airbnb

Marketing Strategy Optimization

BA2 - Group10

- Liz Ji - 84126093
- Bella Liu - 94524360
- Thao Do - 33849365
- Hamy Vu - 15268303

Case Overview

airbnb

Our client

- Global booking platform (Est. 2008)
- Earns platform fees from guests & hosts
- Believes guest preferences differ by markets

Data source

- AirBnb-collected listing & rating data (2000 records / market)

Objective

- Identify key occupancy drivers
- Understand topics influencing ratings
- Tailor marketing for each market

Limitations

- No location variables
- Sentiment score derived from simple dictionary script captures tone, not implying subtle meaning

Analytical method

Regression

Build log-linear models using listing attributes to quantify key occupancy drivers for market benchmarks

Topic Modelling

Apply LDA on review text to identify key topics & check how they relate to guests' satisfaction (rating) scores

Regression: Our initial regression featuring all independent variables showed unexpected coefficients in size and host-type, implying underlying issues in variable selection (1/2)

Result for all-variable-included regression of Miami

Variable	Estimate	Std. Error	Stat. significance
(Intercept)	1.022994	0.347493	**
log(price)	-0.275394	0.064885	***
log(number_of_reviews + 1)	0.072376	0.033515	*
rating	0.211648	0.050238	***
1 log(accommodates)	0.184502	0.104604	.
log(bedrooms + 1)	0.597058	0.211542	**
log(bathrooms + 1)	-0.211748	0.202713	
log(beds + 1)	0.233006	0.125461	.
log(minimum_nights + 1)	0.064659	0.040297	
2 host_is_superhost	0.531454	0.067292	***
pro_host	-0.127173	0.071506	.
entire_home	0.686142	0.089129	***
instant_bookable	0.357479	0.069274	***
sentiment	0.028123	0.008727	**

*Significant codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

Quick observation

1. All 3 "size" variables (beds, bedrooms, bathrooms) move in **different directions** with different **significance**

   → Potential multicollinearity issue

2. Superhost is positive & highly significant while pro_host is negative and statistically insignificant

   → Surprising as both suggests "higher quality", "more credible" hosts yet, pro_host is negatively associated with occupancy

Regression: Our initial regression featuring all independent variables showed unexpected coefficients in size and host-type, implying underlying issues in variable selection (2/2)

Result for all-variable-included regression of Paris

	Variable	Estimate	Std. Error	Stat. significance
	(Intercept)	5.508207	0.156947	***
	log(price)	-0.390092	0.032716	***
	log(number_of_reviews + 1)	-0.073523	0.013657	***
	rating	0.080071	0.020787	***
1	log(accommodates)	0.158723	0.039463	***
	log(bedrooms + 1)	-0.038212	0.037866	
	log(bathrooms + 1)	-0.061180	0.089098	
	log(beds + 1)	0.070456	0.056035	
	log(minimum_nights + 1)	-0.103262	0.020784	***
2	host_is_superhost	0.046019	0.042201	
	pro_host	-0.329771	0.046185	***
	entire_home	0.254844	0.043395	***
	instant_bookable	0.069222	0.032056	*
	sentiment	0.006062	0.002753	*

*Significant codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

Quick observation

1 Bedrooms & bathrooms are statistically **insignificant** and **negative**, which contradicts our thinking: bigger houses should not lower occupancy

→ Potential multicollinearity issue

2 Superhost coefficient is **tiny** and statistically **insignificant**, unlike Miami

→ Super host performed differently across markets?

2 Pro host remains **significantly negative**, same as Miami

→ Surprising as it contradicts our intuition of professional hosts having better occupancy performance

Regression: Therefore, we transformed some listing variables to stabilize variance, reduce skewness, and improve linearity for our regression modelling

Variable	**Definition**	**Transform?**
occupancy	# days in the last 90 that the property was booked	✅log(occupancy + 1)
price	Ave. price in ($) for each night	✅log(price)
number_of_reviews	# reviews since the property was listed	✅log(number_of_reviews+1)
rating	Ave. rating from all reviews (out of 5 stars)	❌
accommodates	Max. # people that can stay at the property	✅log(accommodates)
minimum_nights	Min. # nights for each booking	✅log(minimum_nights + 1)
bedrooms	# bedrooms in the property	✅log(bedrooms + 1)
bathrooms	# bathrooms in the property	✅log(bathrooms + 1)
beds	# beds in the property	✅log(beds + 1)
host_is_superhost	1 if the host is a super host; 0 if not	❌
pro_host	1 if the host is a professional property manager; 0 if not	❌
entire_home	1 if the rental includes the entire home; 0 if not	❌
instant_bookable	1 if the property is instantly bookable; 0 if not	❌
sentiment	Score for how positive/ negative a listing's recent guest reviews are	❌

