

Data Analysis Project

Airbnb Los Angeles

Karen Adjoa Selasi Koomson (G#: G01541691)

2026-02-22

1. Introduction

The consulting company seeks to understand the factors that contribute to Airbnb rental prices in Los Angeles. Specifically, the goal of this project is to identify which property characteristics allow hosts to charge higher prices.

The dataset consists of 45,533 Airbnb listings in Los Angeles as of September 4, 2024, sourced from Inside Airbnb. The dataset includes listing-level information such as neighborhood, room type, number of bedrooms, review scores, availability, and amenities.

This analysis addresses the following big questions:

Does room_type affect price? Does property_type affect price? Do neighbourhoods affect price? Does the number of accommodates (guests) affect price? Does being a superhost (host_is_superhost) affect price? Do review_scores_rating and number_of_reviews correlate with price?

The remainder of the report presents the data, methods, results, and conclusions derived from exploratory data analysis.

2. Data and Models

2.1 Data

The dataset consists of 45,533 Airbnb listings in Los Angeles and 25 variables describing listing characteristics, host information, geographic location, and pricing.

Before analysis, the data were cleaned and reformatted to ensure consistency. Date variables were converted to proper date format, response rates were converted to numeric values, and binary indicators such as superhost status and instant bookable were converted to logical variables. Categorical variables including room type, property type, and neighborhood were converted to factors to facilitate grouping and summary analysis. Missing values were retained where appropriate.

These preparation steps ensured that the dataset was suitable for exploratory data analysis.

Question 1: Does Room Type Affect Price?

Rationale: Entire homes, private rooms, and shared rooms likely have different price ranges.

Data: The dataset contains 45,533 Airbnb listings in Los Angeles with the following relevant variables: price – listing price in USD room_type – type of room: Entire home/apt, Private room, Shared room, Hotel room Other variables exist, but for this question we are focusing on price and room_type.

Method: To determine whether room type influences Airbnb pricing, we compared listing prices across the four room categories: Entire home/apt, Private room, Shared room, and Hotel room.

Listings with missing or zero prices were removed. Extreme price outliers above \$2,000 were excluded to prevent distortion of summary statistics and visualizations. The room_type variable was treated as a categorical factor.

We calculated summary statistics (mean, median, standard deviation, and listing count) for each room type and created a boxplot to compare price distributions.

Analysis and Results:

```
# Clean data for this question
Airbnb_clean1 <- Airbnb %>%
  filter(!is.na(price), price > 0, price <= 2000) %>%
  mutate(room_type = as.factor(room_type))

# Summary table
price_by_room <- Airbnb_clean1 %>%
  group_by(room_type) %>%
  summarise(
    n_listings = n(),
    avg_price = mean(price, na.rm=TRUE),
    median_price = median(price, na.rm=TRUE),
    sd_price = sd(price, na.rm=TRUE)
  )

kable(price_by_room,
      caption = "Summary of Airbnb Prices by Room Type",
      digits = 0)
```

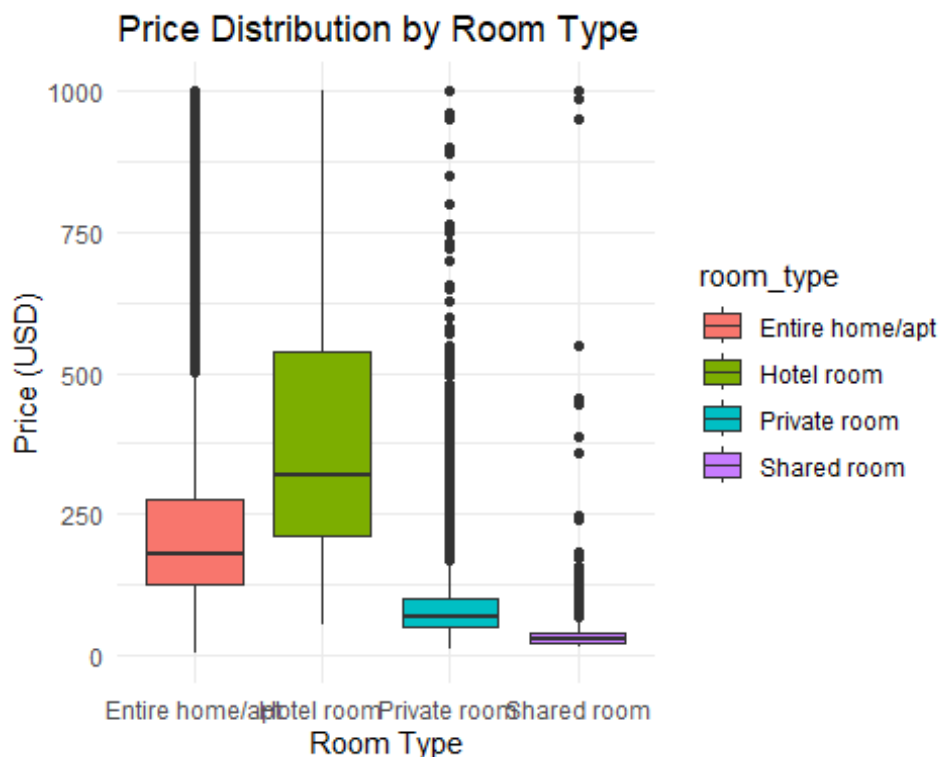
Summary of Airbnb Prices by Room Type

room_type	n_listings	avg_price	median_price	sd_price
Entire home/apt	28302	270	184	268
Hotel room	111	588	346	485
Private room	7865	95	69	102

