OF BERLIN HIGHLIGHTS SOME CLEAR HOTSPOTS IN TERMS OF LISTINGS

EXPLORATORY ANALYSIS



Berlin is notorious for having a very **stretched out city area**

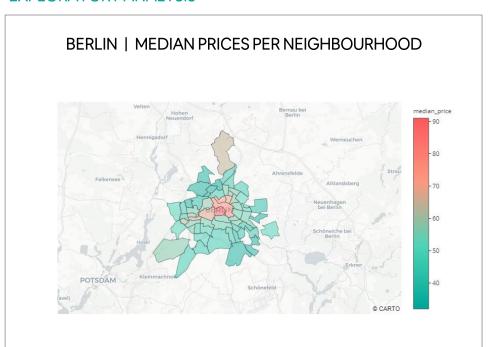
Very few neighbourhoods cover a very large portion of listings

The biggest hotspots are located in the **centre** and east side of Berlin



HIGH PRICES ARE NOT DIRECTLY RELATED TO LISTING HOTSPOTS AND HIGHER IN THE WEST

EXPLORATORY ANALYSIS



Median prices are also **high in the center**, but are more evenly **spread among east and west**

The central area with the **highest median price** is not one of the listing hotspots

Certain **outside areas** with few listings also have relatively **high median prices**

Similarly, when analyzing **zipcodes** there is a clear span in median prices (see Appendix)



ONE KEY FACTOR WITH HIGH CORRELATION TO PRICE IS THE NUMBER OF POTENTIAL GUESTS

EXPLORATORY ANALYSIS



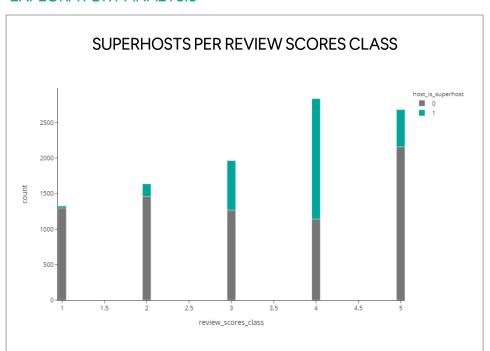
Unsurprisingly, the number of accommodated people per listing appears to be one of the key predictors for price

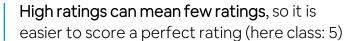
There is a visible and strong **positive relationship** between the number of people hosted and listing price



HIGH RATINGS ARE NOT AUTOMATICALLY RELATED TO MORE SUCCESSFUL LISTINGS

EXPLORATORY ANALYSIS





The **low occurrence of superhosts** in class 5 seems to strengthen this assumption

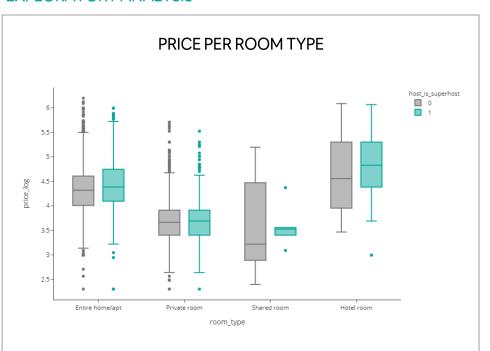
Further analysis reveals perfect ratings are related to newer listings (see Appendix)

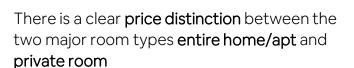
Such insights need to be **considered and potentially treated** during feature engineering



ROOM TYPE, ON THE OTHER HAND, IS CLEARLY A STRONG PREDICTOR OF PRICE

EXPLORATORY ANALYSIS





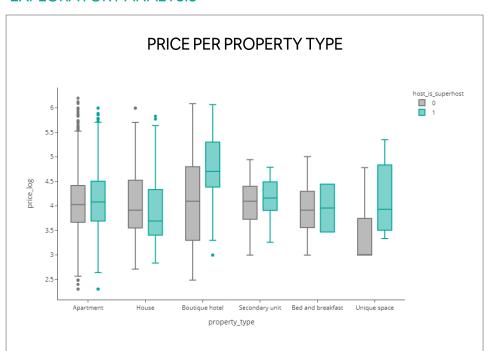
While this follows basic intuition, the difference is quite substantial