Highlight

- We take the ***log*** of some variables e.g., price, # reviews, beds as they they tend to be right-skewed or because they are count-like

  → Log stabilizes variance & linearizes relationship

- We **add +1** before the *log* of e.g., #, beds, baths as they have 0 values

- We do not transform the rest as they are **binary** indicators (e.g., pro_host), within **bounded range** with little variation (e.g., rating), or have **symmetric** distribution already (e.g., sentiment)

  → Log will not improve anything

Regression: We found high correlations with 4 house-size variables (>0.5), thus proposing to keep accommodates as the only representative for the regression

Correlation result for Miami

	accommodates	bedrooms	bathrooms	beds
accommodates	1.00	0.87	0.75	0.85
bedrooms	0.87	1.00	0.82	0.83
bathrooms	0.75	1.00	1.00	0.71
beds	0.85	0.83	0.71	1.00

Correlation result for Paris

	accommodates	bedrooms	bathrooms	beds
accommodates	1.00	0.01	0.00	-0.02
bedrooms	0.01	1.00	0.52	0.67
bathrooms	0.00	0.52	1.00	0.63
beds	-0.02	0.67	0.63	1.00

Highlight

- Miami shows stronger correlations among 4 house-size variables while Paris shows slightly weaker ones i.e., accommodates don't move together 3 other variables at all
- Intuitively, ***accommodates*** is the closest proxy to guest capacity

  → We keep it as the only size variable in the regression

Regression: Additionally, we choose pro_host (and drop super_host) to represent host quality as it shows a clearer & more consistent relationship with occupancy in 2 markets

1. By definition, pro_host has more economic influence on occupancy

- **Pro_host:** business-like, multi-listing host who provides efficient, consistent, hotel-style service across properties*

  → More supply side levers to influence occupancy e.g., pricing, min. night, availability

- **Super_host:** who perform exceptionally historically with low cancellation, fast response, 4.8+ rating, high bookings**

  → Just signal service quality

2. pro_host has consistent coefficients on occupancy in both markets

- **Super_host** is significant and positive in Miami but insignificant in Paris
- Meanwhile, **pro_host** is always negative across both markets

3. Introduce interaction term: pro_host × minimum_nights

- **Pro hosts** often has stricter minimum-night rules

  → May lower occupancy for short-stay demand

- Adding this interaction allows us to see how # required min night can affect pro more negatively than casual hosts

*) https://www.airhostagency.com/post/individual-vs-professional-hosts-on-airbnb-understanding-your-options
**) https://www.airbnb.ca/resources/hosting-homes/a/how-to-become-a-superhost-702

Final regression function

$$\ln(\text{Occupancy}_i + 1) = \beta_0 + \beta_1 \ln(\text{Price}_i) + \beta_2 \ln(\text{Reviews}_i + 1) + \beta_3 \text{Rating}_i + \beta_4 \ln(\text{Accommodates}_i) + \beta_5 \ln(\text{MinNights}_i + 1)$$
$$+ \beta_6 \text{ProHost}_i + \beta_7 \text{EntireHome}_i + \beta_8 \text{InstantBookable}_i + \beta_9 \text{Sentiment}_i + \beta_{10} \left[\text{ProHost}_i \times \ln(\text{MinNights}_i + 1)\right] + \varepsilon_i$$

Miami regression result: Occupancy is driven by expected factors, with strong signals from guest reviews and listing features revealing Miami guests preferences