room_type	n_listings	avg_price	median_price	sd_price
Shared room	506	49	30	91

```
# Boxplot
ggplot(Airbnb_clean1, aes(x = room_type, y = price, fill = room_type)) +
  geom_boxplot() +
  scale_y_continuous(limits=c(0,1000)) +
  labs(title="Price Distribution by Room Type", x="Room Type", y="Price (USD)") +
  theme_minimal()

## Warning: Removed 911 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```



Answer:

The summary table shows that Entire home/apt listings are the most common with 28,302 listings, and have an average price of \$270 and a median price of \$184, with a standard deviation of \$268. Hotel rooms, while much less common (111 listings), have the highest average price at \$587, though the median price is \$346, indicating that a few very expensive hotel listings are skewing the average. Private rooms are

significantly cheaper, with an average price of \$95 and a median of \$69, while shared rooms are the least expensive, averaging \$49 with a median of \$30.

The boxplot further confirms these patterns, showing that Entire home/apt listings generally command higher prices than private or shared rooms, while hotel rooms, though rare, can be very expensive.

Conclusion: Room type is a significant factor influencing Airbnb prices in Los Angeles. Entire homes and apartments consistently command higher prices compared to private and shared rooms, while hotel rooms, though rare, have the highest potential prices. This analysis highlights the importance of room type as a key determinant of rental price.

Question 2: Does Property Type Affect Airbnb Price?

Data: For this analysis, we examine the `property_type` variable, which indicates the type of listing such as Entire condo, Entire home, Private room in home, and other property types. The price variable remains the focus, measured in USD. As before, listings with missing or zero prices, as well as extreme outliers above \$2,000, were excluded to maintain a consistent and meaningful analysis.

Methods: We grouped the listings by `property_type` and calculated summary statistics for each group, including the number of listings, average price, median price, and standard deviation of price. This allows us to quantify how prices differ across property types. In addition, a bar chart was created to visualize the average price for each property type, helping to identify which types of properties command higher prices.

Analysis and Results:

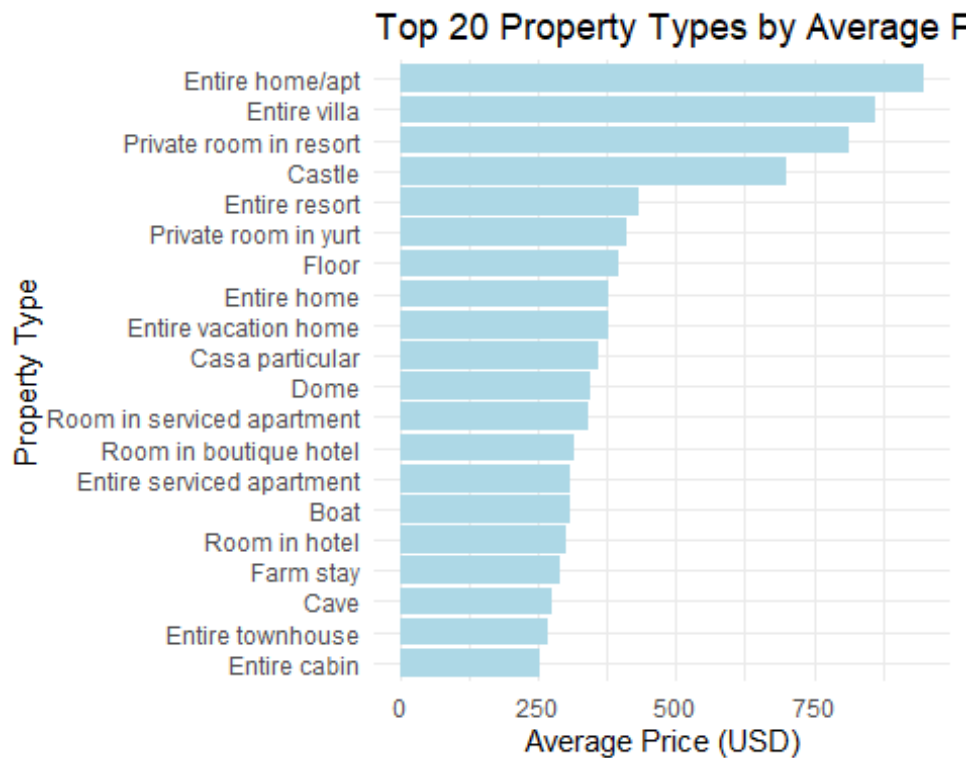
Top 20 Property Types by Average Price

property_type	n_listings	avg_price	median_price	sd_price
Entire home/apt	5	951	1180	837
Entire villa	472	863	834	483
Private room in resort	3	814	473	601
Castle	15	702	551	501
Entire resort	1	434	434	NA
Private room in yurt	2	412	412	88
Floor	1	395	395	NA
Entire home	10724	379	268	330
Entire vacation home	54	377	248	331
Casa particular	10	361	200	338

property_type	n_listings	avg_price	median_price	sd_price
Dome	4	343	156	406
Room in serviced apartment	5	341	346	7
Room in boutique hotel	127	314	207	365
Entire serviced apartment	402	307	233	258
Boat	12	307	244	291
Room in hotel	392	302	212	283
Farm stay	23	290	159	376
Cave	1	275	275	NA
Entire townhouse	624	267	229	156
Entire cabin	48	253	186	219
Entire place	53	248	150	341
Entire condo	1445	230	185	164
Entire bungalow	427	218	175	159
Tower	3	217	245	124
Shared room in bed and breakfast	11	212	52	306
Entire cottage	147	203	175	109
Tent	10	199	173	83
Cycladic home	1	192	192	NA
Room in aparthotel	65	189	179	142
Campsite	11	183	163	128
Entire loft	265	183	150	121
Private room in serviced apartment	9	182	75	273
Earthen home	7	181	177	51
Shepherd's hut	1	180	180	NA
Entire rental unit	9530	180	152	124
Train	3	175	181	23
Ranch	3	170	126	81
Room in bed and breakfast	1	166	166	NA
Hut	5	160	69	203
Private room in vacation home	2	154	154	76
Private room in bed and breakfast	139	152	90	145
