



# Airbnb Listings Bangkok Data Analysis

Muhammad Sultan Pasya

# TABLE OF CONTENT

- BACKGROUND
- PROBLEMS TO ADDRESS
- DATA UNDERSTANDING
- GEOSPATIAL COMPETITION ANALYSIS
- SEASONAL TRENDS AND PRICING ANALYSIS
- OPTIMAL MINIMUM NIGHTS POLICY ANALYSIS
- LISTING DESCRIPTION AND KEYWORD ANALYSIS
- CONCLUSION



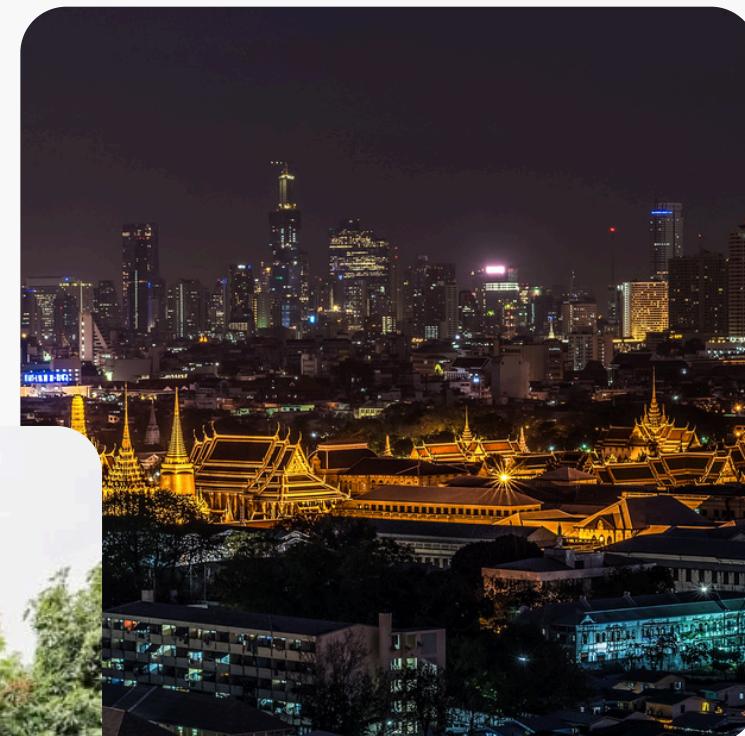
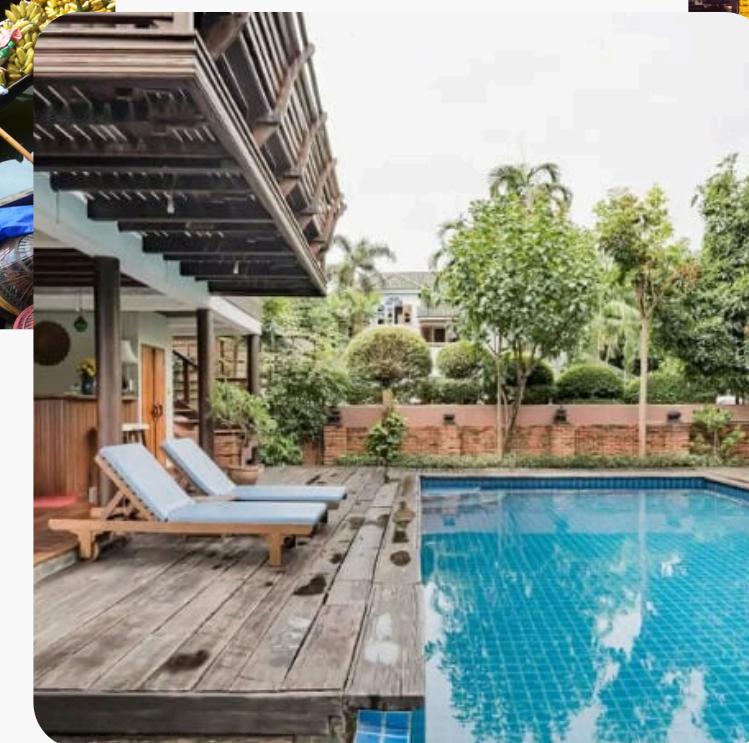
# BACKGROUND

- GROWTH OF BANGKOK'S TOURISM INDUSTRY
- AIRBNB'S ROLE
- CHALLENGES FOR AIRBNB HOSTS
- DATA-DRIVEN ANALYSIS
- ACTIONABLE STRATEGIES

Bangkok, Phuket and Pattaya were among the top 20 most visited cities in the world this year, according to a report by the payment-processing corporation Mastercard.