Hotel rooms post the highest prices while shared rooms are on the low end



APARTMENTS ON AVERAGE ARE HIGHER PRICED COMPARED TO HOUSES

EXPLORATORY ANALYSIS



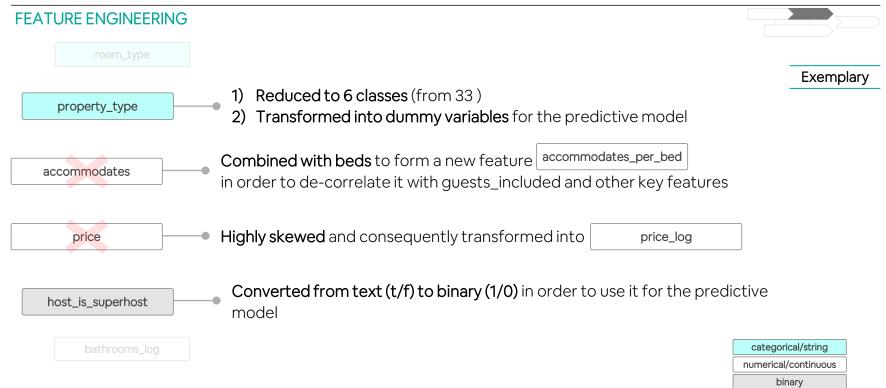


superior location of the former

Superhosts are not necessarily associated with higher prices, which indicates that they are spread across the whole pricing spectrum

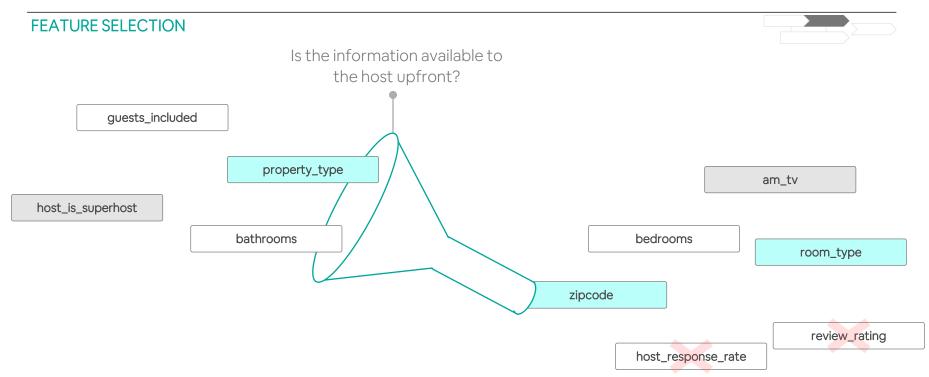


VARIOUS FEATURES HAVE BEEN ADAPTED OR COMBINED BASED ON EXPLORATORY ANALYSIS





INFORMATION AVAILABLE TO USERS PLAYED A KEY ROLE IN SELECTING THE FINAL FEATURES





THE SELECTED FEATURES ENABLE A MODEL CAPABLE OF PROVIDING A FIRST INDICATION

FEATURE IMPORTANCE Room type: Private room Guests included in price **Bedrooms** Room type: Shared room Property type: Boutique hotel Minimum nights



01 BACKGROUND
02 APPROACH

03 OUTCOMES

What has been achieved?





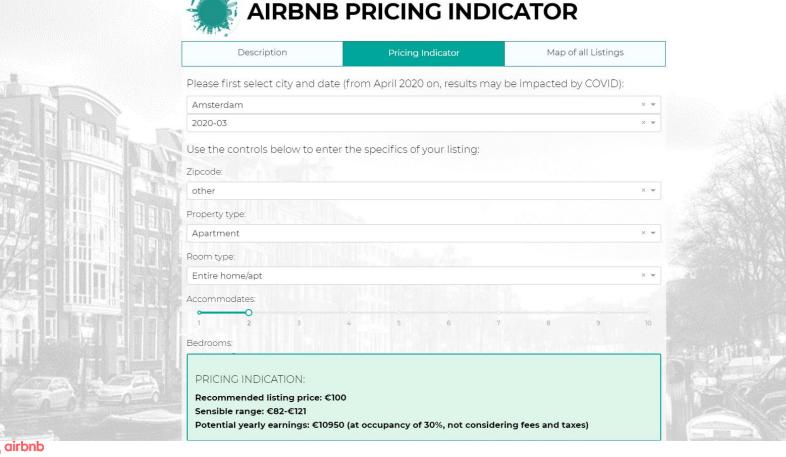
THE MODEL IS APPLIED IN AN INTERACTIVE WEB APPLICATION ACCESSIBLE VIA ANY BROWSER

WEB APPLICATION





THE MODEL IC ADDITED IN AN INTEDACTIVE MED



01 BACKGROUND

02 APPROACH

03 OUTCOMES

04 NEXT STEPS

Where does it go from here?





THE RESULTS OPEN UP COUNTLESS POSSIBILITIES FOR FURTHER EXPLORATION

FUTURE WORK

Predictive modeling

Apply further models and adapt current ones (e.g. NN) Examine other prediction targets (e.g. occupancy rate)

Feature engineering

Explore NLP for text fields (descriptions, reviews, ...)
Scrape listing photos and analyze quality
Enhance current feature set

Lean structure

Remove remaining redundancies wherever possible (e.g. pack repeated steps into functions, apply more pipelines, ...)