Private room in hostel	21	151	129	82
Private room in ranch	1	150	150	NA
Minsu	1	148	148	NA
Entire guesthouse	2481	148	128	92

property_type	n_listings	avg_price	median_price	sd_price
Shipping container	7	145	125	91
Bus	2	144	144	21
Treehouse	7	140	131	44
Entire guest suite	1140	139	120	87
Tiny home	153	138	114	98
Tipi	1	135	135	NA
Private room in cabin	5	128	81	78
Barn	9	125	86	84
Yurt	10	118	120	56
Private room	28	116	88	98
Private room in loft	16	115	94	79
Camper/RV	128	114	97	68
Shared room in farm stay	1	110	110	NA
Private room in villa	196	106	65	168
Private room in guest suite	163	104	89	56
Lighthouse	1	100	100	NA
Private room in treehouse	5	100	110	24
Private room in guesthouse	83	100	85	54
Private room in earthen home	4	98	73	71
Private room in bungalow	102	97	65	155
Private room in barn	1	96	96	NA
Private room in condo	347	95	77	59
Private room in cottage	4	85	82	27
Private room in rental unit	1348	84	68	73
Private room in home	4594	83	65	91
Private room in townhouse	326	81	70	53
Private room in nature lodge	1	80	80	NA
Private room in casa particular	10	80	71	38
Private room in castle	2	76	76	1
Shared room in bungalow	7	74	19	126
Private room in dome	1	73	73	NA
Shared room in guesthouse	7	73	60	53
Private room in cave	2	70	70	14
Shared room in hotel	4	70	64	17
Shared room in condo	4	68	45	55

property_type	n_listings	avg_price	median_price	sd_price
Private room in tiny home	12	66	65	21
Shared room	2	66	66	49
Private room in farm stay	7	65	56	30
Shared room in boutique hotel	3	62	61	2
Private room in camper/rv	1	62	62	NA
Private room in riad	1	60	60	NA
Shared room in villa	3	58	40	36
Room in hostel	2	55	55	0
Shared room in rental unit	163	46	26	93
Shared room in cottage	1	45	45	NA
Shared room in hostel	47	42	35	63
Shared room in home	248	42	30	70
Private room in tent	1	40	40	NA
Private room in houseboat	1	36	36	NA
Private room in minsu	1	35	35	NA
Shared room in serviced apartment	5	33	24	25



Answer:

The bar chart of the top 20 property types by average price shows that property type strongly influences Airbnb pricing in Los Angeles. The most expensive listings are luxury or unique properties, including Entire villas, Castles, Entire resorts, and high-end Entire homes, with average prices well above \$700 per night.

Mid-range listings, such as Entire homes, Entire condos, and Entire guesthouses, tend to have average prices between \$150–\$400 per night. These represent the bulk of the market and offer typical accommodations for guests.

Cheaper options, including most private rooms and shared rooms, are priced between roughly \$35–\$150 per night, reflecting smaller spaces, lower privacy, or more basic amenities.

Overall, the chart demonstrates a clear pricing hierarchy by property type: luxury and entire properties command premium prices, while smaller or shared accommodations are more affordable. The top 20 visualization highlights the categories that most strongly influence pricing without overwhelming the reader with all 96 property types.

Question 3: Do neighborhoods affect Airbnb price?

Data Relevant variables for this question: price – listing price in USD
neighbourhood_cleansed – the cleaned neighborhood name where the listing is located

The goal is to examine how Airbnb prices vary across different neighborhoods in Los Angeles.

Methods Data Cleaning: Exclude listings with missing or zero prices. Remove extreme price outliers (above \$2,000) to focus on typical pricing. Convert neighbourhood_cleansed to a factor for grouping.

Summarization: Group listings by neighbourhood_cleansed and calculate: Number of listings (n_listings) Average price (avg_price) Median price (median_price)

Used a bar chart of average price by neighborhood. Display only the top 20 neighborhoods with the highest average prices. Flip the x-axis (coord_flip) for readability.

Analysis and Results:

```
# Summarize price by neighborhood
price_by_neighborhood <- Airbnb_clean1 %>%
  group_by(neighbourhood_cleansed) %>%
  summarise(
    n_listings = n(),
    avg_price = mean(price, na.rm=TRUE),
    median_price = median(price, na.rm=TRUE)
```



```

) %>%
  arrange(desc(avg_price))

# Display top 20 neighborhoods by average price
top_neighborhoods <- price_by_neighborhood[1:20, ]

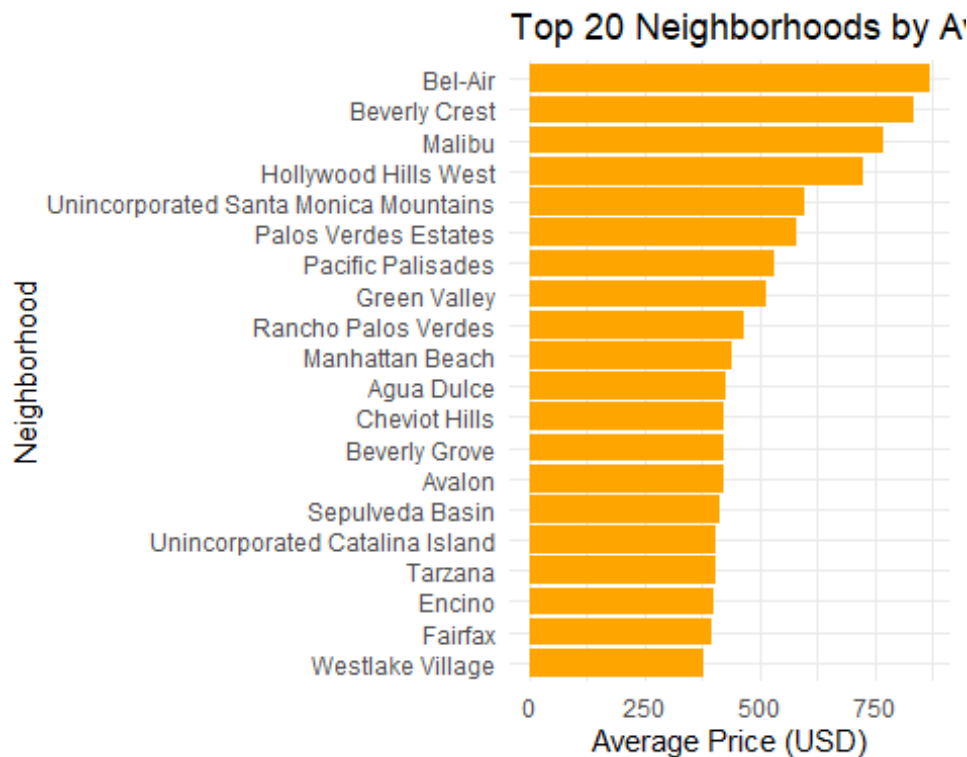
kable(top_neighborhoods,
      caption = "Top 20 Neighborhoods by Average Price",
      digits = 0)

```

Top 20 Neighborhoods by Average Price

neighbourhood_cleansed	n_listings	avg_price	median_price
Bel-Air	44	872	796
Beverly Crest	145	837	713
Malibu	313	769	675
Hollywood Hills West	487	725	650
Unincorporated Santa Monica Mountains	169	597	446
Palos Verdes Estates	16	582	342
Pacific Palisades	154	531	376
Green Valley	4	515	107
Rancho Palos Verdes	80	468	244
Manhattan Beach	341	438	348
Agua Dulce	19	426	150
Cheviot Hills	19	424	259
Beverly Grove	473	422	219
Avalon	208	420	378
Sepulveda Basin	4	414	440
Unincorporated Catalina Island	9	407	268
Tarzana	150	404	250
Encino	197	402	242
Fairfax	197	398	200
Westlake Village	13	379	262

```
# Bar chart of top 20 neighborhoods by average price
ggplot(top_neighborhoods, aes(x=reorder(neighbourhood_cleansed, avg_price),
y=avg_price)) +
  geom_bar(stat="identity", fill="orange") +
  coord_flip() +
  labs(title="Top 20 Neighborhoods by Average Airbnb Price", y="Average Price
(USD)", x="Neighborhood") +
  theme_minimal()
```



Answer:

Neighborhood is definitely a major factor in Airbnb pricing. I looked at average prices by neighborhood, using the same cleaned dataset with listings under \$2,000. Then I grouped by neighbourhood_cleansed and calculated the average and median prices.

The results make a lot of sense. The priciest neighborhoods are Bel-Air and Beverly Crest, averaging \$872 and \$837 per night. Malibu and Hollywood Hills West are also very expensive at \$769 and \$725. Other upscale areas like Pacific Palisades and the Santa Monica Mountains still have high averages, \$530–\$597. Even neighborhoods with fewer listings, like Green Valley and Sepulveda Basin, can have high prices, which

shows that exclusivity matters. Basically, location is one of the strongest drivers of price the nicer the neighborhood, the more you can charge.

Question 4: Does the number of guests (accommodates) affect Airbnb price?

Data: For this analysis, I use the price and accommodates columns from the Airbnb Los Angeles dataset. The goal is to examine whether listings that can host more guests tend to command higher prices.

Methods: I first filtered the dataset to remove listings with missing or zero prices and excluded extreme outliers above \$2,000. The accommodates variable was treated as numeric. To explore the relationship between the number of guests and price, i calculated the average price for each value of accommodates. A line plot was then created to visualize the trend of average price as the number of accommodated guests increases.

Analysis and Results:

```
# Summarize average price by number of guests accommodated
price_by_accommodates <- Airbnb_clean1 %>%
  group_by(accommodates) %>%
  summarise(
    n_listings = n(),
    avg_price = mean(price, na.rm = TRUE),
    median_price = median(price, na.rm = TRUE)
  )