Thailand's capital topped the ranking with 22.78 million international visitors. Following Bangkok are Paris (19.10 million visitors) and London (19.09 million visitors).

Phuket ranked 14th with 9.89 million visitors, while Pattaya ranked 15th with 9.44 million visitors.





## PROBLEMS TO ADDRESS

- GEOSPATIAL COMPETITION AND DENSITY
- SEASONAL TRENDS AND PRICING
- OPTIMAL MINIMUM NIGHTS POLICY
- LANGUAGE AND KEYWORD IMPACT

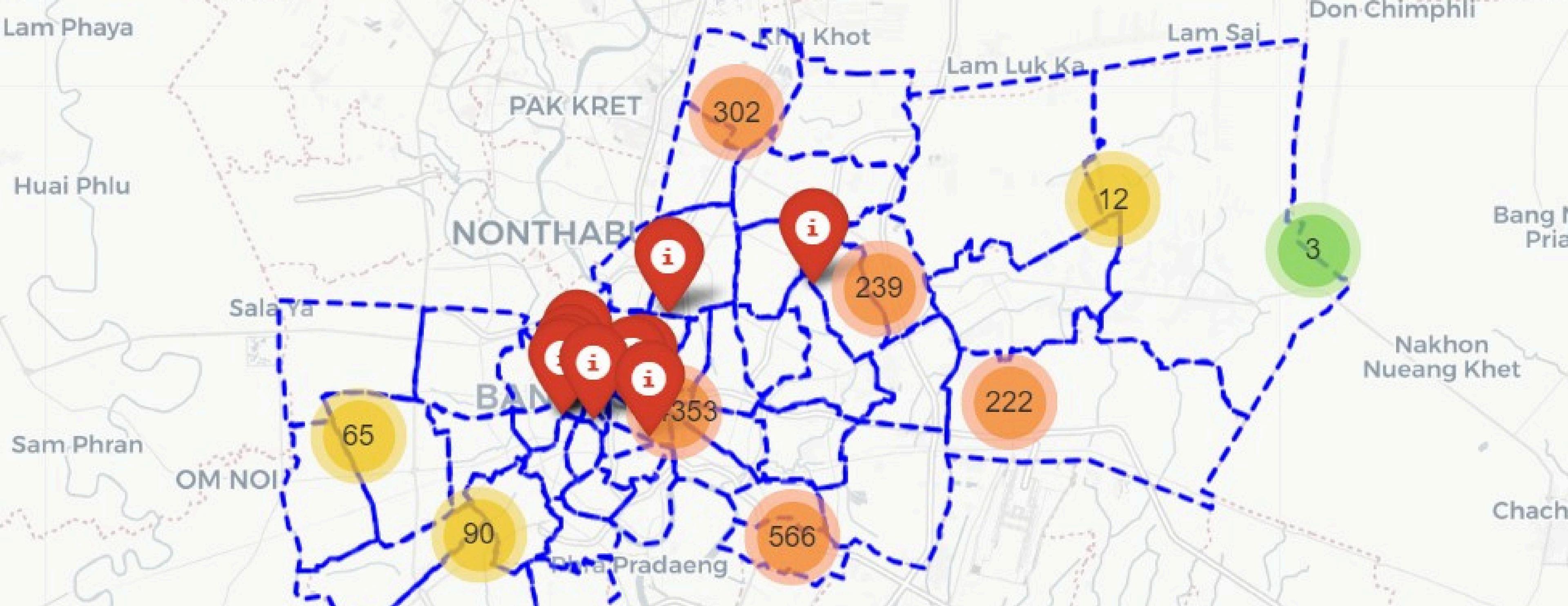
# DATA UNDERSTANDING

## DATA DICTIONARY

Feature	Description
<code>id</code>	Airbnb's unique identifier for the listing.
<code>name</code>	Name of the listing.
<code>host_id</code>	Airbnb's unique identifier for the host/user.
<code>host_name</code>	Name of the host, usually just the first name(s).
<code>neighbourhood</code>	The neighbourhood is geocoded using latitude and longitude.
<code>latitude</code>	Latitude of the listing (WGS84 projection).
<code>longitude</code>	Longitude of the listing (WGS84 projection).
<code>room_type</code>	Type of room: Entire home/apt, Private room, Shared room, or Hotel.
<code>price</code>	Daily price in local currency.
<code>minimum_nights</code>	Minimum number of nights for a stay.
<code>number_of_reviews</code>	Total number of reviews the listing has.
<code>last_review</code>	Date of the last/newest review.
<code>calculated_host_listings_count</code>	Number of listings the host has in the city/region.
<code>availability_365</code>	Number of days the listing is available in the next 365 days.
<code>number_of_reviews_ltm</code>	Number of reviews in the last 12 months.