Result for final regression of Miami

R^2: 0.1526

	Variable	Estimate	Std. error	p-value	Stat. significance
	(Intercept)	0.342699	0.358844	0.33969	
	log(price)	-0.149419	0.059985	0.01282	*
	log(min_nights+1)	0.066630	0.065414	0.30853	
	pro_host	-0.192991	0.142044	0.17441	
1	log(number_of_reviews+1)	0.161973	0.032760	8.29e-07	***
	rating	0.280799	0.050980	4.10e-08	***
2	log(accommodates)	0.447432	0.078446	1.35e-08	***
	entire_home	0.652388	0.088799	2.95e-13	***
	log(min_nights+1):pro_host	0.057870	0.077876	0.45750	
	instant_bookable	0.329572	0.070612	3.25e-06	***
1	sentiment	0.025772	0.008899	0.00382	**

*Significant codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

Highlight

1 **log(number_of_reviews+1), rating, sentiment**

→ Even with small coefficients, number of reviews, ratings and sentiment are highly reliable predictors of high occupancy, showing that better guest feedback consistently drives stronger booking performance in Miami

2 **log(accommodates), entire_home**

→ Relatively high coefficients show Miami travelers value space and privacy, which is consistent with a market dominated by families, groups, and leisure travelers

- **Surprising Findings**

log(price) has only a modest negative effect (-0.149) on occupancy, suggesting that Miami travelers are relatively less price sensitive. This aligns with a leisure-oriented market where travelers prioritize space, comfort, and overall experience over prices

Pro_host shows an insignificant effect on occupancy, indicating that host professionalism does not influence booked days in Miami
→ Counter-intuitive. **Hypothesis**: Travelers may care more about the property's features than if it is operated by a professional host

Paris regression result: Most key drivers behave as expected while some counter-intuitive nuances highlight dynamics around guest preference, host type & listing appeal

Result for final regression of Paris

R^2: 0.1799

	Variable	Estimate	Std. error	p-value	Stat. significance
	(Intercept)	5.537914	0.144338	< 2e-16	***
1	log(price)	-0.384047	0.032178	< 2e-16	***
1	log(min_nights+1)	-0.149636	0.023929	4.90e-10	***
1	pro_host	-0.586206	0.083305	2.70e-12	***
	log(number_of_reviews+1)	-0.065984	0.013050	4.67e-07	***
	rating	0.080908	0.020674	9.40e-05	***
2	log(accommodates)	0.155069	0.039105	7.58e-05	***
2	entire_home	0.245193	0.043152	1.53e-08	***
2	log(min_nights+1):pro_host	0.167153	0.045256	0.000227	***
	instant_bookable	0.074185	0.031865	0.020006	*
3	sentiment	0.005930	0.002742	0.030678	*

*Significant codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

Highlight

1 **Higher prices, stricter min-night requirements & pro-host listings** all reduce occupied nights

→ Price sensitivity, demand for flexible stays, and potential drawbacks of pro-host

2 **Entire-home** & **interaction** btw. pro-host and min-night requirements increase occupancy

→ Privacy = highly valued. Pro-hosts can partially offset long-stay constraints through more efficient ops

3 **Sentiment:** small, positive, but significant → Even a slight boost in guest tone can improve # bookings

- **Surprising Findings**

Accommodates has a modest upside on occupancy. Though large listings often attract group travellers, it could be that Paris has more non/small group demand

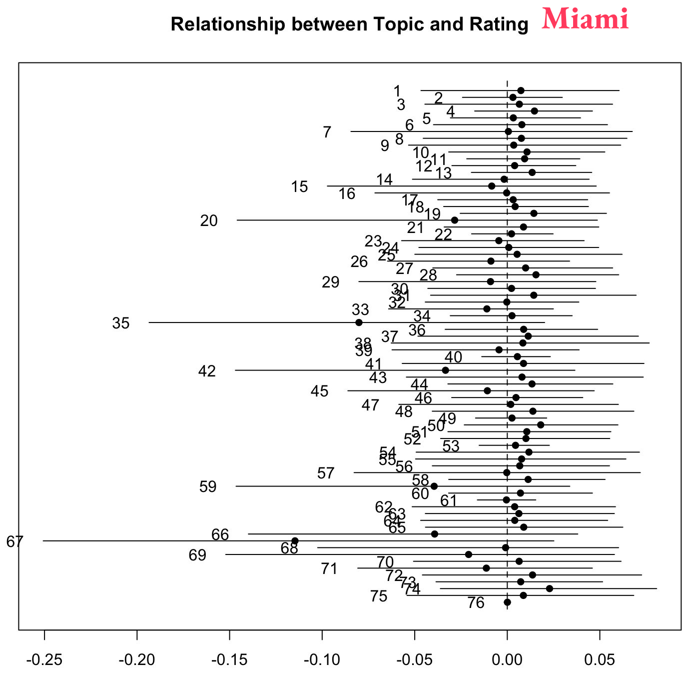
Pro_host's negative coefficient: Potential reasons - Pro, hotel-like operators tend to price higher, ask for longer min stays, and less personalize listings → Lower conversion