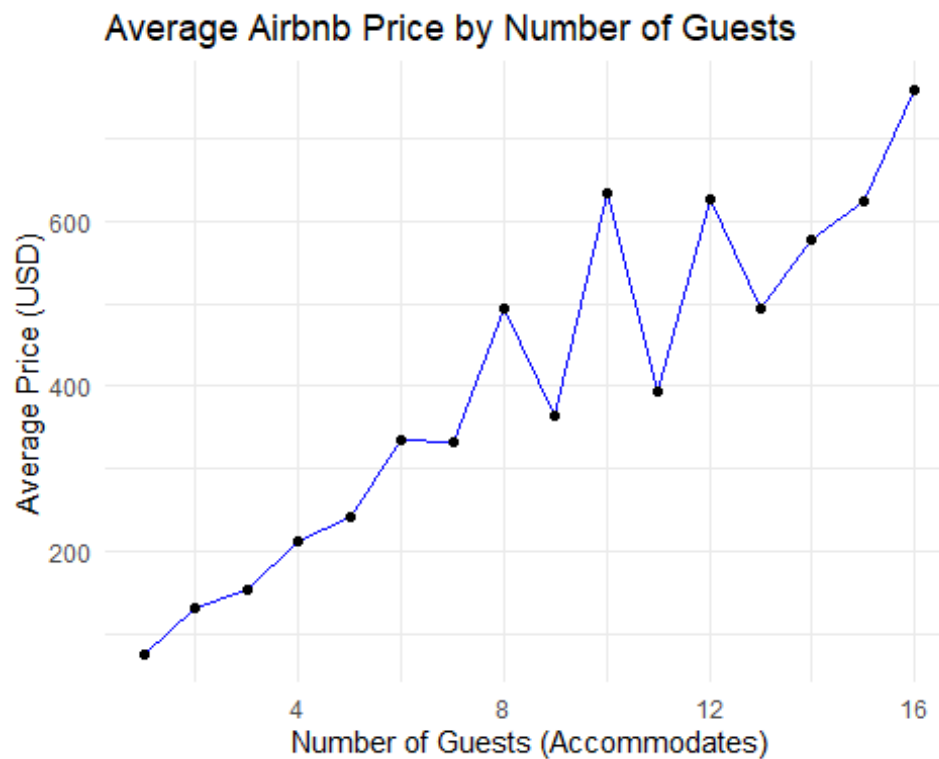
kable(price_by_accommodates,
      caption = "Average Price by Number of Guests Accommodated",
      digits = 0)
```

Average Price by Number of Guests Accommodated

accommodates	n_listings	avg_price	median_price
1	3213	75	60
2	12387	130	109
3	2824	154	133
4	6687	211	174
5	1816	242	199
6	4160	336	255
7	758	333	264

accommodates	n_listings	avg_price	median_price
8	2372	493	350
9	331	364	263
10	1104	634	450
11	121	393	307
12	457	626	436
13	64	493	406
14	184	577	450
15	46	625	470
16	260	758	617

```
# Line plot
ggplot(price_by_accommodates, aes(x = accommodates, y = avg_price)) +
  geom_line(color = "blue") +
  geom_point() +
  labs(title = "Average Airbnb Price by Number of Guests",
       x = "Number of Guests (Accommodates)",
       y = "Average Price (USD)") +
  theme_minimal()
```



Answer:

The analysis shows a clear positive relationship between the number of guests a listing can accommodate and its price. Listings for 1–2 guests have average prices below \$150, while listings accommodating 6–8 guests average \$330–\$490. Large listings for 10–16 guests show even higher average prices, with some exceeding \$750. Median prices follow a similar trend, confirming that higher-capacity listings typically command higher nightly rates.

Conclusion: The number of guests is a significant determinant of Airbnb price. Larger listings, which accommodate more guests, consistently achieve higher average prices.

Question 5: Does being a superhost affect Airbnb price?

Data: For this analysis, we focus on the price and host_is_superhost columns in the Airbnb Los Angeles dataset. The goal is to investigate whether listings hosted by superhosts tend to charge higher prices compared to those hosted by regular hosts.

Method: The dataset was filtered to remove listings with missing or zero prices and extreme outliers above \$2,000. The host_is_superhost column, which contains “t” for true and “f” for false, was converted to a logical variable (TRUE/FALSE). We then summarized average, median, and standard deviation of prices for superhosts versus non-superhosts. A boxplot was also generated to visualize the distribution of prices between the two groups.

Analysis and Results:

```
# Summarize price by superhost status
price_by_superhost <- Airbnb_clean1 %>%
  group_by(host_is_superhost) %>%
  summarise(
    n_listings = n(),
    avg_price = mean(price, na.rm = TRUE),
    median_price = median(price, na.rm = TRUE),
    sd_price = sd(price, na.rm = TRUE)
  )

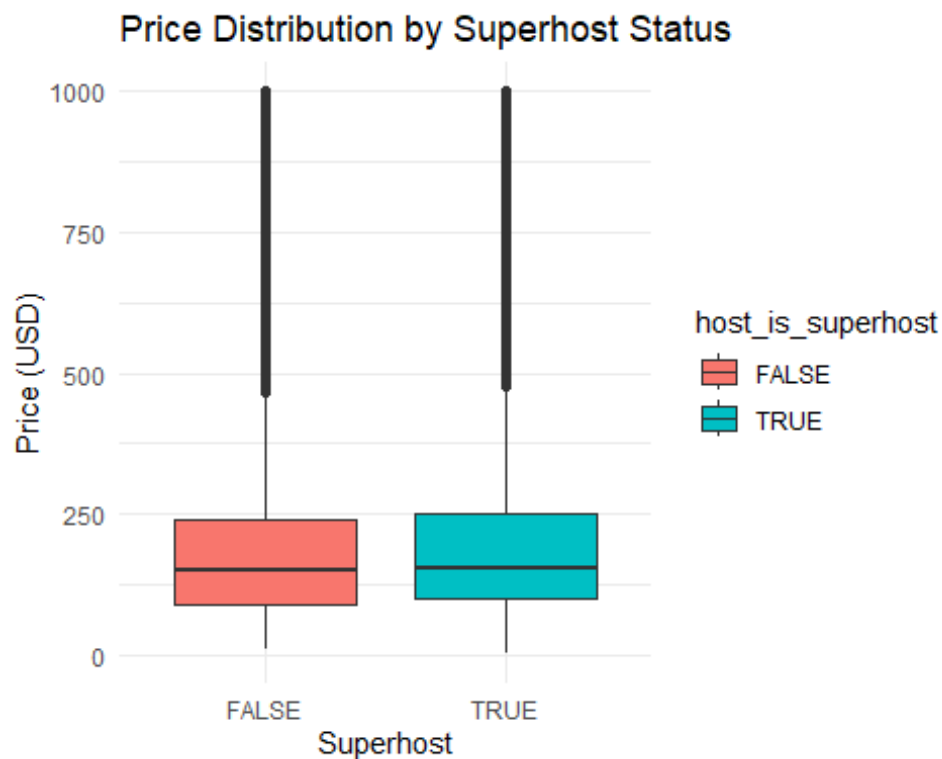
kable(price_by_superhost,
      caption = "Airbnb Prices by Superhost Status",
      digits = 0)
```

Airbnb Prices by Superhost Status

host_is_superhost	n_listings	avg_price	median_price	sd_price
FALSE	21205	228	150	260
TRUE	15579	233	160	244

