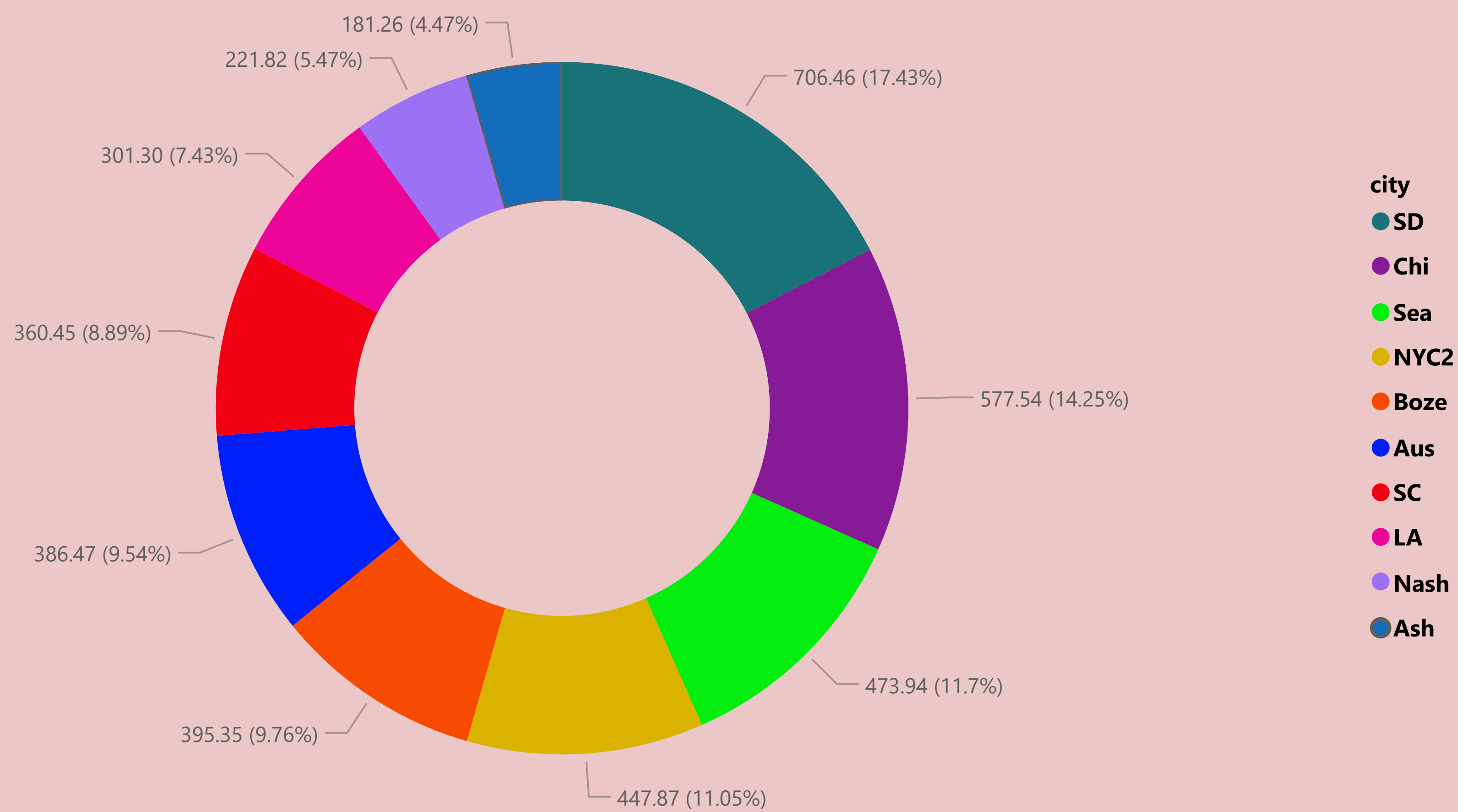
Power BI Sandbox: Airbnb Edition

experimentation and play

Average of price by city



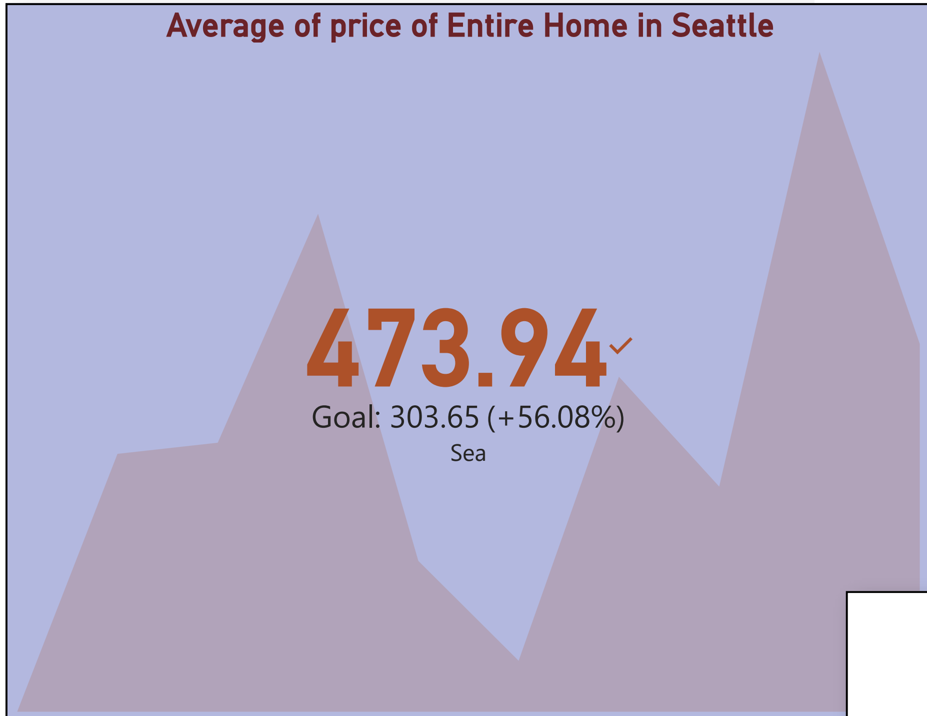
Abeviation	Location
Ash	Ashville, NC
Austin	Austin, TX
Boze	Bozeman, MT
Chi	Chicago, IL
LA	Los Angeles, CA
Nash	Nashville, TN
NYC2	New York, NY
SC	Santa Clara, CA
SD	San Diego, CA
Sea	Seattle, WA

Susan Schnitzel · Portfolio Draft with PowerBI · October 2025

Quick measure

My Initial Fun with KPIs

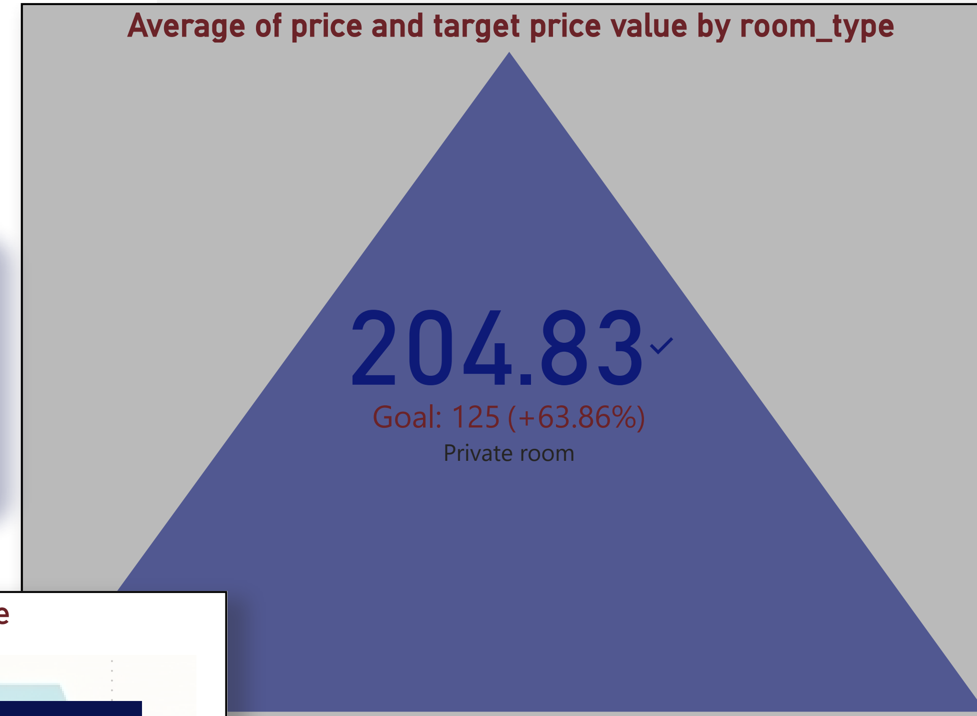
These early KPIs are just the beginning and are quick snapshots, set up to explore how values shift as more data and insights emerge. They're not final, but they help sketch the landscape and spark questions. As cleaning progresses and patterns surface, these metrics will evolve to reflect deeper learning and refined logic.