Cloud

Move both model creation and app into the cloud (GCP)

Automatization

Build a workflow to automatically retrain model monthly with new datasets

Use automated outlier detection

Use automated feature engineering

Let user input exact address and determine location

Replicability

Apply analysis to even more cities and compare results



Book your next vacation in ...

... a Windmill close to Amsterdam



https://www.airbnb.com/rooms/2107131

... an airstream



... a potato (?)



Thank You!

Mauricio Malzer

linkedin.com/in/mauriciomalzer github.com/Rurbinasal



APPENDIX



01 BACKGROUND

Why are we here today?

02 APPROACH

03 OUTCOMES





01 BACKGROUND

02 APPROACH

How was the problem tackled?

03 OUTCOMES

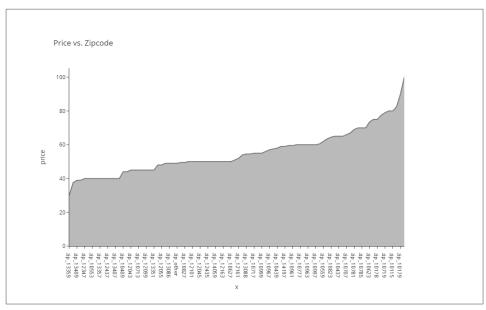




MEDIAN PRICE VARIES SUBSTANTIALLY AMONG DIFFERENT ZIPCODES

EXPLORATORY ANALYSIS

MEDIAN PRICES PER ZIPCODE



- Most zipcodes have a median price between €40 and €60
- A few primarily central and western zipcodes stand out in terms of high median prices

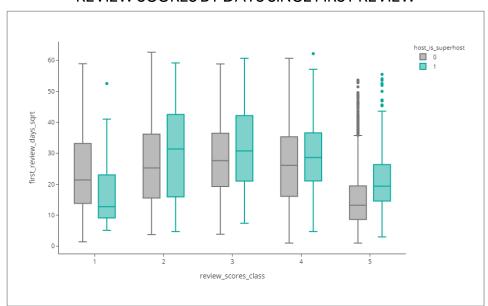


HIGH RATINGS MOST FREQUENTLY OCCUR WITH VERY NEW LISTINGS

EXPLORATORY ANALYSIS



REVIEW SCORES BY DAYS SINCE FIRST REVIEW



- High ratings (class 5) also appear to be the newest listings
- Superhosts have, on average, been on the platform longer than their peers





WALKTHROUGH

Overview Setup Safety Financials Hosting services

The following pages lead through the listing creation process, highlighting the corresponding features in the predictive model

Features



Why host on Airbnb?

No matter what kind of home or room you have to share, Airbnb makes it simple and secure to host travelers. You're in full control of your availability, prices, house rules, and how you interact with guests.

We have your back

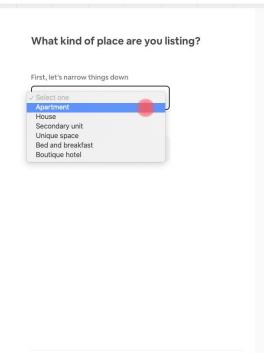
To keep you, your home, and your belongings safe, we cover every booking with \$1M USD in property damage protection and another \$1M USD in insurance against accidents.