```
# Boxplot
ggplot(Airbnb_clean1, aes(x = host_is_superhost, y = price, fill =
host_is_superhost)) +
  geom_boxplot() +
  scale_y_continuous(limits = c(0, 1000)) +
  labs(title = "Price Distribution by Superhost Status",
       x = "Superhost",
       y = "Price (USD)") +
  theme_minimal()

## Warning: Removed 911 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```



Answer:

Out of the dataset 15,579 are hosted by superhosts and 21,205 by non-superhosts. Focusing on price, we see that listings by superhosts have a slightly higher average price of \$233.39 compared to \$228.38 for non-superhosts. Median prices follow a similar pattern: \$160 for superhosts versus \$150 for non-superhosts. The standard deviation indicates that both groups have substantial variability, though superhost listings show a slightly lower spread (SD = 243.76) compared to non-superhosts (SD =

259.86), suggesting that while most superhost listings cluster around higher prices, non-superhost listings include more extreme values.

The boxplot of price by superhost status confirms that superhost listings tend to be priced higher overall, with the median and upper quartile elevated relative to non-superhosts. While the difference is modest, it is consistent, indicating that superhost status is positively associated with price.

Conclusion: Listings managed by superhosts command higher prices than those by non-superhosts, demonstrating that trust and reliability perceived by guests can justify a pricing premium. This insight can guide hosts in understanding the value of superhost certification when setting their nightly rates.

Question 6: Do review scores and number of reviews correlate with Airbnb price?

Data: For this analysis, we use three variables from the Airbnb dataset: price – listing price in USD review_scores_rating – average rating of the listing (1–5 scale) number_of_reviews – total reviews received

The goal is to determine whether listings with higher review scores or more reviews tend to command higher prices.

Methods: We first filtered the dataset to remove listings with missing or zero prices, missing review scores, and extreme price outliers above \$2,000. For the analysis, we calculated:

Correlation between review_scores_rating and price to quantify the relationship.
Correlation between number_of_reviews and price.

We also created scatter plots to visualize the relationships, including a log transformation of number of reviews to reduce skew in the plot.