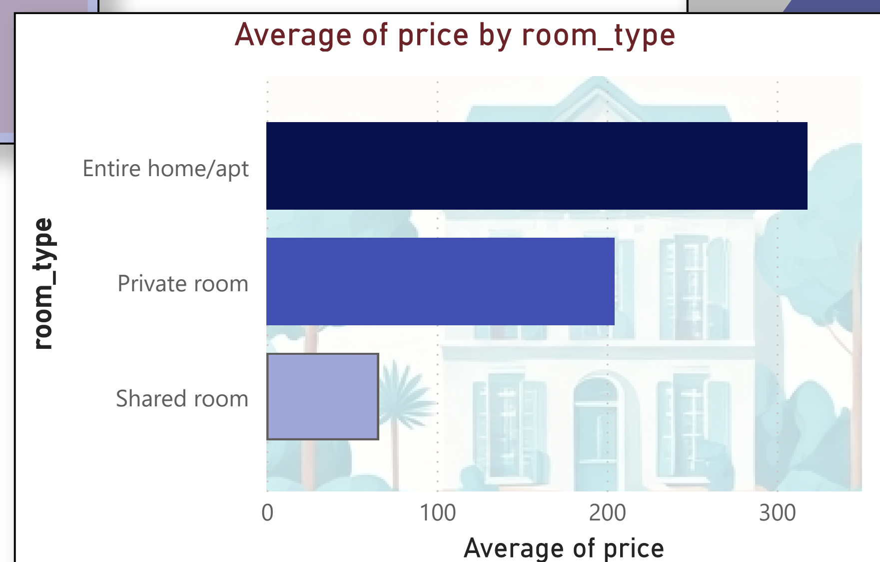


Sandbox metrics;
curious now, smarter later.

KPI

Pricing Performance





Quick measure

Airbnb Pricing Influencers -Decrease

Overview:

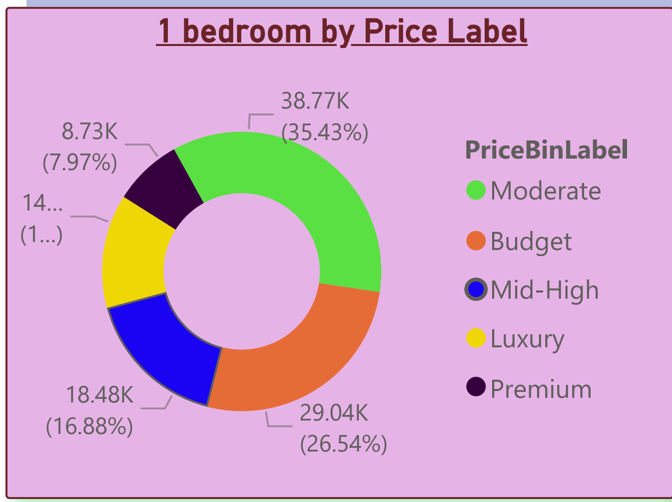
Moderate-tier listings now dominate the pricing landscape, suggesting a strategic sweet spot for hosts balancing affordability and profitability. Studio and 1-bedroom units remain influential in shaping price tiers, though 1-bedrooms show a slightly stronger downward impact on price. This may reflect listing conventions or pricing strategies in urban markets.

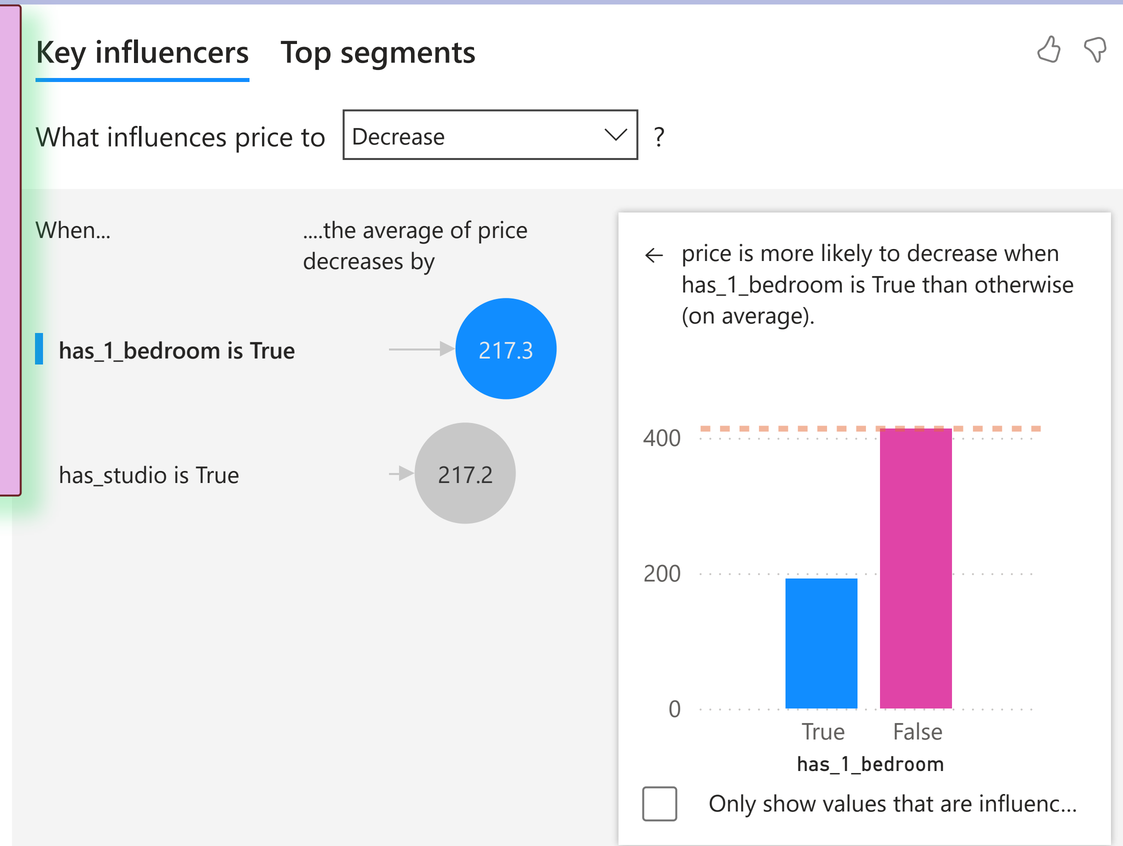
Visual Insight:

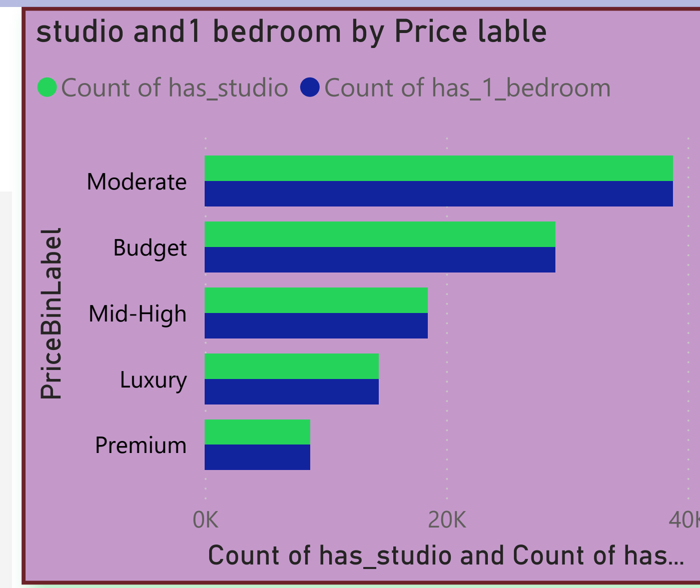
The donut chart reveals a peak in the "Moderate" bin, followed by "Budget" and "Mid-High." "Luxury" and "Premium" bins represent the smallest segments, indicating a skew toward mid-range pricing. Bar charts show studio and 1-bedroom listings share nearly identical pricing patterns, implying semantic overlap and possibly due to shared amenities or regional classification norms.

Influence Summary:

- **1-Bedroom units exert the strongest downward pressure on price, as shown in the Key Influencer visual.**
- **Studios, while close in impact, may be priced higher in premium urban markets or classified interchangeably with 1-bedrooms.**
- **Premium listings, though rare, may represent outliers with unique features or high-demand locations that are ideal for a follow-up outlier analysis.**

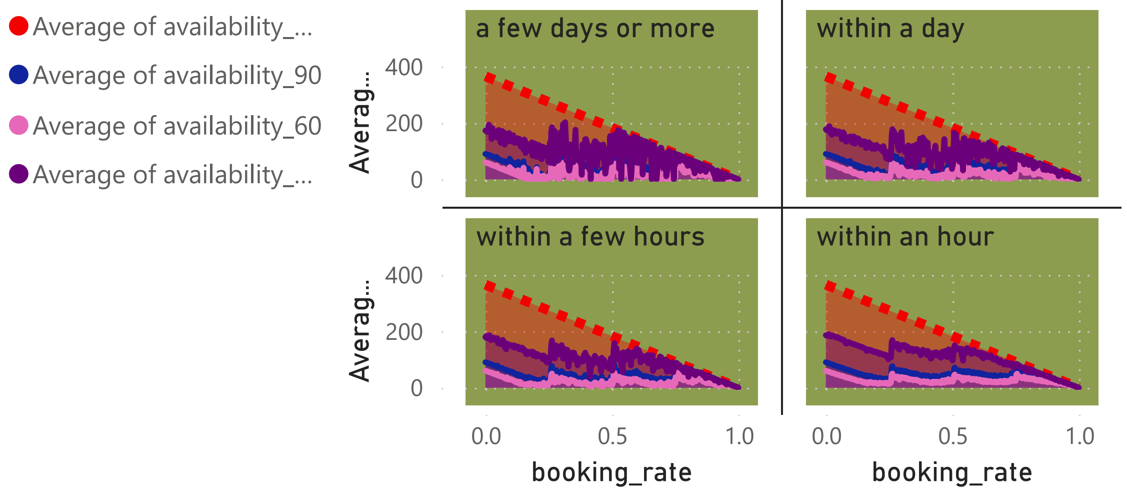






Quick measure

Days of Availability vs. Booking Rate



RED FLAG ALERT: 365-Day Line looks too clean

The regression line for **365-day availability** is perfectly straight and tightly correlated with booking rate, that could signal:

- **Synthetic or default values**: Some platforms auto-fill availability as 365 for listings that don't specify a calendar, especially inactive or placeholder listings.
- **Overfitting or lack of variance**: If most 365-day listings are professionally managed, they may behave similarly—leading to a tight, linear pattern.
- **Data artifact**: If blanks or partial calendars were removed, you may have unintentionally filtered out the "messy middle," leaving only high-performing listings.