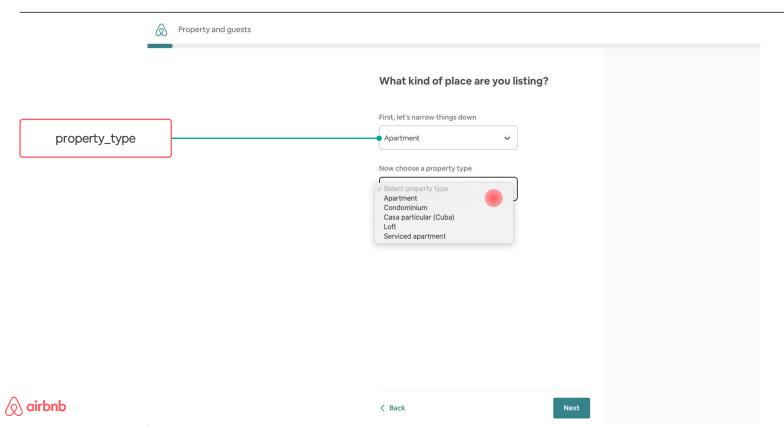


Property and guests

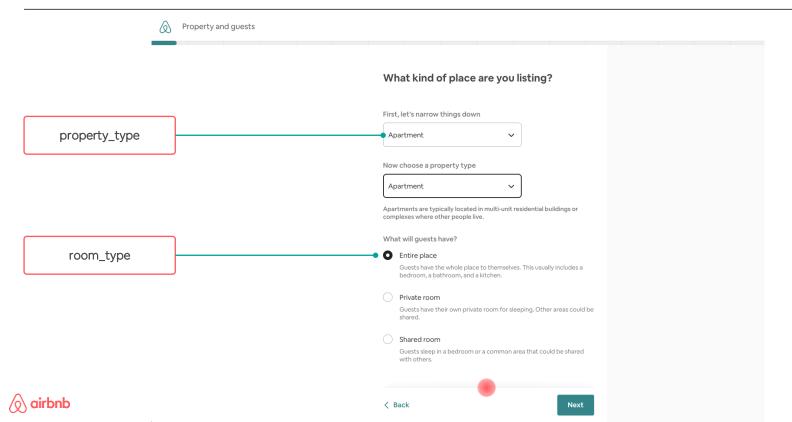




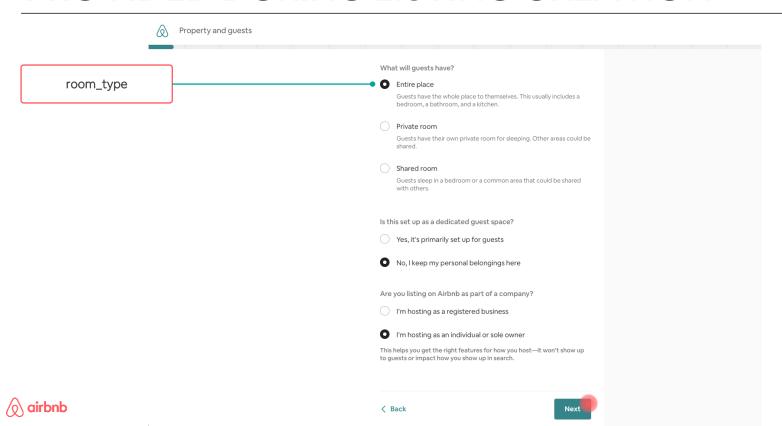




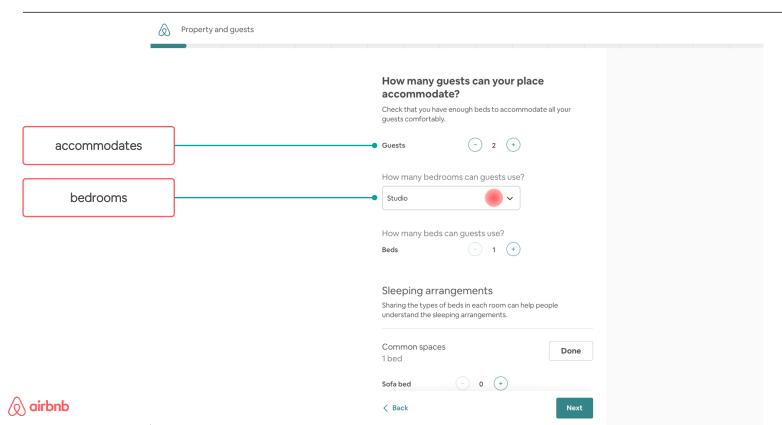




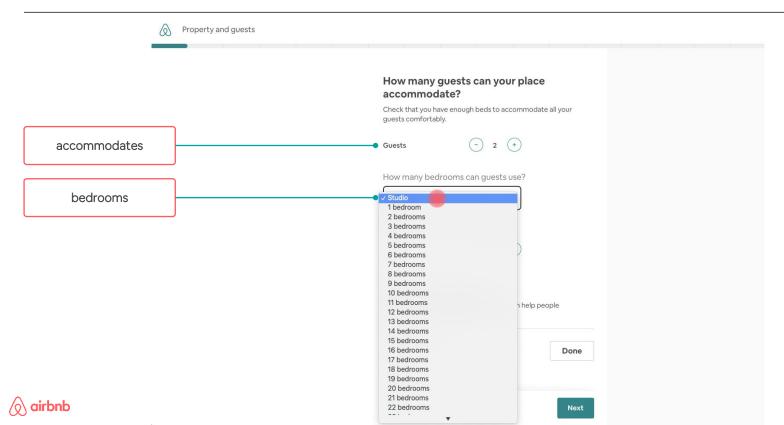












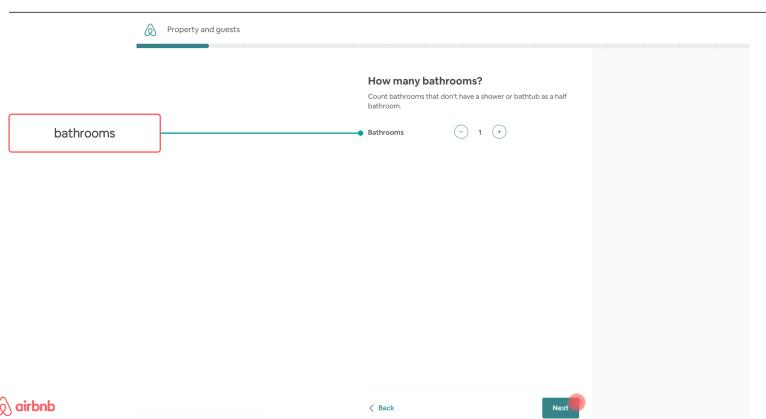


<u> </u>	Property and guests			
		Studio	•	
		How many beds can g	guests use?	
		Sleeping arranger Sharing the types of beds understand the sleeping a	in each room can help peo	pple
beds	•	Common spaces 0 beds		Done
)	Sofa bed	0 (+)	
		Couch	0 (+)	
		Floor mattress	0 (+)	
		Bunk bed	-) o 🔘	
		Add another bed	~	
\ airbnb		< Back		Next



\lambda	Property and guests			
		Studio	v	
		How many beds can	guests use?	