## DATASET SUMMARY

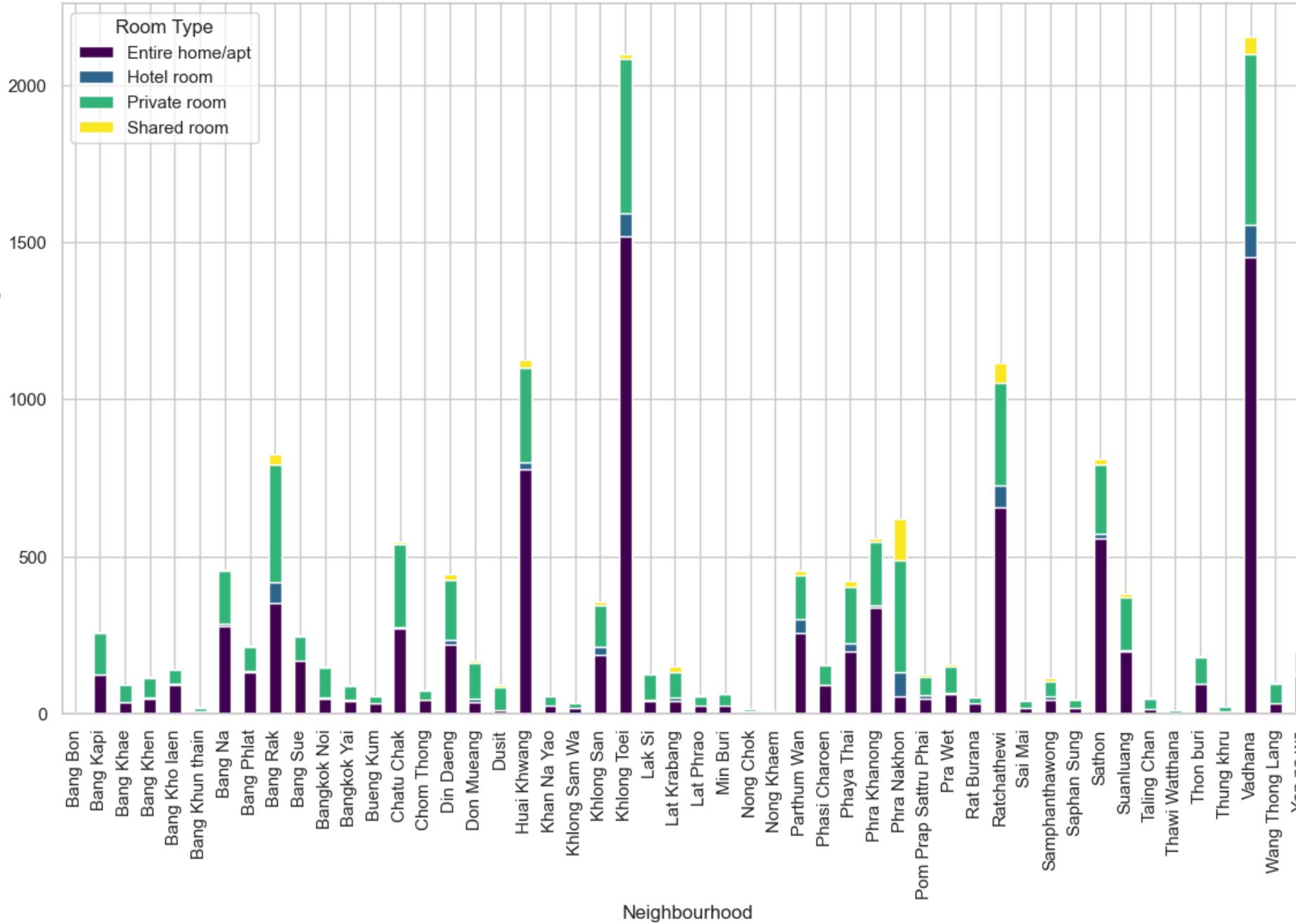
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 15854 entries, 0 to 15853
Data columns (total 17 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Unnamed: 0        15854 non-null   int64  
 1   id               15854 non-null   int64  
 2   name              15846 non-null   object  
 3   host_id            15854 non-null   int64  