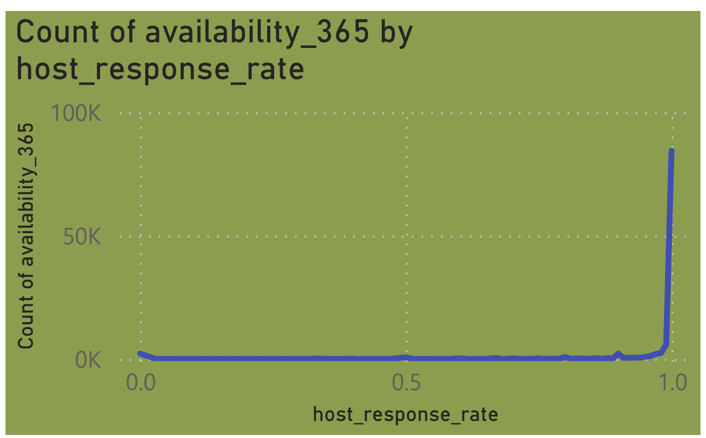
INVESTIGATIONS/OTHER ANALYSIS NEEDED AND MIGHT NEED TO CLEAN THE VARIABLE..... OR DROP

Quick measure

Graph Behavior: Host Rate vs. 365 Availability

- **The count of 365-day listings vs. host response rate** hugs the axis until it nears a host rate of 1, then **shoots up to 80K**.
- That spike suggests **heavy weighting or clustering which is** likely a small number of high-response hosts managing a large volume of always-available listings.
- Could be professional hosts, property managers, or platform artifacts.

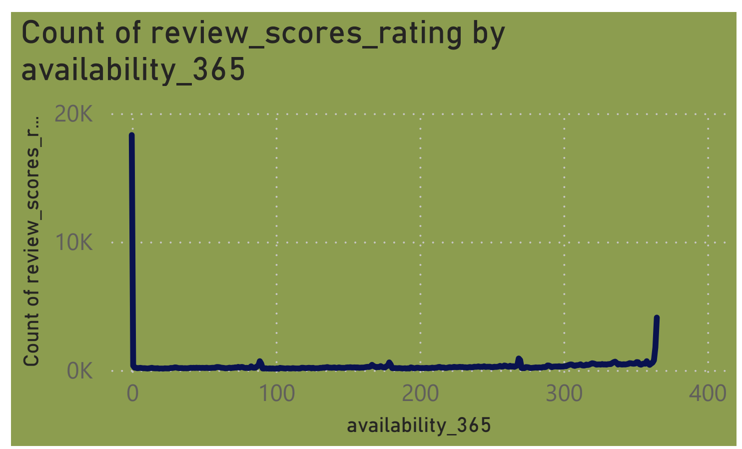
*This is a classic case of **data distortion by outlier***



Segmenting by Property Type & City

- **Unusual property types** (caves, trains, tipis, boats, dorms, etc.) are disproportionately represented in the 365-day availability bucket.
- These are clustered in **Seattle, LA, and NYC which are** markets known for eclectic listings and high tourism.
- But when segment by **city alone**, the highest 365-day counts show up in **Nashville, Asheville, and Chicago which is** suggesting a different dynamic, possibly tied to regional hosting norms or platform defaults.

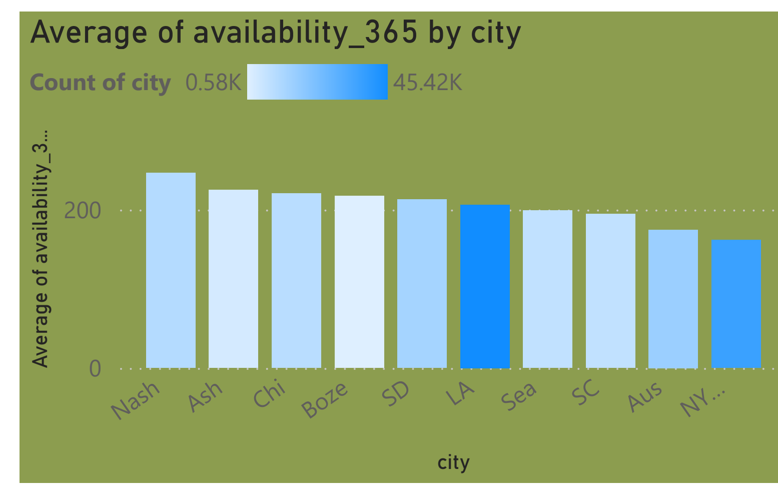
*This contrast between **property type vs. city-level behavior** shows how semantic quirks and geographic context can skew metrics.*



Reviews vs. 365 Availability

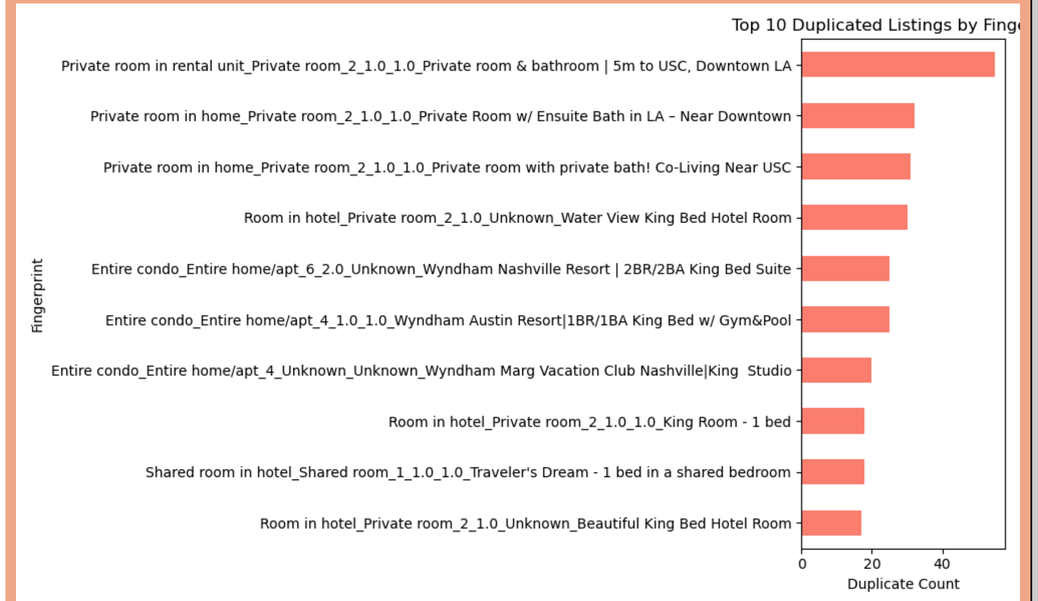
- At **review count = 0**, the graph **spikes**, then drops and hugs the axis until ~300 reviews, where it climbs again and nudges upward at 356.
- This pattern suggests:
  - A large number of **inactive or new listings** with 365-day availability and zero reviews.
  - A second tier of **established listings** with high review counts and consistent availability.

This dual behavior could be a great visual for a "Lifecycle of Listings" slide by showing how availability and engagement evolve. (Will look at in future.)



Detecting Repeat Listings & Platform-Managed Hosts

DUPLICATE LISTINGS - YIKES

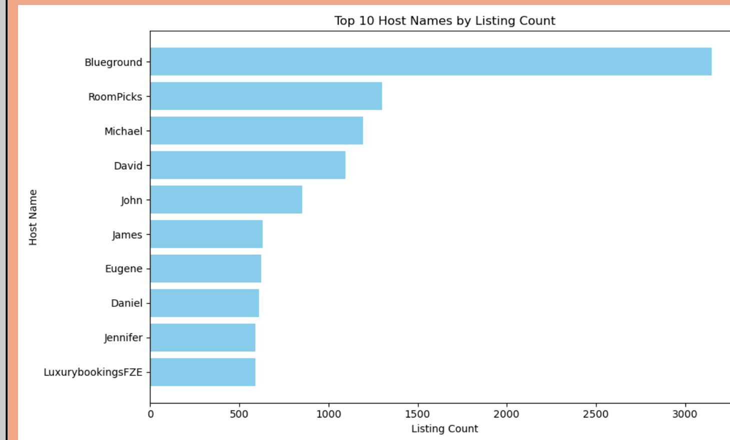


host_id	Count of id
107434423	3148
3223938	598
446820235	589
501999278	589
533234561	335
501999514	334
19303369	254
338667310	251
468914943	245
162280872	239
200239515	226
204704622	215
126644161	213
51501835	183
Total	**145515**