Analysis and Results:

```
# Filter data for valid review scores
reviews_data <- Airbnb_clean1 %>%
  filter(!is.na(review_scores_rating), !is.na(number_of_reviews))

# Correlations
cor_price_rating <- cor(reviews_data$price,
  reviews_data$review_scores_rating, use="complete.obs")
cor_price_reviews <- cor(reviews_data$price, reviews_data$number_of_reviews,
  use="complete.obs")

cor_price_rating
## [1] 0.06011306
```

```

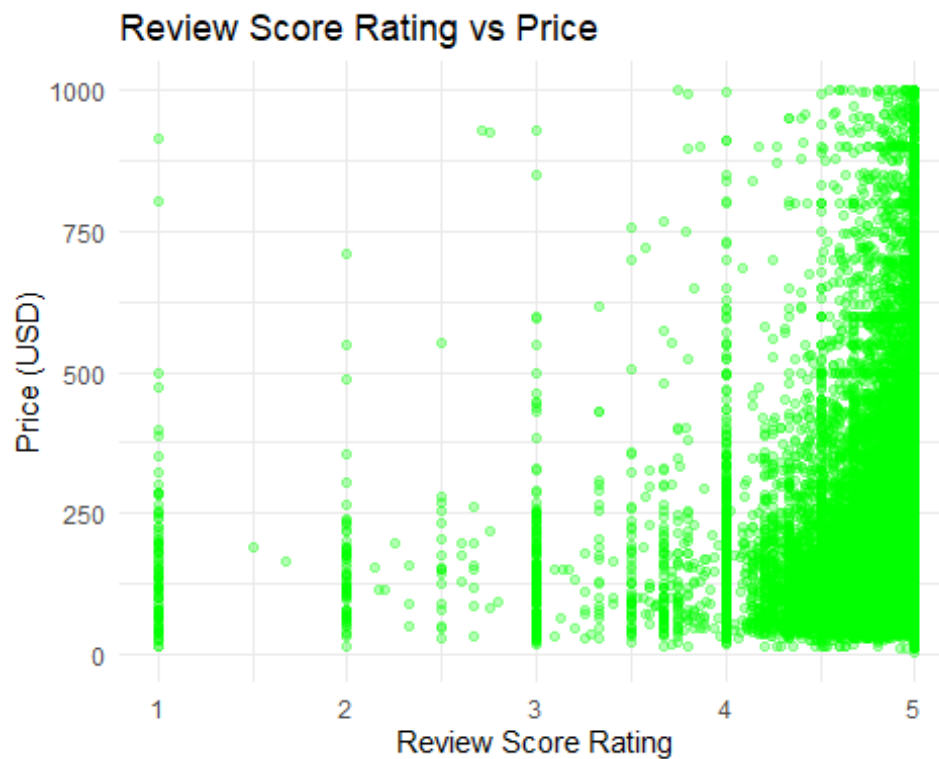
cor_price_reviews

## [1] -0.06727936

# Scatter plot: review score vs price
ggplot(reviews_data, aes(x = review_scores_rating, y = price)) +
  geom_point(alpha = 0.3, color = "green") +
  scale_y_continuous(limits = c(0, 1000)) +
  labs(title = "Review Score Rating vs Price", x = "Review Score Rating", y =
"Price (USD)") +
  theme_minimal()

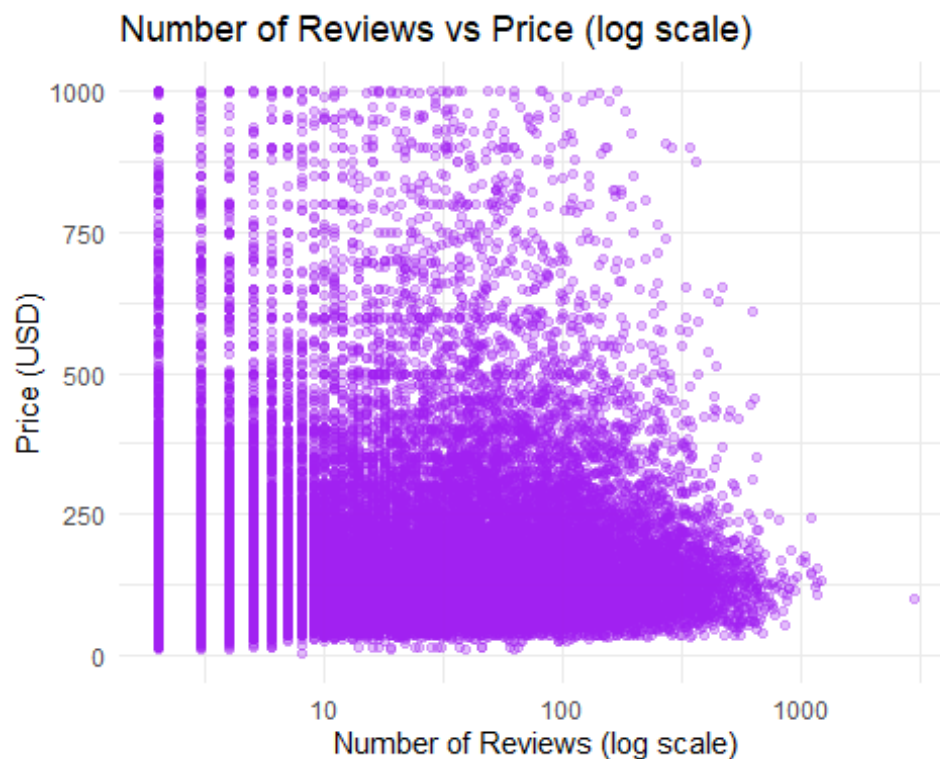
## Warning: Removed 506 rows containing missing values or values outside the
scale range
## (`geom_point()`).

```




```
# Scatter plot: number of reviews vs price (log scale)
ggplot(reviews_data, aes(x = number_of_reviews + 1, y = price)) +
  geom_point(alpha = 0.3, color = "purple") +
  scale_x_log10() +
  scale_y_continuous(limits = c(0, 1000)) +
  labs(title = "Number of Reviews vs Price (log scale)", x = "Number of
Reviews (log scale)", y = "Price (USD)") +
  theme_minimal()

## Warning: Removed 506 rows containing missing values or values outside the
scale range
## (`geom_point()`).
```



Answer:

The correlation between price and review scores is 0.06, indicating a very weak positive relationship. This suggests that listings with slightly higher ratings tend to have marginally higher prices, but the effect is minimal. Conversely, the correlation between price and number of reviews is -0.067, a very weak negative relationship, indicating that listings with more reviews do not necessarily have higher prices and in fact show a slight tendency to have lower prices.

Scatter plots reinforce these findings. Prices are highly variable across all review scores, and the number of reviews spans a wide range with no clear trend toward higher prices. While highly-rated listings can be expensive, many expensive listings have only a few reviews, and popular listings with many reviews may still have moderate pricing.

Conclusion: Review scores and number of reviews have minimal influence on Airbnb price in Los Angeles. While they may reflect quality and popularity, other factors such as room type, property type, neighborhood, and accommodates have much stronger impacts on pricing.

3. Discussion and Analysis

The exploratory analysis of Airbnb listings in Los Angeles reveals several factors that strongly influence rental prices. Across six key questions, we examined the impact of room type, property type, neighborhood, accommodates, superhost status, and review metrics.

Room type is a major determinant of price. Entire homes and apartments consistently command higher rates than private or shared rooms. Hotel rooms, though rare, are the most expensive on average, reflecting premium offerings. This indicates that guests are willing to pay more for privacy and full-unit accommodations.

Property type further explains price differences. Luxury or unique properties, such as villas and guest suites, achieve higher prices compared to standard apartments or houses. While most listings are apartments and houses with moderate pricing, premium properties create significant variability in the market.

Neighborhood location is another critical factor. High-end neighborhoods like Bel-Air, Beverly Crest, Malibu, and Hollywood Hills West have the highest average prices, often exceeding \$700 per night. In contrast, less exclusive neighborhoods tend to have substantially lower prices. This demonstrates that location strongly drives pricing, reflecting both prestige and demand.