 4   host_name          15853 non-null   object  
 5   neighbourhood       15854 non-null   object  
 6   latitude            15854 non-null   float64 
 7   longitude           15854 non-null   float64 
 8   room_type           15854 non-null   object  
 9   price              15854 non-null   int64  
 10  minimum_nights     15854 non-null   int64  
 11  number_of_reviews   15854 non-null   int64  
 12  last_review         10064 non-null   object  
 13  reviews_per_month   10064 non-null   float64 
 14  calculated_host_listings_count  15854 non-null   int64  
 15  availability_365    15854 non-null   int64  
 16  number_of_reviews_ltm  15854 non-null   int64  
dtypes: float64(3), int64(9), object(5)
memory usage: 2.1+ MB
```



# GEOSPATIAL COMPETITION ANALYSIS

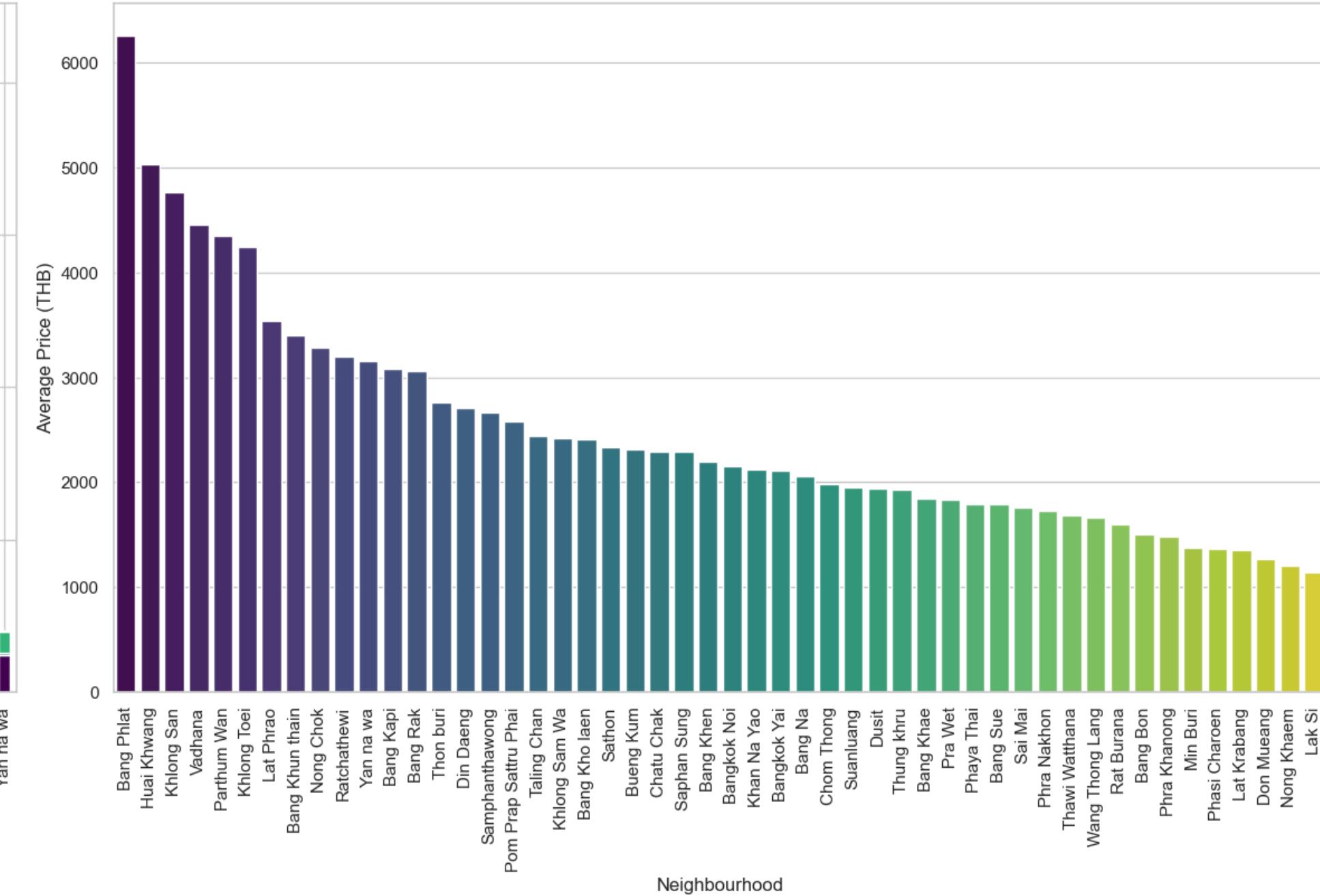
# DISTRIBUTION OF ROOM TYPES AND NEIGHBOURHOOD AVERAGE PRICE

Distribution of Room Types Across Neighbourhoods



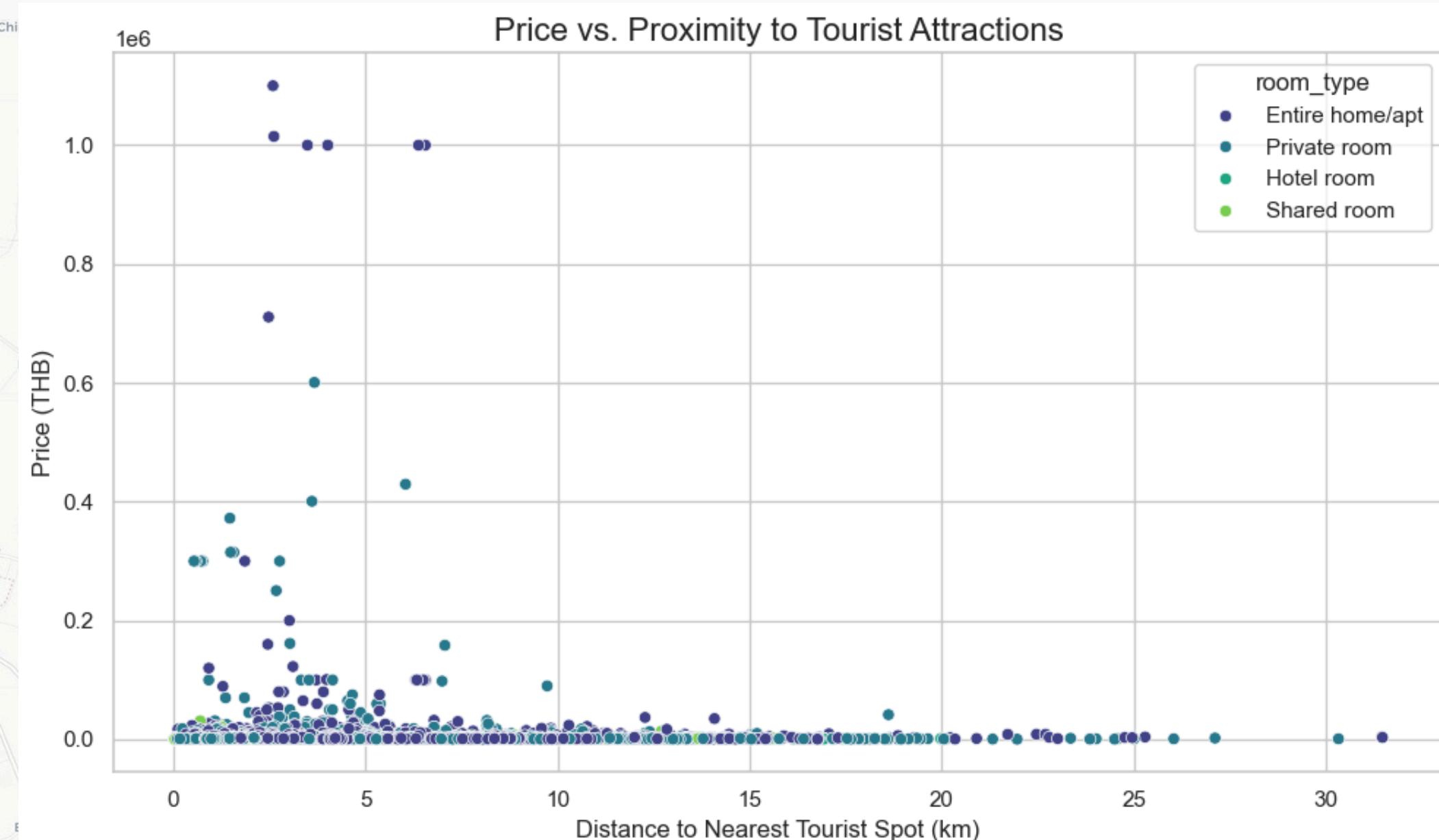
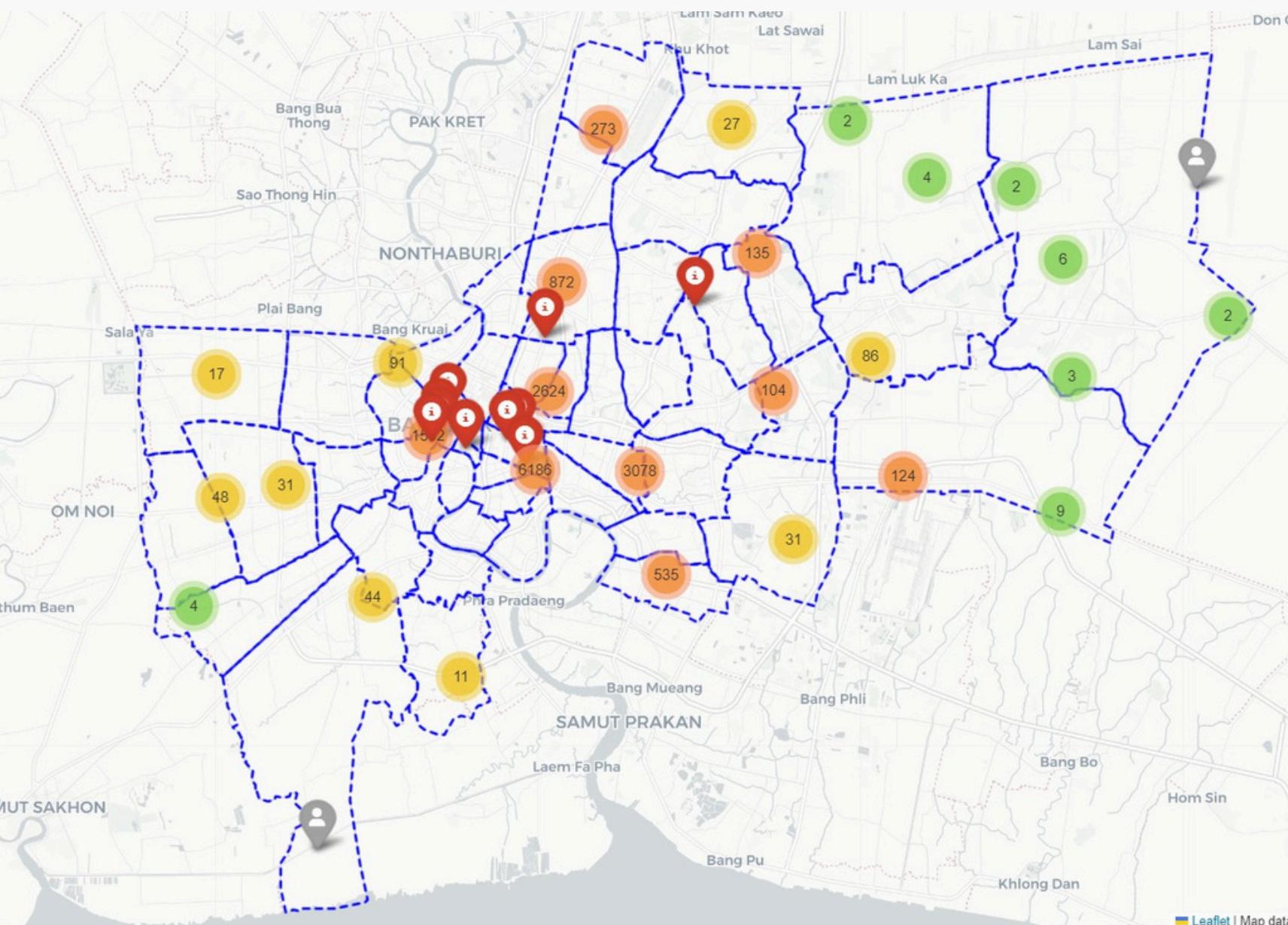
DOMINATED BY ENTIRE HOME/APT

Average Price of Listings by Neighbourhood



HIGHER AVG. PRICE NEIGHBOURHOOD  