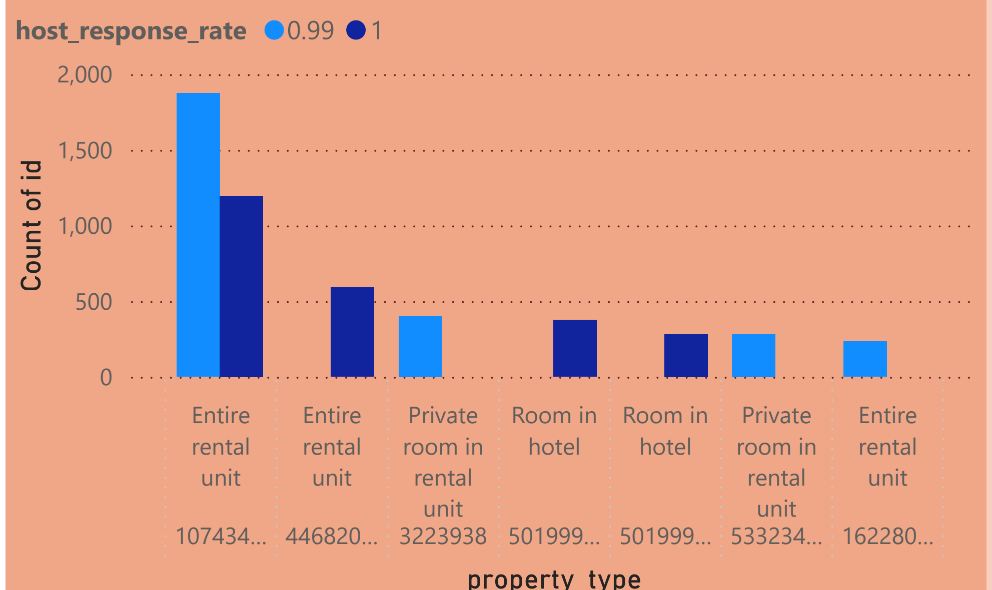
Listing Count and Red Flags TOP 4

3,148
🚩 Very likely platform-managed or large property group
598
🚩 High volume will need check for generic host name or serviced units
589
🚩 Same as above and possible duplication or commercial ops
501999278
589
🚩 Duplicate count and again could be same host or mirrored listings

TOP Ten Hosts (Number of Listings) Python Visual



Few Listings w/ Hosting Rates

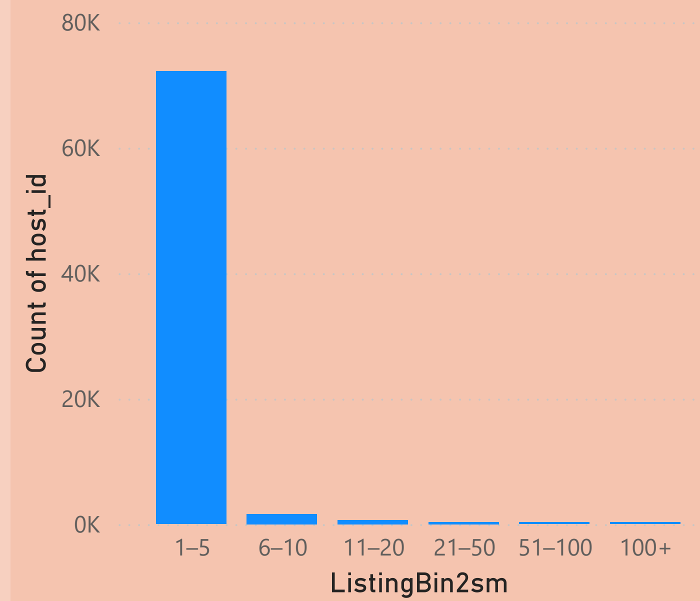


While analyzing Airbnb listing data, I identified several hosts with **extraordinarily high listing counts with** some exceeding 3,000 listings. These hosts exclusively offer **entire rental units**, and their listings often share identical or near-identical attributes.

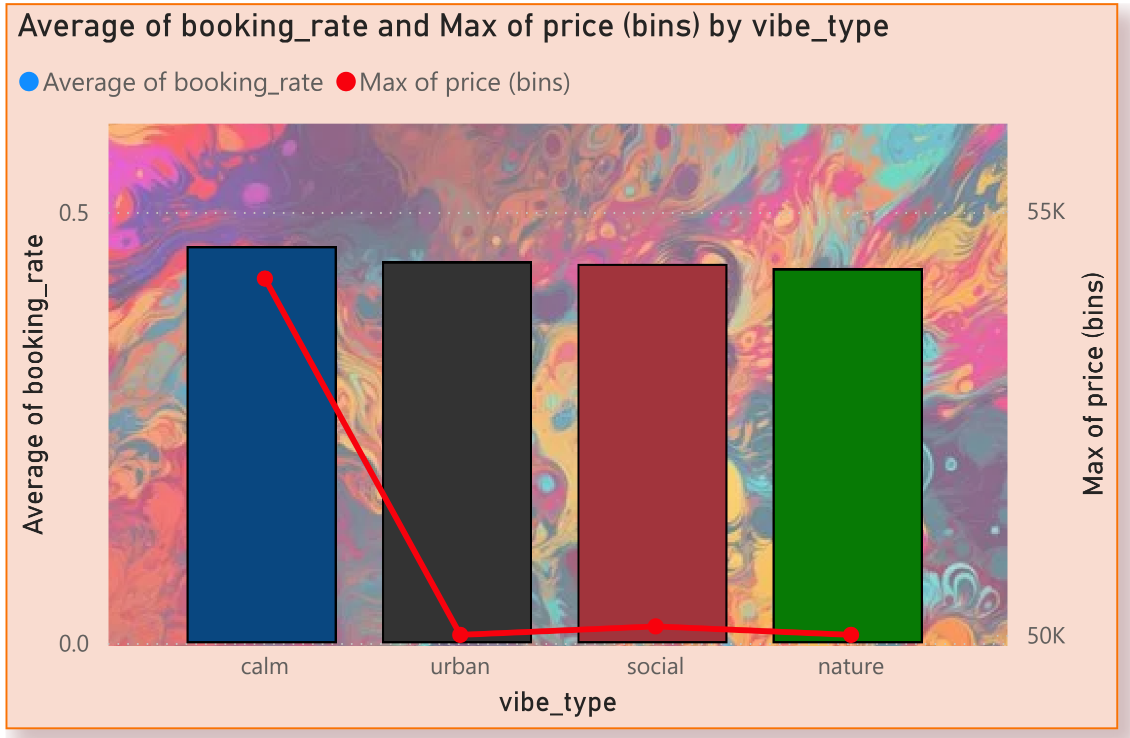
Using Python in Power BI,

- Grouped listings by host_name and host_id to flag reused names across multiple IDs
- Created a semantic **fingerprint** using property_type, room_type, accommodates, bedrooms, bathrooms, and name
- Identified **duplicate fingerprints**, indicating templated or cloned listings
- Visualized top duplicated listings and high-volume hosts to surface potential **platform-managed behavior**

Count of host_id by ListingBin2sm



« Quick measure

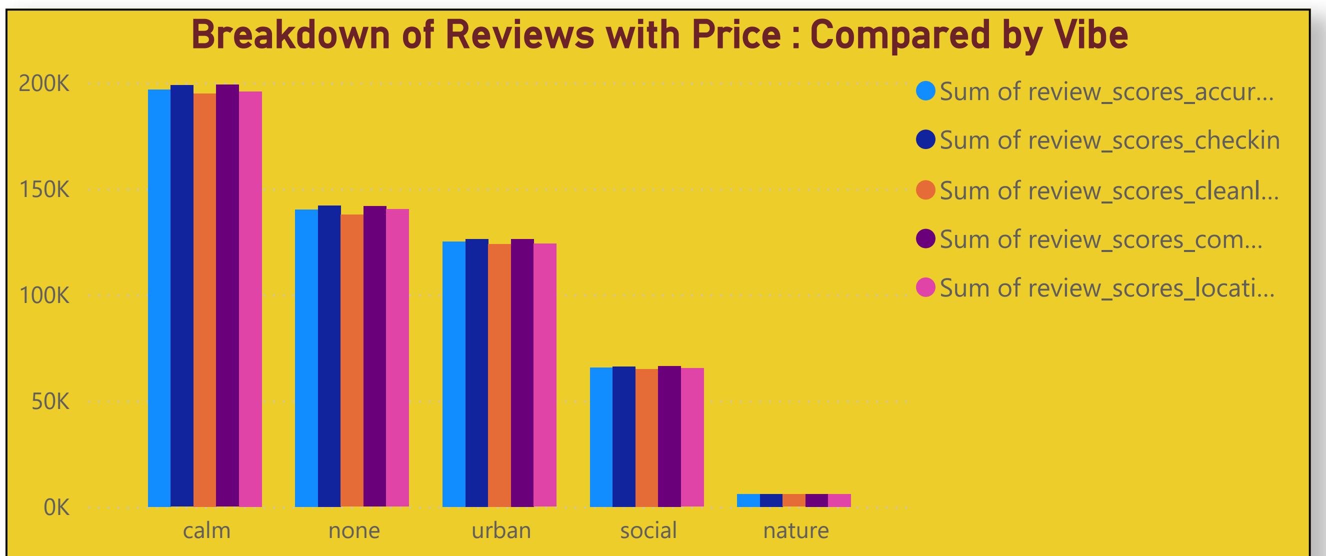


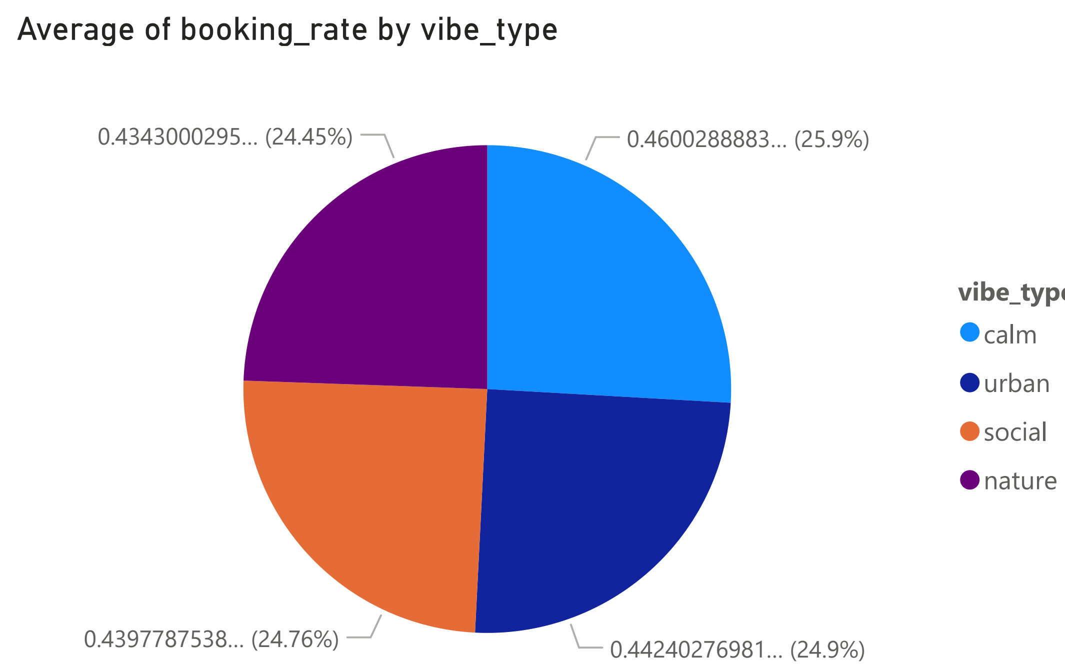
<u>Emotional Tone Insight: The “Calm” Premium</u>

Listings associated with the **“calm” vibe** consistently outperform others across key metrics:

- **Higher review scores** and more emotionally positive feedback
- **Slight lead in booking rate**, suggesting stronger guest resonance
- **Significantly higher max price**, indicating pricing elasticity tied to emotional framing.