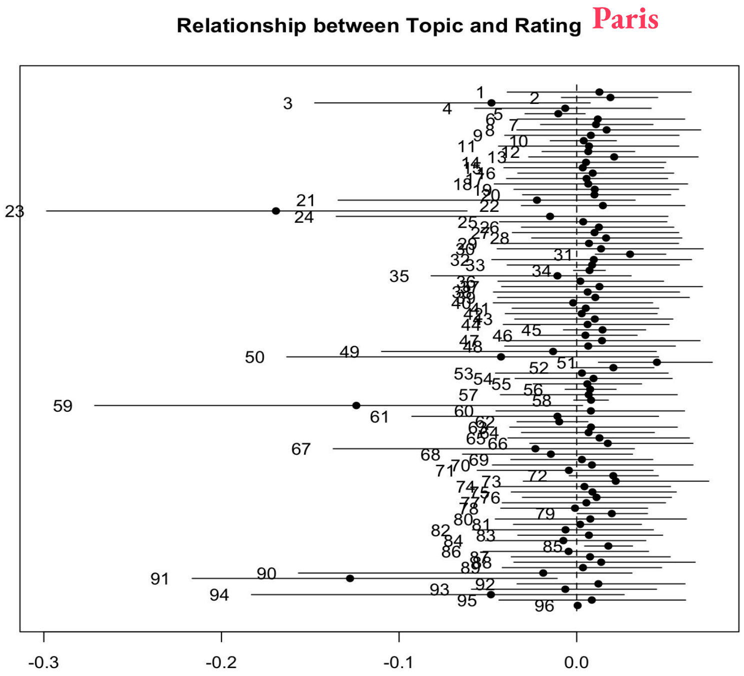
More #reviews should increase host's credibility & bookings yet its coefficient is negative → This reflects historical popularity only. Older listings may have more reviews over time yet still lose bookings to newer, trendier ones

Key Differences between Two Markets: Paris exhibits robust and consistently significant patterns across predictors, while Miami shows weaker significance and noisier patterns

Dimension	Miami	Paris	Insight	Potential reasons
Model Predictability	Lower R^2 (0.1526)	Higher R^2 (0.1799)	Higher predictability in Paris vs more variability in Miami	Paris tourism is more homogeneous, while Miami serves a wider mix of guest types (families, groups, etc.)
Guest Feedback (reviews, rating, sentiment)	All three variables remains highly significant with relative larger effects	All three variables are significant with relative smaller effects	Feedback matters in both markets but matters even more in Miami	Miami guests face more uncertainty across neighborhoods and room style → Rely more heavily on reviews, ratings and sentiment when choosing listings
Impact of Price & Minimum Nights	Price affects occupancy, but min_nights becomes insignificant	Both price and min_nights strongly reduce occupancy	Highly suppresses occupancy in Paris; modest effect in Miami	Paris is dominated by short-stay tourists; Miami has more flexible trip patterns
Listing Characteristics (accommodates, entire_home)	**Strong effects:** big listings & entire homes greatly increase occupancy	Both variables are significant with relative smaller effects	Space and privacy are more valued in Miami	Miami attracts families, groups, and beach travelers; Paris attracts smaller groups and solo tourists
Host type (pro_host)	Insignificantly negative effect on Occupancy	Significantly negative effect on Occupancy	Pro host deters occupancy in Paris but has little impact in Miami	In Paris, pro hosts often pair higher price & stricter rules. In Miami, people tend to stay longer and are less sensitive to host rules, weakening the pro-host effect

Topic modelling result overview





- Total # of topics identified**:** 76
- Topics linked to higher ratings: 28, 37, 50
- Topics significantly associated with lower ratings: 35, 59, 67

- Total # of topics identified**:** 96
- Topics linked to higher ratings: 23, 59, 91
- Topics significantly associated with lower ratings: 30, 51, 73