		Sleeping arrange Sharing the types of bed understand the sleeping	ls in each room can help pe	ople
beds	•	Common spaces 1 bed		Done
		Sofa bed	- o (+)	
		Couch	- o (+)	
		Floor mattress	- o +	
		Bunk bed	<u> </u>	
		Add another bed	~	
\ airbnb		< Back		Next

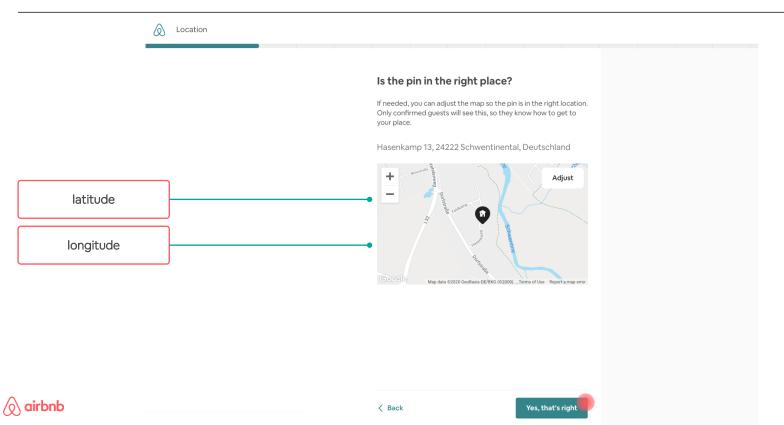




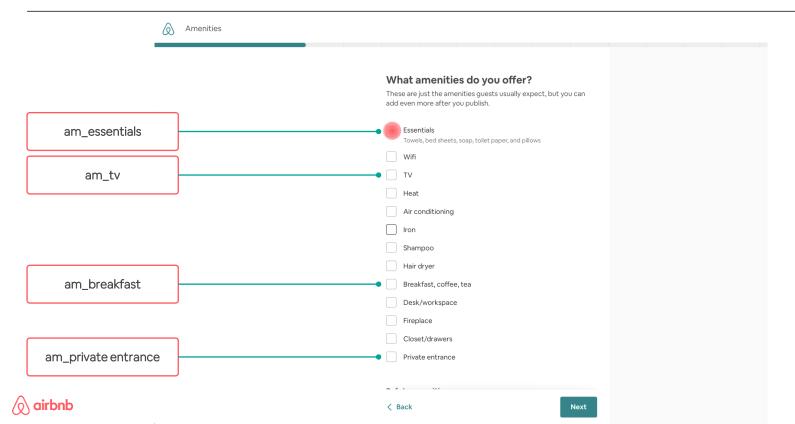


<u> </u>	Location		
		Where's your place is Guests will only get your exact a reservation. Use current location	
		Country / Region Germany	•
		Street address	
		Hasenkamp 13 z. B Kurfürstendamm 67	
		Apt, suite. (optional)	
		z. B Gebäude 1 City	ZIP code
zipcode		Schwentinental z. B Berlin	24222 z. B 10719
_s airbnb		< Back	Next