This positions “calm” as a semantic lever and where emotional language drives both engagement and revenue.





« Quick measure

Insights from "Reviews, Estimated Occupancy 365, Price "Bins and Cities Comparisons

Key Insights from the Data

1. **Occupancy Drives Review Quality** Listings with higher **estimated occupancy over 365 days** consistently show **higher average review scores**. This suggests that sustained guest engagement correlates with better experiences or at least more favorable feedback.
2. **Price Elevates Perceived Quality** As nightly prices increase, **average review scores also rise**, with **luxury-tier listings** earning the highest ratings. However, this trend reflects perceived value rather than volume, because luxury listings, while highly rated, tend to receive fewer reviews due to lower booking frequency.
3. **Review Volume Peaks Mid-Tier** The **total number of reviews** peaks in the **moderate price bin**, then drops sharply in higher tiers. This indicates that while luxury listings are rated highly, they're reviewed less frequently and likely due to lower booking volume or exclusivity.

City-Level Highlights

Nashville

Highest average occupancy and suggesting strong year-round demand

Asheville

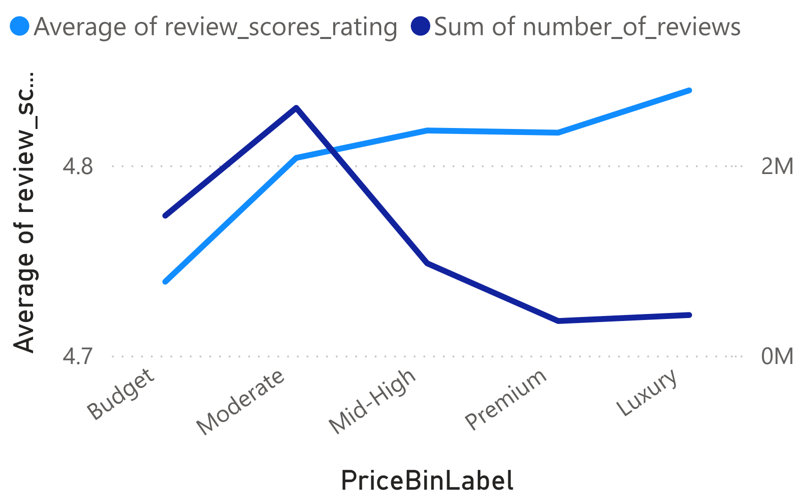
Highest average review count and indicating high guest turnover or engagement

San Diego & Chicago

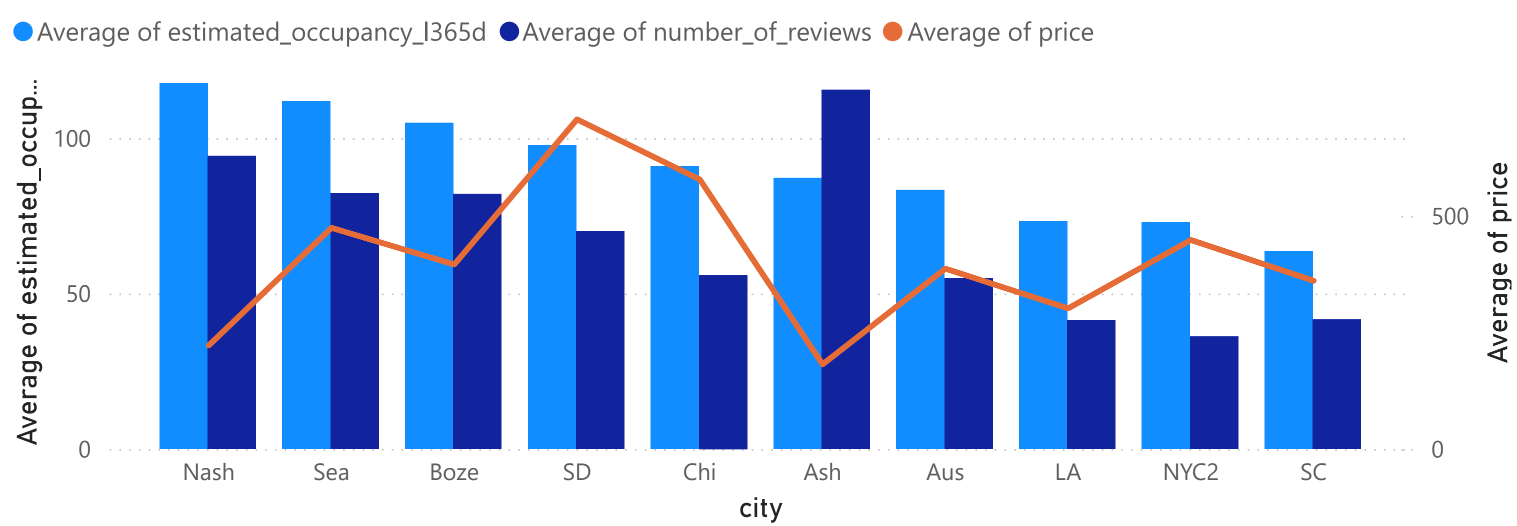
Highest pricing relative to occupancy and suggesting premium positioning or market inflation

Next Steps for Deeper Diagnostics

- **Normalize review scores by occupancy** to isolate quality from volume
- **Segment by property type and vibe** to see how emotional tone interacts with pricing and reviews
- **Map pricing elasticity** across cities to identify where high prices still yield strong occupancy and ratings
- **Scaffold mentoring assets** that teach how semantic signals (like "calm" or "luxury") shape guest perception and host strategy



Average of estimated_occupancy_l365d, Average of number_of_reviews and Average of price by city