Miami topic result: Top 3 most **positive** topics show that higher ratings are driven by hospitality, cleanliness, and convenient location

Topic 28: Convenient location

Topic 37: Family-friendly atmosphere

Topic 50: Cleanliness, quietness & privacy

Highlight

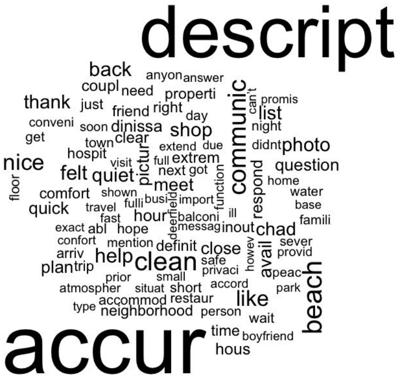
1. Convenient location: Key words like “locat,” “walk,” “within,” “distance,” “near,” “downtown,” “beach,” “restaur,” “min”
→ Emphasize convenience and accessibility of local attractions: Guests could easily reach shops, beaches, and restaurants by foot

2. Family-friendly atmosphere: Frequent words like “welcom,” “famili,” “stay,” “feel,” “comfort”
→ Guests perceived the space as warm, safe, and comfortable for group or family stays

3. Cleanliness, quietness & privacy: Words including “quiet,” “privat,” “space,” “clean,” “safe,” “room”
→ Guests found the space well-maintained and valued the sense of privacy during their stay

Miami topic result: Top 3 most **negative** topics reveal guest dissatisfaction with limited space, inaccurate descriptions, and lack of cleanliness

Topic 35: Different from Description



Topic 59: Cleanliness and Furnishing

Topic 67: Space and Amenities Issues

Highlight

1. **Different from description**: Dominant words like “descript,” “photo,” “accur”
→ Guests felt the property did not match the online description or photos

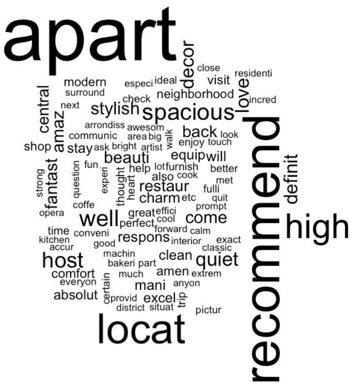
2. **Cleanliness and furnishing**: Repeated terms such as “clean,” “comfort,” “modern,” “furnish,” “docor,” “equip,” “keyless”
→ Property did not meet guests’ cleanliness expectations; guests are dissatisfied with room modernity & visual appearance

3. **Space and amenities issues**: Frequent terms like “bedroom,” “bathroom,” “master,” “size,” “space”
→ Dissatisfaction with room layout, size, or usability; experienced inconveniences with core amenities

Paris topic result: Top 3 most positive topics show that guests value exceptional hosts, convenient locations, and beautiful, well-equipped homes

Topic 30: Great host qualities

Topic 51: Stylish, well-equipped places



Topic 73: Great, convenient locations

Highlight

1. **Great host experiences:** Words e.g., "host", "respond", "communic", "prompt", "welcome", "comfort"
   → Guests favor positive interactions, good communication, and supportive hosts
2. **Stylish, well-equipped apartments:** Words e.g., "spacious", "stylish", "decor", "charm", "clean", "well", "equip", "beautiful", "cozy"
   → Guests value aesthetic appeal, Parisian charm, well-maintained interior design
3. **Convenient location & neighborhoods:** Words e.g., "locat", "central", "metro", "neighborhood", "area", "nearby", "restaur", "bar"
   → Guests consistently praise proximity to transport, restaurants, cool districts

Paris topic result: Top 3 most **negative** topics show that show frustrations center around cleanliness lapses, malfunction, and check-in/ booking management difficulties

Topic 23: Apartment condition & cleanliness issues

Topic 59: Maintenance & functionality issues

Topic 91: Check-in, guest support & communication issues

Highlight

1. **Apartment condition & cleanliness issues**: Repeated words are “apart,” “clean,” “stay,” “towel,” “toilet”
   → Common housekeeping problems e.g., worn or insufficient towels, toilet upkeeping
2. **Maintenance & functionality issues**: “Shower,” “issue,” “door,” “small” “broken” are often mentioned
   → Highlight water problems, broken items, and functional inconveniences during the stay
3. **Check-in, guest support & communication issues**: Key terms are “check,” “time,” “help,” “didnt,” “answer”, “key", “inform", “communic"
   → Hint at check-ins, key handover issues, and host responsiveness