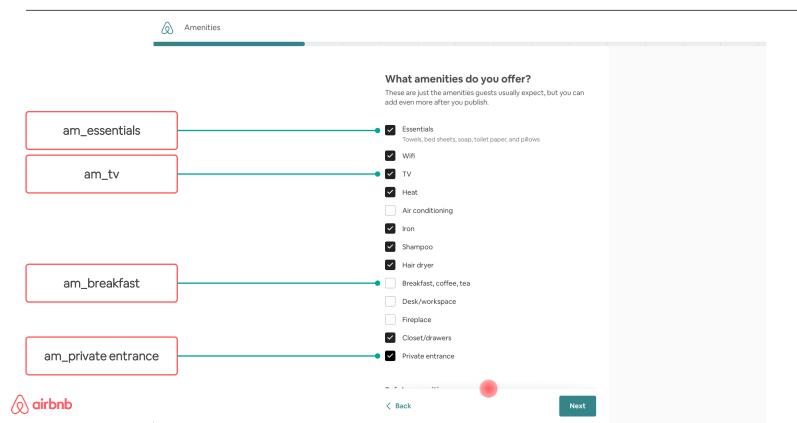




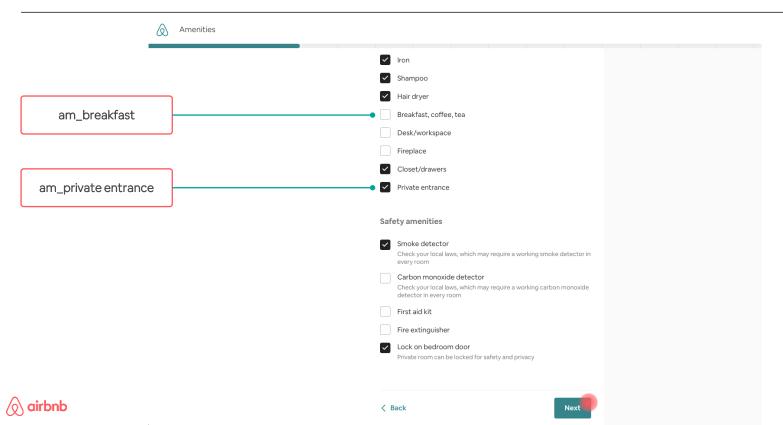




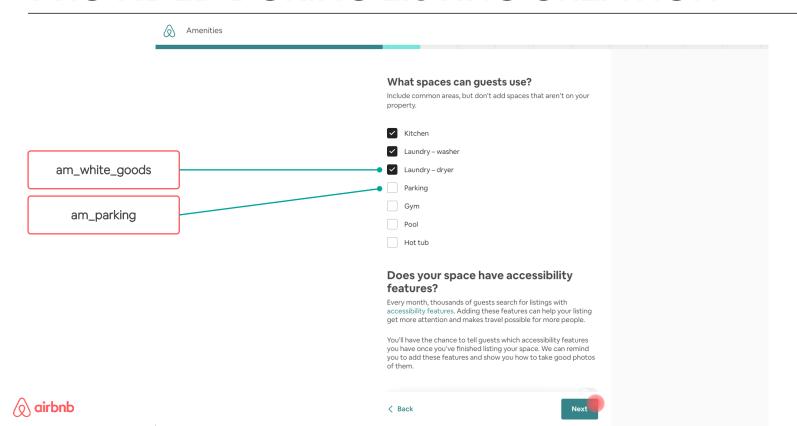
















Liven up your listing with photos

Take photos using a phone or camera. Upload at least one photo to publish your listing—you can always add more or edit later.

Upload Photos

Or drag them in from your desktop



Quick tips for quality photos

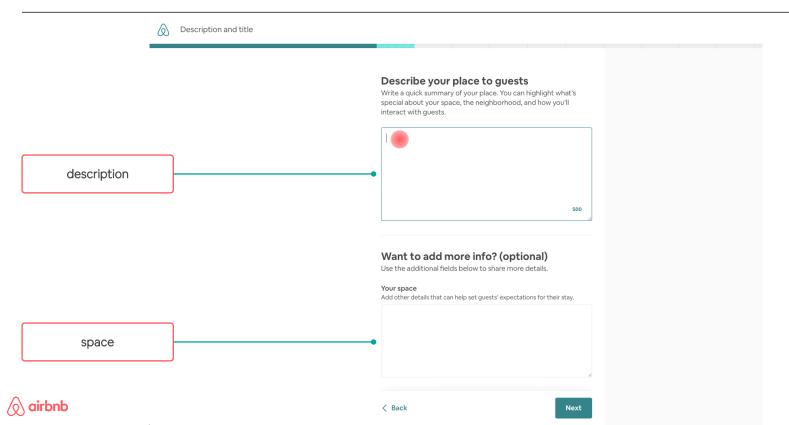
- · Declutter your space
- Use natural daylight and avoid flash
- Shoot from room corners in landscape mode
- Balance visible floor and ceiling space
- Highlight special decor and amenities
- Add photos of every room quests can access