«

Quick measure

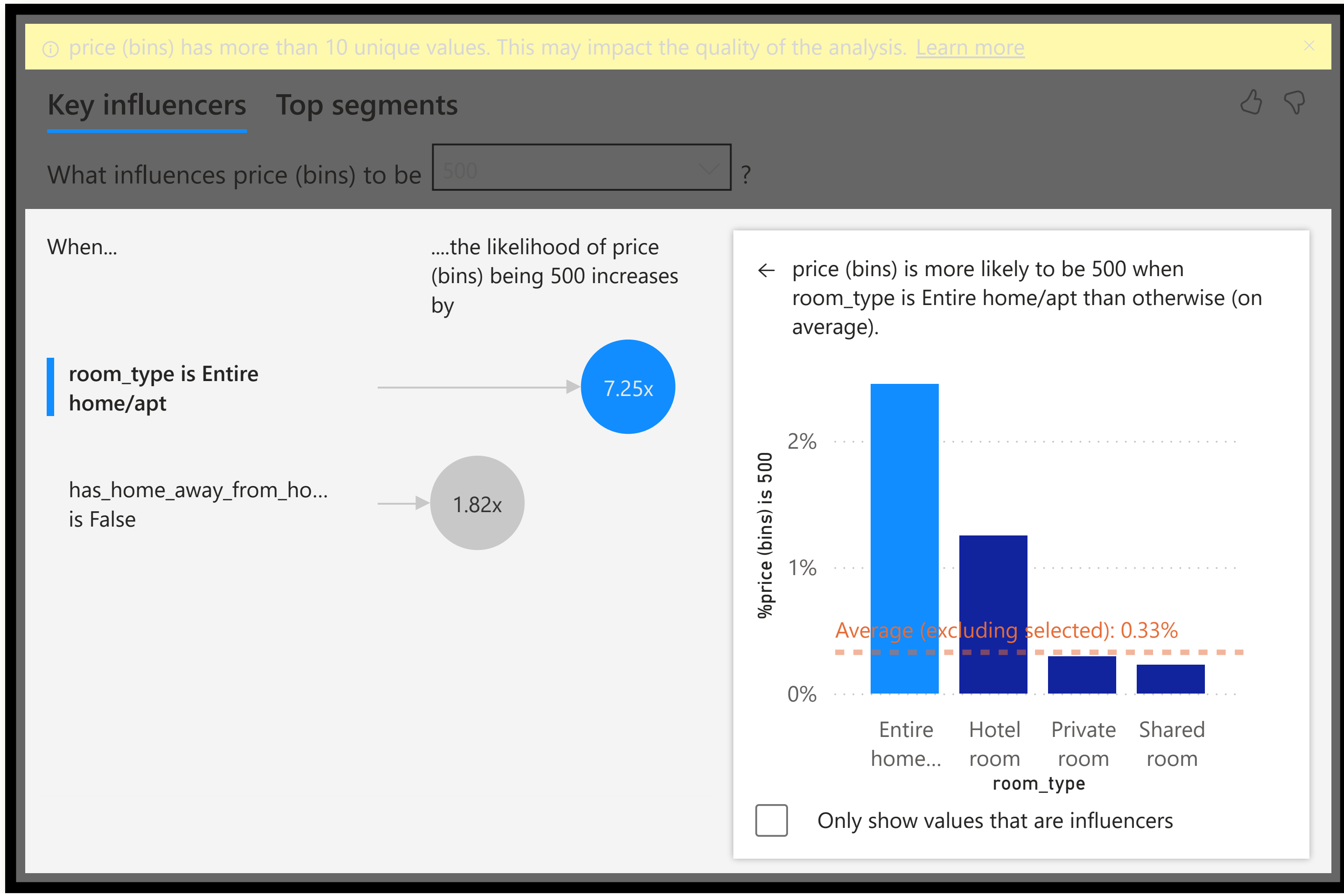
Room Types - Key Influencers

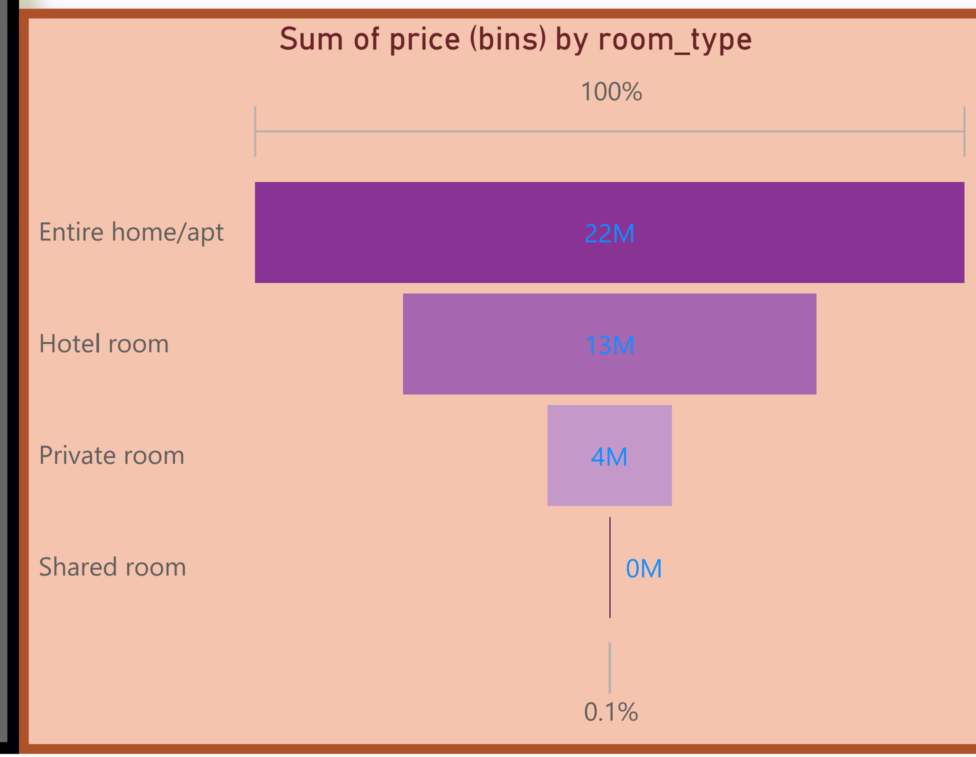
Room Type Insights & Early Observations

The sum of prices by room type confirms what we'd expect: **Entire home/apt listings dominate total revenue**. However, **Hotel Room** listings also show a surprisingly high price sum. This isn't an anomaly at first glance, but it warrants deeper investigation. A follow-up slide will explore whether this spike is due to:

- High-end hotel listings in premium areas
- Misclassified room types or property types
- Regional quirks in labeling or pricing strategy

Additionally, early analysis suggests that listings using the phrase **"home away from home"** do **not** correlate with higher prices. This could be a key insight for hosts: perhaps travelers aren't seeking familiarity, but rather **novelty and escape**. Messaging that emphasizes **uniqueness or adventure** may resonate more with guests and influence pricing potential.





Quick measure

Key Influencers of Price: Bathrooms, Bedrooms & Property Type

Core Insight

The strongest predictors of higher pricing in Airbnb listings are **bathroom and bedroom counts**, with a clear upward trend in price as these features increase:

- Listings with **more than 4.5 bathrooms**, **more than 4 bedrooms**, and **more than 4 beds** show the highest price lifts.
- These variables exhibit a **strong positive correlation** with price, confirming that scale and capacity are key drivers of value perception.

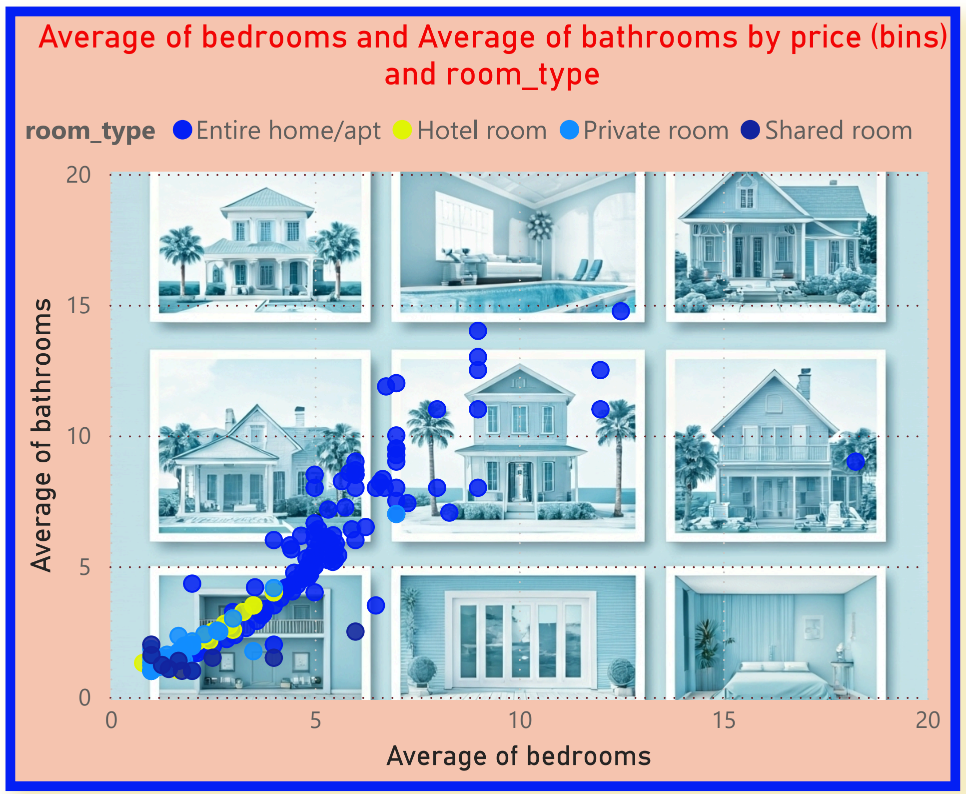
Property Type Matters

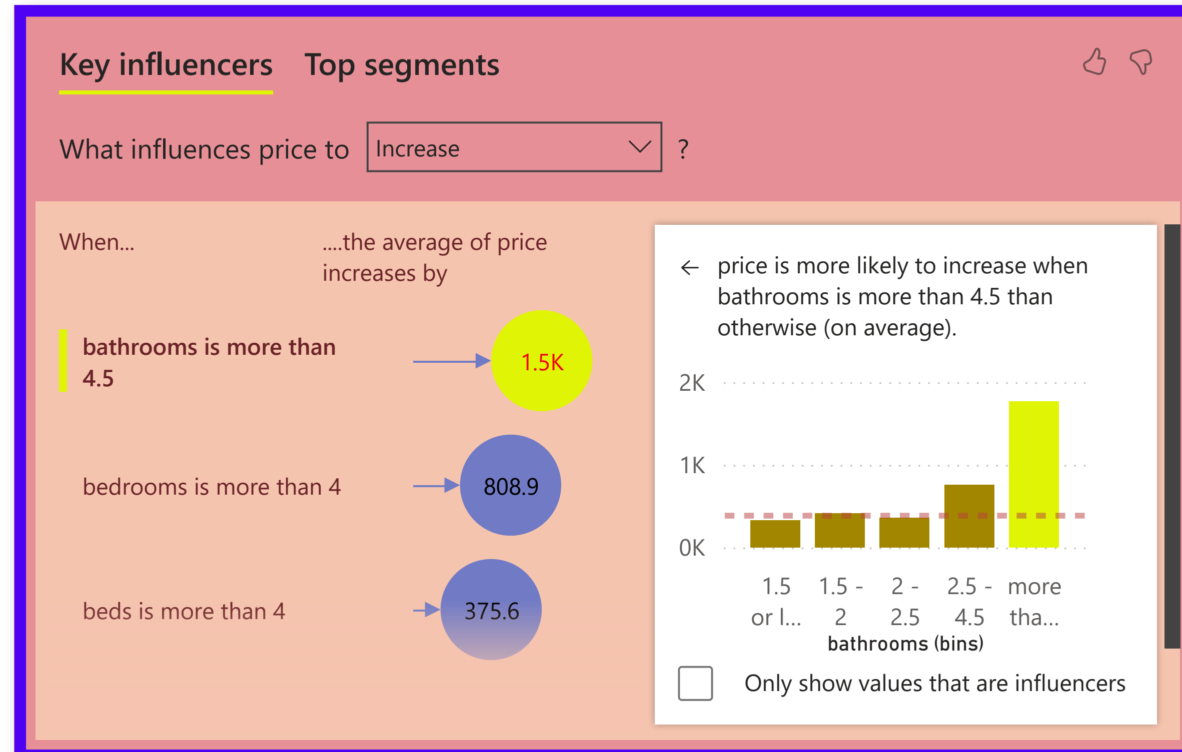
Unsurprisingly, listings categorized as **entire homes or apartments** dominate the upper tiers of bathroom and bedroom counts.