The number of guests a listing accommodates also has a substantial impact. Larger listings, designed for six or more guests, consistently command higher prices, with the highest-capacity properties sometimes exceeding \$750 per night. This trend reflects the additional space and amenities required for larger groups.

Superhost status provides a modest premium. Listings managed by superhosts average around \$233 per night, slightly higher than \$228 for non-superhosts. While the difference is small, it suggests that reliability and reputation are valued by guests and can justify a minor price increase.

Finally, **review scores** and number of reviews have minimal impact on price. Correlations between price and review score (0.06) and between price and number of reviews (-0.067) are very weak. Although high review scores or many reviews may reflect quality or popularity, they are not primary drivers of price.

In summary, room type, property type, neighborhood, and accommodates are the strongest factors affecting Airbnb prices in Los Angeles. Superhost status offers a small additional premium, while review metrics have little influence. Hosts and clients seeking to optimize pricing should focus on upgrading property type, targeting premium neighborhoods, and accommodating larger groups to maximize revenue.

4. Conclusions

The analysis of Airbnb listings in Los Angeles demonstrates that several key factors significantly influence rental prices. Room type, property type, neighborhood, and the number of guests accommodated emerge as the strongest drivers of pricing. Entire homes and apartments command higher prices than private or shared rooms, while luxury or unique properties, such as villas or guest suites, achieve premium rates. Location plays a critical role, with high-end neighborhoods like Bel-Air, Beverly Crest, Malibu, and Hollywood Hills West consistently commanding the highest average nightly rates. Similarly, listings designed to accommodate larger groups naturally command higher prices, reflecting the added value of space and amenities.

Superhost status offers a modest pricing advantage, with superhost-managed listings showing slightly higher average and median prices than those managed by non-superhosts. In contrast, review scores and number of reviews have minimal impact on price, indicating that while quality and popularity are important to guests, they do not strongly determine pricing.

Overall, the findings suggest that Airbnb hosts and property managers can maximize revenue by focusing on upgrading their property type, targeting premium neighborhoods, and accommodating larger groups. Superhost status provides a small additional premium, but other factors, particularly location and property features, are far more influential.

These insights provide a clear roadmap for clients looking to understand the drivers of Airbnb pricing and make data-informed decisions about listing strategies and pricing optimization. Future analyses could explore interactions between factors, seasonal price variation, or additional amenities to further refine pricing strategies.

5. Appendix

Summary Tables

1. Room Type Summary

Airbnb Price Summary by Room Type

room_type	n_listings	avg_price	median_price	sd_price
Entire home/apt	28302	270	184	268
Hotel room	111	588	346	485
Private room	7865	95	69	102
Shared room	506	49	30	91

2. Top 20 Neighborhoods by Average Price

Top 20 Neighborhoods by Average Airbnb Price

neighbourhood_cleansed	n_listings	avg_price
Bel-Air	44	872
Beverly Crest	145	837
Malibu	313	769
Hollywood Hills West	487	725
Unincorporated Santa Monica Mountains	169	597
Palos Verdes Estates	16	582
Pacific Palisades	154	531
Green Valley	4	515
Rancho Palos Verdes	80	468
Manhattan Beach	341	438
Agua Dulce	19	426
Cheviot Hills	19	424
Beverly Grove	473	422
Avalon	208	420
Sepulveda Basin	4	414
Unincorporated Catalina Island	9	407
Tarzana	150	404
Encino	197	402

neighbourhood_cleansed	n_listings	avg_price
Fairfax	197	398
Westlake Village	13	379

3. Accommodates vs Price

Airbnb Price by Number of Guests Accommodated

accommodates	n_listings	avg_price	median_price
1	3213	75	60.0
2	12387	130	109.0
3	2824	154	133.0
4	6687	211	174.0
5	1816	242	199.0
6	4160	336	255.0
7	758	333	263.5
8	2372	493	350.0
9	331	364	263.0
10	1104	634	450.5
11	121	393	307.0
12	457	626	436.0
13	64	493	406.5
14	184	577	449.5
15	46	625	470.0
16	260	758	617.0

4. Superhost vs Price

Airbnb Price Summary by Superhost Status

host_is_superhost	n_listings	avg_price	median_price	sd_price
FALSE	21205	228	150	260
TRUE	15579	233	160	244

5. Correlations for Reviews

Correlation Between Price and Review Metrics

Metric	Correlation_with_Price
Review Score Rating	0.060
Number of Reviews	-0.074

5. Key R Commands Used

Load data

```
Airbnb <- read.csv("Airbnb Los Angeles Listings.csv")
```

Clean data (remove NAs and outliers)

```
Airbnb_clean <- Airbnb %>% filter(!is.na(price) & price > 0 & price <= 2000)
```

Summarize room type

```
Airbnb_clean %>% group_by(room_type) %>% summarise(n_listings = n(), avg_price = mean(price), median_price = median(price), sd_price = sd(price))
```

Summarize neighborhood

```
Airbnb_clean %>% group_by(neighbourhood_cleansed) %>% summarise(n_listings = n(), avg_price = mean(price)) %>% arrange(desc(avg_price))
```

Summarize accommodates

```
Airbnb_clean %>% group_by(accommodates) %>% summarise(n_listings = n(), avg_price = mean(price), median_price = median(price))
```

Summarize superhost

```
Airbnb_clean %>% group_by(host_is_superhost) %>% summarise(n_listings = n(), avg_price = mean(price), median_price = median(price), sd_price = sd(price))
```

Correlations

```
cor(Airbnb_cleanprice, Airbnb_cleanreview_scores_rating, use="complete.obs")  
cor(Airbnb_cleanprice, Airbnb_cleannumber_of_reviews, use="complete.obs")
```