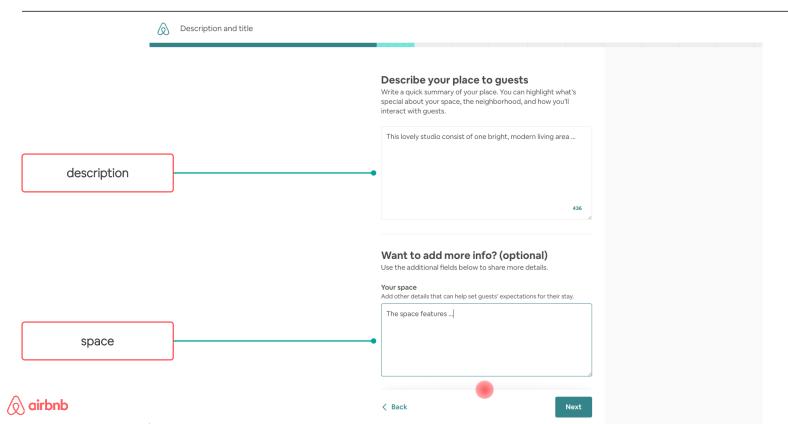




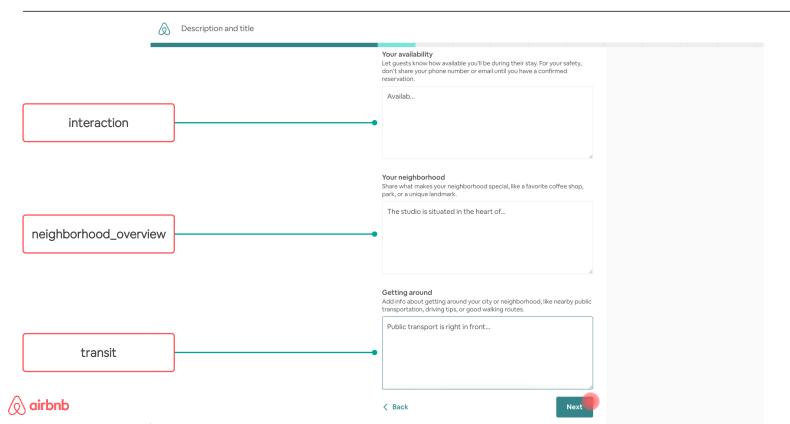




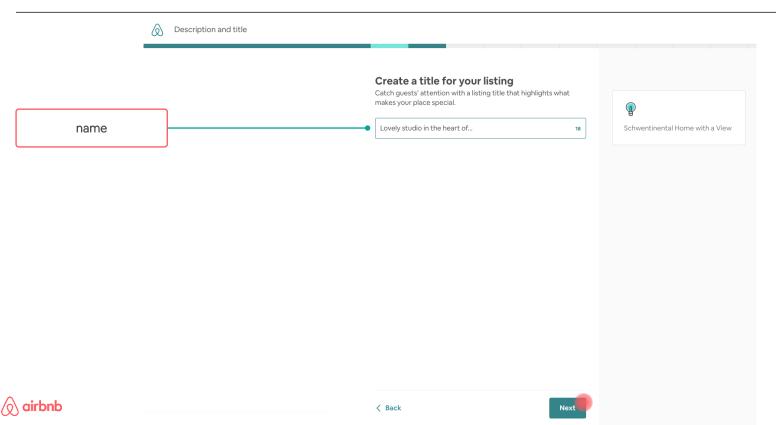
















Booking settings

Review Airbnb's guest requirements

Airbnb has requirements that all guests must meet before they book.

All Airbnb guests must provide:

- Email address
- Confirmed phone number
- Payment information

Before booking your home, each guest must:

- ✓ Agree to your House Rules
- Message you about their trip
- Let you know how many guests are coming
- Confirm their check-in time if they're arriving within 2 days