Summary & Recommendation | Miami

Key Insights

- **Higher occupancy is predicted by number of reviews, rating, sentiment, accommodates, and entire home**

  Travelers prefer well-reviewed, spacious, and private listings, consistent with leisure-oriented travel

- **Price has only a mild negative effect on occupancy**

  Miami guests are less price-sensitive and prioritize comfort, experience, and amenities over cost

- **Higher coefficients for accommodates and entire_home highlight the importance of space**

  Bigger listings and full-privacy units are especially appealing for Miami's family/group leisure market

- **Guest satisfaction is strongly driven by hospitality, cleanliness, and location**

- **Negative sentiment is centered around space limitations, inaccurate descriptions, and cleanliness/furnishing issues**

Marketing Strategy Recommendations

- **Introduce stricter listing accuracy standards with a "Verified accuracy" icon**
  - Reduce expectation mismatches by requiring clearer photos and more precise, up-to-date descriptions
  - If feasible, conduct periodic specialist inspections to verify room conditions

- **Emphasize space, privacy, and family/group travel needs for the Miami market**
  - Highlight entire-home and spacious listings in Miami-targeted campaigns, aligning with the preferences of family- and group-based leisure travelers

- **Promote high-rating, high-sentiment listings more prominently**
  - Prioritize listings with stronger ratings and positive review sentiment in search rankings
  - Allocate increased ad exposure to amplify visibility

Summary & Recommendation | Paris

Key Insights

- **Paris guests are highly sensitive to price and stay policy restrictions**

- **Privacy and interior quality strongly influence bookings**
  Entire home listings perform better. Guest left good reviews on stylish, clean, well-equipped listings

- **Host warmth & communications shape guest satisfaction**
  Ratings consistently link responsive, welcoming hosts with positive sentiment, which the regression confirms has a small but meaningful impact on occupancy

- **Operational frictions**: cleanliness, maintenance, and check-in are clear detractors to booking potential

- **Pro-host listings underperform likely because of higher prices & less personalization,** which is misaligned with Paris 's preference for warmth, charm, and flexibility

Marketing Strategy Recommendations

- **Promote flexible stay & value-oriented listings more**
  - Personalize badges & search boosts to highlight competitively priced units and flexible min-nights

- **Emphasize entire-home & design in listing brandings**
  - Create a Best-of-Paris collection of listings with strong decor, charm & aesthetic appeal
  - Utilize rich visuals & curated storytelling to showcase the interior quality guests often praise

- **Marketing host warmth & responsiveness in listing**
  - Add a "Hosted With Care" badge or guest quotes e.g., "Our host was incredibly responsive" to highly visible positions of listings to pin warmth as a differentiator vs traditional pro-host hotel-like listings

- **Refresh older listings to stay competitive**
  - Incentivize/ educate hosts with many older, highly-reviewed listings to to update photos, décor, amenities to compete with newer, more stylish ones

Additional research suggestions to validate hypothesis & explore demand drivers

Miami

1. Assess listing accuracy & expectation alignment

- Determine which factors of descriptions travellers often find inaccurate and quantify their impacts on bookings

→ **Method**: Listing audit, guest surveys

- Test if "Verified accuracy" badge increases bookings & satisfaction → **Methods**: A/B testing

2. Deepdive space, privacy & group travel preferences

- Identify which space attributes e.g., size, layout matter the most → **Method**: Conjoint analysis
- Measure willingness to pay for entire home & privacy features → **Methods:**
  - **A/B test Miami search results**: lightly boost listings with +10–20% more square footage or those "entire homes"
  - **Track micro-conversion metrics**: click-through rate, save-rate, and booking rate changes.
  - **Compare against control markets** e.g., Chicago, Toronto to isolate Miami's leisure-travel nature

Paris

1. Improving pro-host listing performance

- Determine why pro-host units feel less personalized
- Test redesigns/ charm-boosting changes for pro-host listings

→ **Methods**: Focus groups, pricing tests, persona segmentation

2. Host warmth & communication expectations

- Identify ideal response times and communication styles
- Evaluate impact of a "Hosted With Care" badge