These property types:

- Offer the most space and amenities
- Align with family travel, group bookings, or luxury positioning
- Are more likely to appear in premium pricing bins

This reinforces the idea that **room count and property type are semantically linked** and together, they shape pricing strategy.





Visual Takeaways

- **Key Influencer visuals** show price increases of:
  - ~$1,500 for listings with >4.5 bathrooms
  - ~$800 for listings with >4 bedrooms
  - ~$375 for listings with >4 beds
- **Bar charts** confirm that entire homes/apartments dominate these high-count categories

Next Steps

1. **Segment by vibe and location** to see how emotional tone interacts with room count
2. **Normalize price by occupancy** to identify high-yield configurations
3. **Scaffold mentoring slides** that teach how structural features influence pricing
4. **Explore outliers**—e.g., listings with high room counts but low prices (potential misclassification or underpricing)

Quick measure

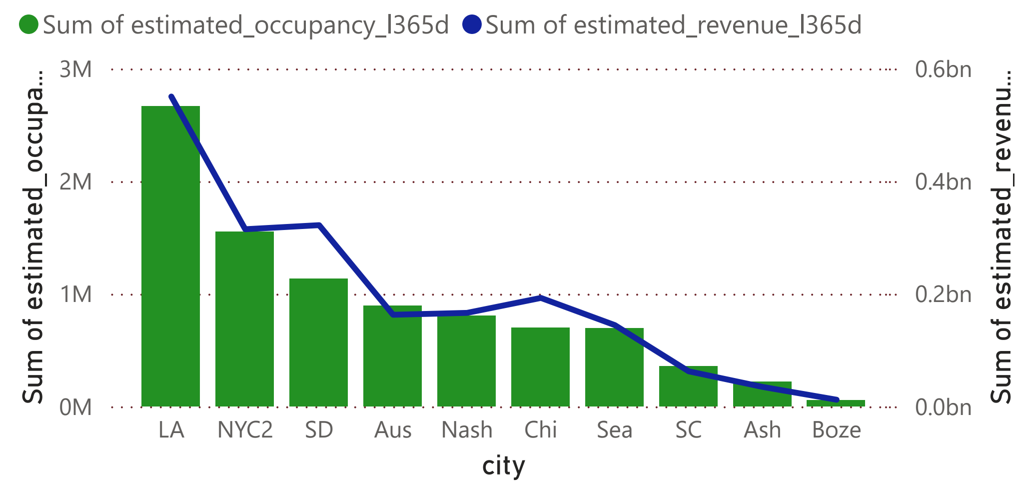
Revenue & Occupancy: Uncovering the Financial Pulse of Airbnb Listings

Key Insight
Incorporating **estimated revenue** into the analysis reveals a strong, direct correlation with **estimated occupancy over 365 days**. Listings that are booked more consistently tend to generate significantly higher revenue and confirming that occupancy is a primary driver of financial performance.

City-Level Revenue Trends
Unlike previous visuals that highlighted **price leaders** like San Diego and Chicago, the revenue analysis tells a different story:

Rank City and Revenue Insight
🥇 1
Los Angeles
Leads in total estimated revenue, driven by high occupancy and mid-to-high pricing
🥈 2
New York City
Strong revenue performance despite competitive pricing, which is likely due to year-round demand
🥉 3
San Diego
High pricing contributes, but occupancy lags slightly behind LA and NYC
This shift from price-centric to revenue-centric analysis highlights the importance of **booking consistency** and **market saturation** over raw nightly rates.

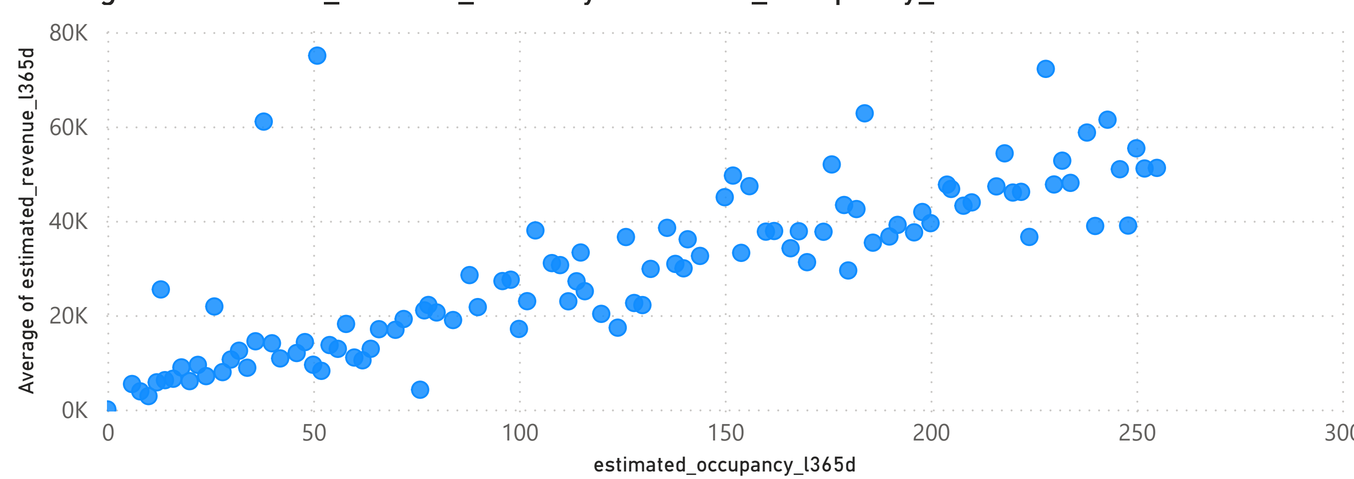
Sum of estimated_occupancy_l365d and Sum of estimated_revenue_l365d by city



Strategic Takeaways
- **Occupancy is the engine:** price alone doesn't guarantee revenue
- **City dynamics matter:** LA and NYC outperform due to volume and demand, not just pricing
- **Revenue ≠ Price:** high-priced listings may underperform if occupancy is low

Next Steps
1. **Normalize revenue by occupancy** to identify high-yield listings
2. **Segment by vibe and property type** to see which emotional tones or formats drive revenue
3. **Map revenue vs. review scores** to explore whether guest satisfaction aligns with financial success
4. **Scaffold mentoring assets** that teach how to pivot from pricing analysis to revenue diagnostics

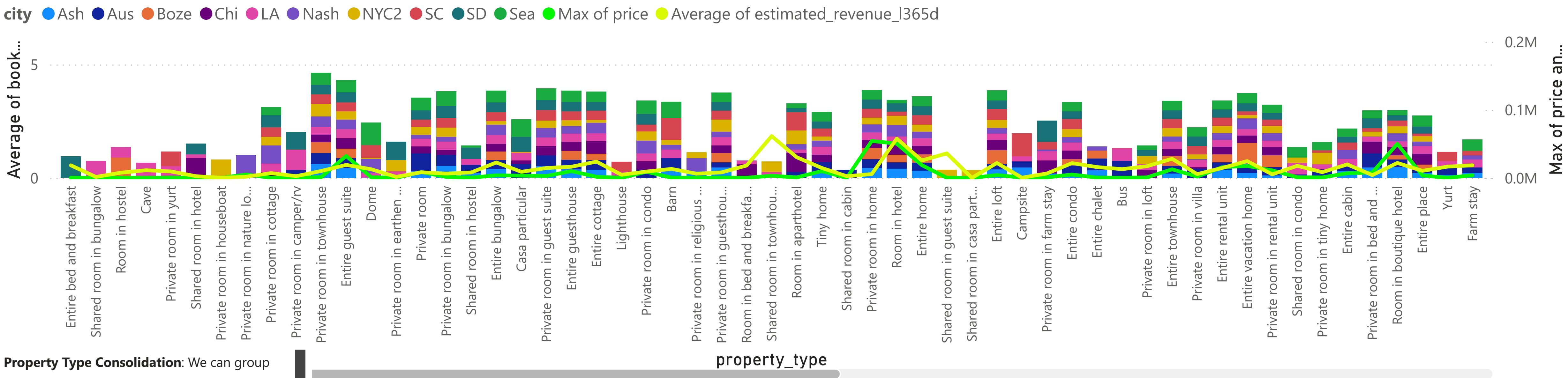
Average of estimated_revenue_l365d by estimated_occupancy_l365d



« Quick measure

Exploring Long-Form Visuals in Power BI: Property Type Trends & Market Signals

This visual may be long and too long for a static export, but that's part of its strength. On the Power BI platform, the scroll bar transforms it from overwhelming to interactive, allowing us to explore a rich landscape of property types, city-specific patterns, and pricing spikes. Beyond its playful aesthetic, this graph is packed with actionable insights.