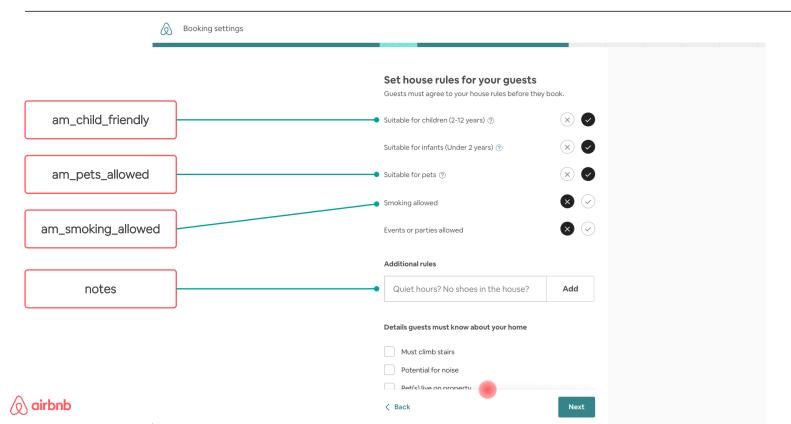
Add additional requirements







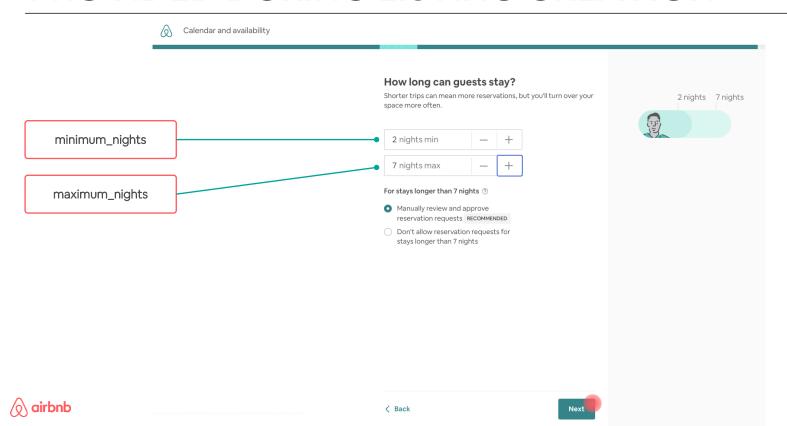






	Booking settings		
		Additional rules	
notes	•	Quiet hours? No shoes in the house?	Add
		Details guests must know about your home	
		Must climb stairs	
		Potential for noise	
		Pet(s) live on property	
		✓ No parking on property	
		Describe the parking situation around your listing	
		Parking on the streets	
		Some spaces are shared	
		Amenity limitations	
		Surveillance or recording devices on property	
		Weapons on property	
		Dangerous animals on property	
.			
) airbnb		< Back	Next

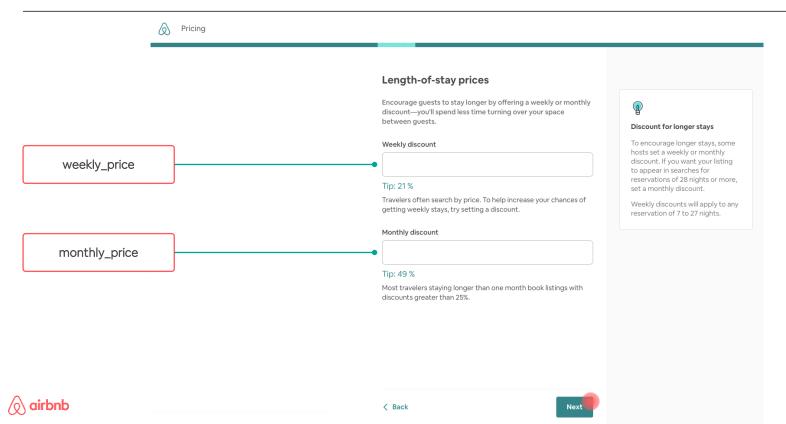






	Pricing	
		Price your space
		Increase your chances of getting booked Set up Smart Pricing to automatically keep your nightly prices competitive as demand in your area changes. Start with a lower price to attract bookings Set up Smart Pricing RECOMMENDED New hosts start with a lower price to attract their first few bookings. Hosts who set prices within 5% of price tips
price		This wil Pricing. Tips are based on your listing's qualities, nearby prices and demand. Tips: €35 ②
		Minimum price When demand for your space is low, nightly prices will drop to attract more guests to book. What's the lowest price you're comfortable with?
		Tip: €24 ⑦
		Maximum price If demand is high, e.g. If there's a popular festival in town, what's the highest price you're willing to charge guests per night?
		Tip: €105 ③
y airbnb		< Back Next ■





THIS SUMMARY SHOWS THE FEATURES RETRIEVED DURING LISTING CREATION



LISTING CREATION

property_type

room_type

accommodates

bedrooms

beds

bathrooms

zipcode

latitude

longitude

am essentials

am_tv

am_breakfast

am_private entrance

am_white_goods

am_parking

description

space

interaction

neighborhood_overview

transit

name

am_child_friendly

am_pets_allowed

am_smoking_allowed

notes

minimum_nights

maximum_nights

price

weekly_price

monthly_price

OTHER SOURCES (E.G. PROFILÉ)

host_is_superhost

host_identity_verified

is_location_exact

security_deposit

cleaning_fee

availability_365

review_scores_rating

instant_bookable

cancellation_policy

calc_host_listings



numerical/continuous

categorical/string

binary

01 BACKGROUND
02 APPROACH

03 OUTCOMES

What has been achieved?





THE MODEL PREDICTS A PRICING INDICATION WITH MEDIAN PERCENTAGE ERROR OF 5,18%



01 BACKGROUND

02 APPROACH

03 OUTCOMES

04 NEXT STEPS

Where does it go from here?