→ **Methods**: NLP analysis of host messages; Sentiment studies, guest surveys

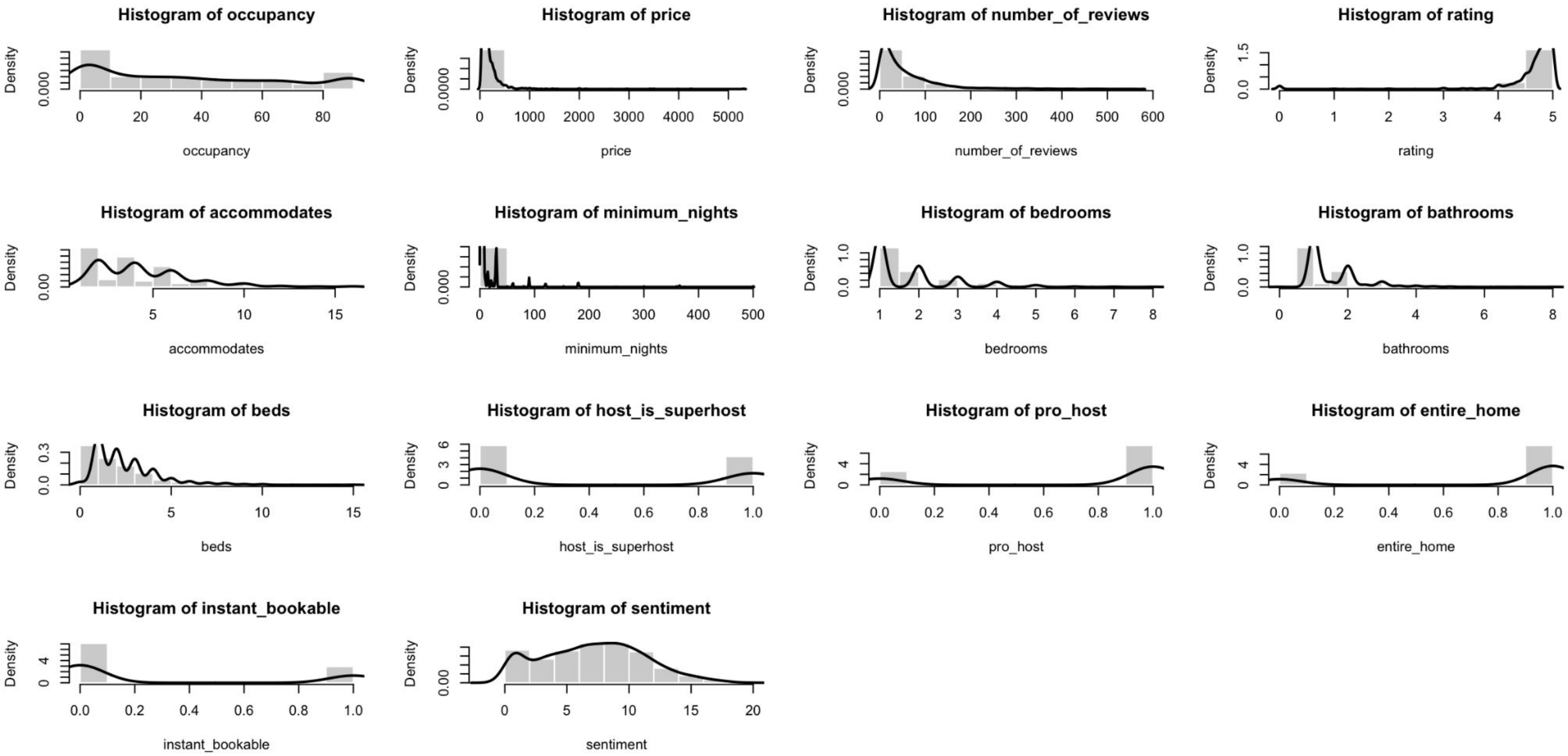
3. Price sensitivity & flexible-stay preferences

- Quantify impact of pricing and minimum-night restrictions on bookings
- Test messaging for flexible-stay offerings

→ **Methods**: Booking funnel analysis, A/B tests, traveler surveys

Appendix

Skewness check: Miami



Skewness check: Paris