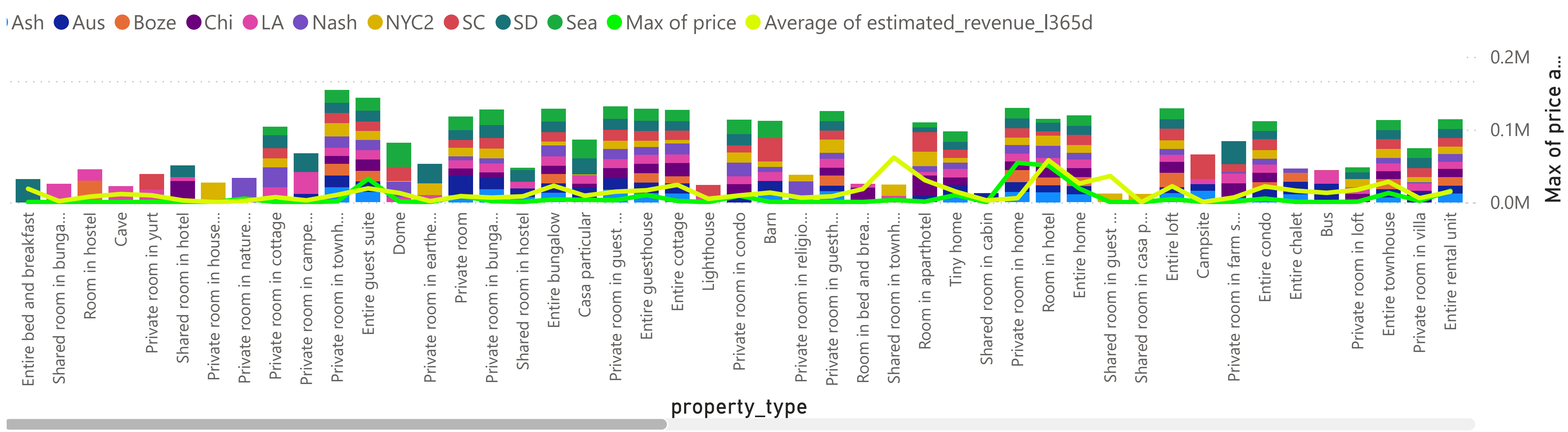


Property Type Consolidation: We can group similar shared room types (standalone, in-home, hostel, bungalow) to better understand broader trends while preserving granularity for deeper dives.

City-Level Uniques: Certain cities show distinct property type distributions and pricing behaviors. Segmenting by city could reveal localized market dynamics and help tailor strategies.

Revenue & Price Spikes: These anomalies are goldmines for market targeting. Investigating the causes such as seasonality, events, or platform behavior, which could uncover opportunities.

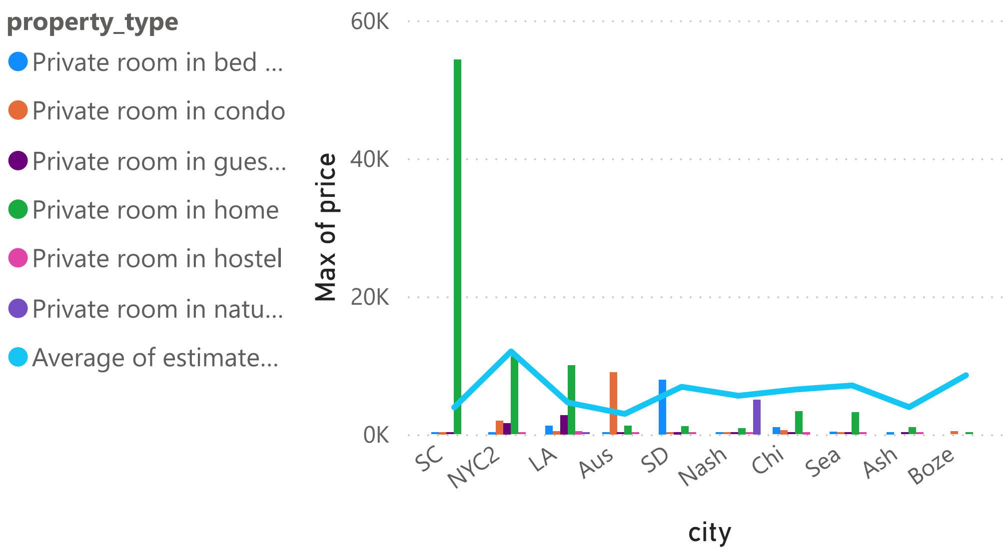
Unusual Property Types: Listings like caves and castles add a whimsical layer, but they also challenge conventional pricing models. Based on our previous findings, they likely diverge from the "Home Away from Home" category, which showed a negative price impact already.

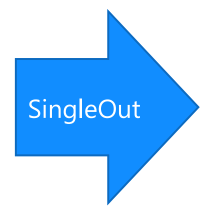


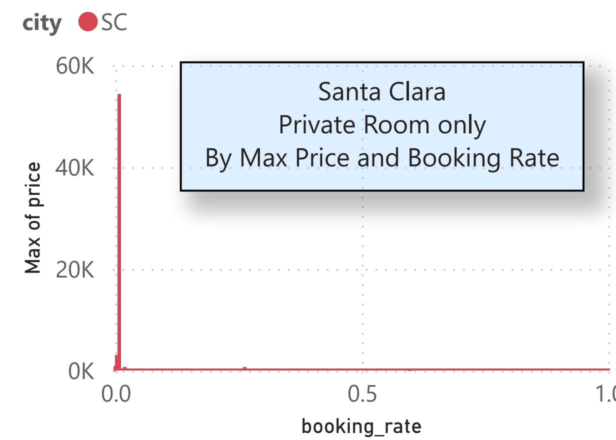
Quick measure

Private Rooms Across Dwellings: A Quick Diagnostic Sweep

A high-level scan of Private Room listings across various dwelling types revealed a standout anomaly







property_type	Max of price	Average of estimated_revenue_l365d	Average of estimated_occupancy_l365d
Private room in home	54280	5621.52	57.87
Private room in rental unit	11078	5230.97	33.21
Private room in condo	8998	9342.75	56.85
Private room in bed and breakfast	7920	4130.18	35.50
Private room in nature lodge	5000	1800.00	7.14
Private room in guest suite	2800	14537.92	99.55
Private room in villa	1940	5562.79	48.16
Private room in resort	1508	2717.45	8.48
Total	**54280**	**6072.80**	**47.91**

OVERVIEW

A **Private Room in a Home** priced at **$54,000 and** located in **Santa Clara**.
I filtered for *Private Room in Home* listings in Santa Clara and compared them by **booking rate**.
RESULTS:

- A **single spike** in both price and booking activity, with the rest of the listings clustering at far more reasonable levels.

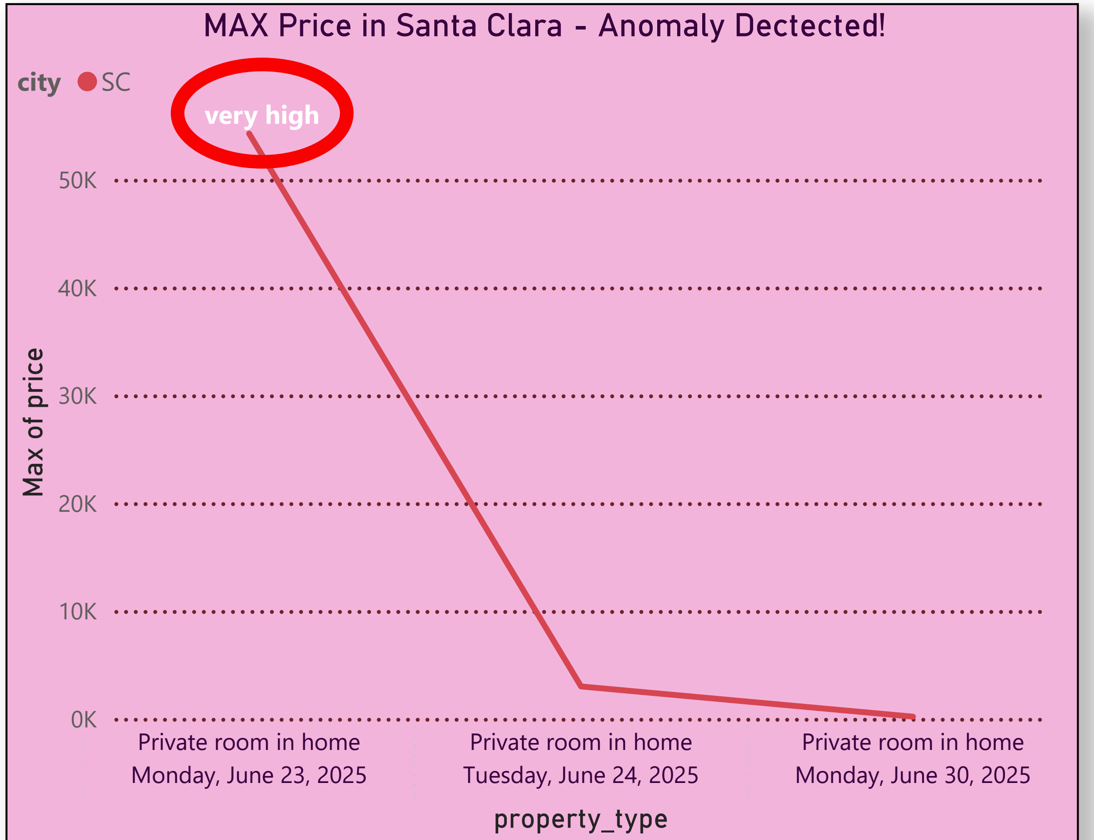
This strongly suggests an **outlier**, possibly due to:

- Data entry error (e.g., extra zero)
- Misclassified property type
- A one-off luxury listing with atypical demand

Ongoing Data Hygiene

As always, this is a reminder that:
Cleaning data is like cleaning a house; no one notices until it's a mess, and the job is never really done.
This anomaly is now flagged for further investigation, and it's a great candidate for a reusable **anomaly detection module** or **semantic fingerprinting guide**.

Quick measure



MORE TO COME....MORE TO CLEAN....MORE TO ANALYSIS AND ......MORE VISUALIZATIONS TO CREATE.

MORE FUN WITH DATA!

Quick measure