USUALLY NEAR MAJOR TOURISM AREA

# DENSITY AND PRICE VS PROXIMITY



**RED ICON: MAJOR TOURISM AREA**

**ORANGE: DENSELY POPULATED**

**YELLOW: MEDIUM DENSITY**

**GREEN: LOW-DENSITY**

**THE CLOSER IT IS TO THE MAJOR TOURISM AREA, THE HIGHER THE RENT PRICE.**

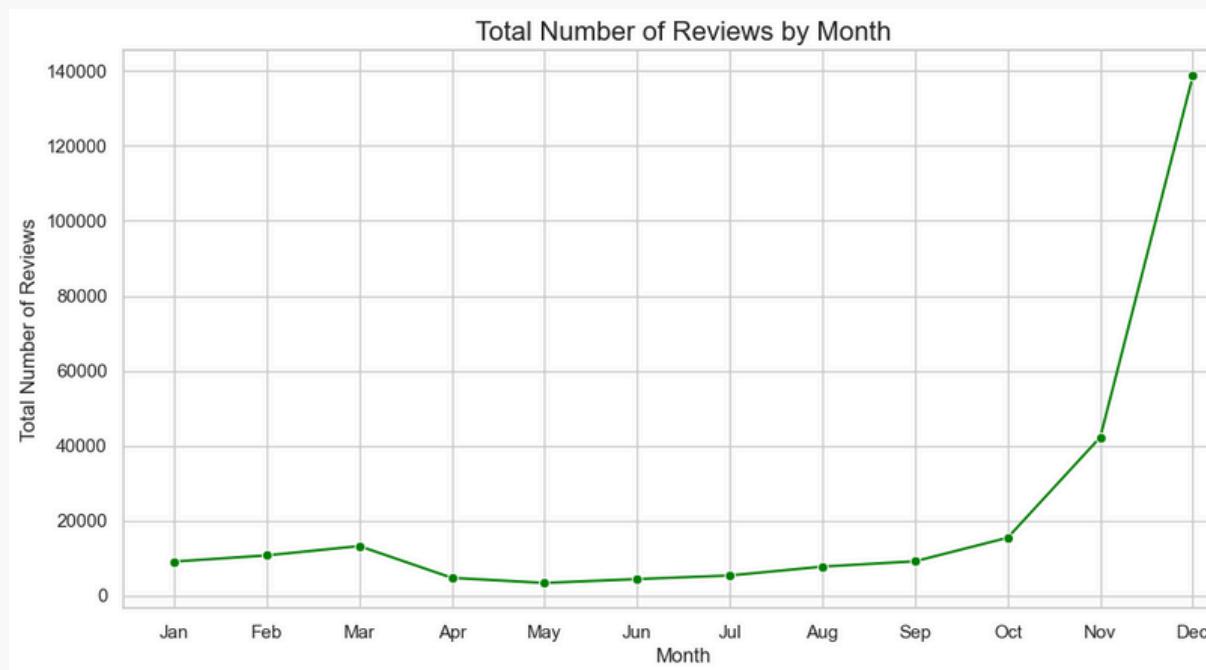
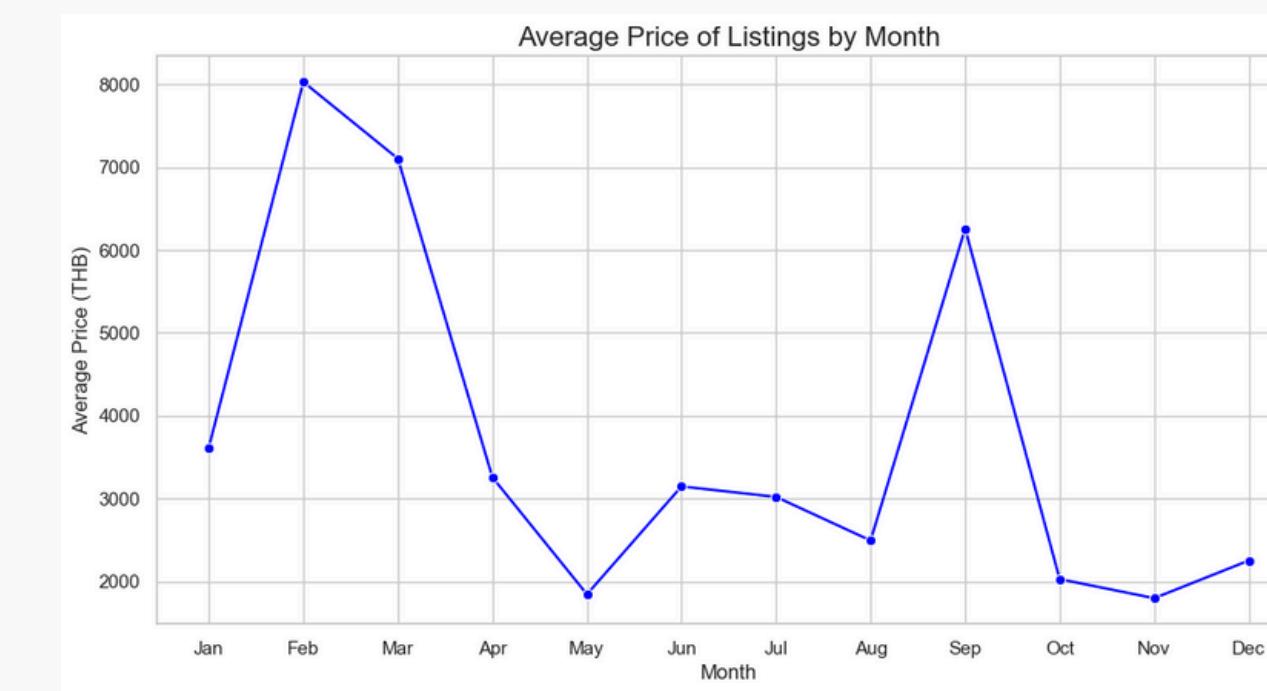
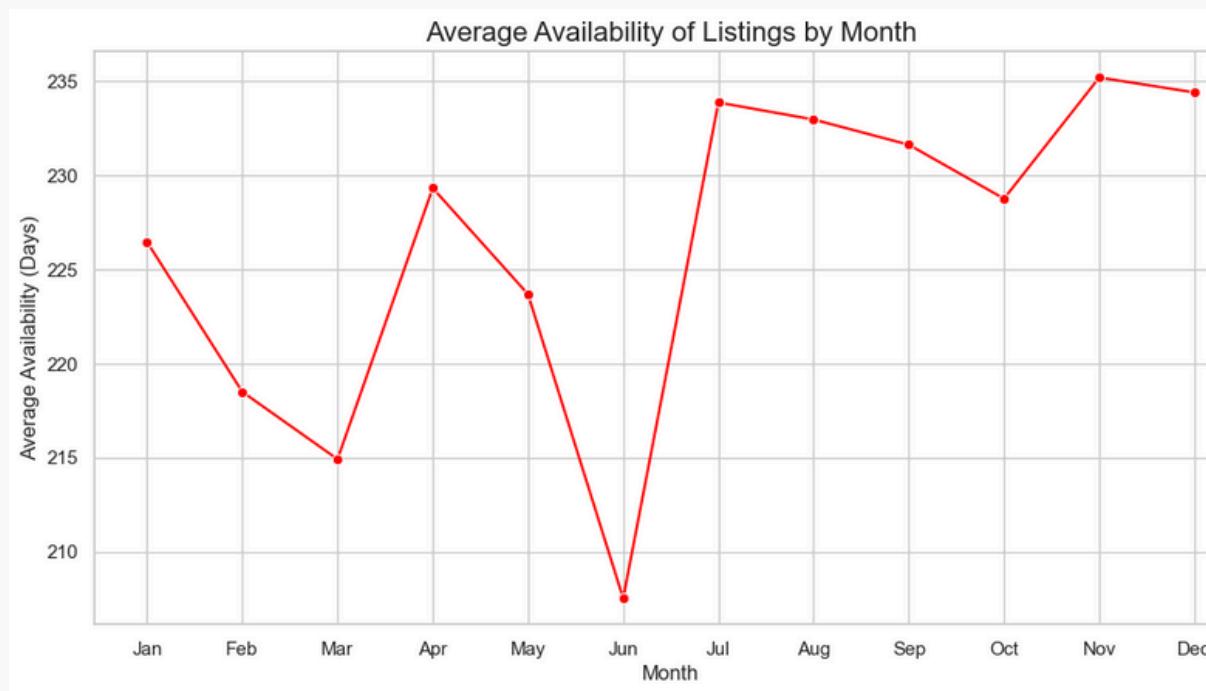
# RECOMENNDATIONS

- DIFFERENTIATION IN COMPETITIVE AREAS
- TAILORED OFFERINGS BY ROOM TYPE
- LEVERAGE PROXIMITY TO ATTRACTIONS
- EXPLORE UNDERREPRESENTED AREAS



# SEASONAL TRENDS AND PRICING ANALYSIS

# AVERAGE AVAILABILITY, AVERAGE PRICE, AND TOTAL REVIEWS BY MONTH



## PRICING FLUCTUATIONS:

- PEAK TOURIST MONTHS DRIVE HIGHER PRICES.
- LOWER PRICES ARE NEEDED IN OFF-PEAK SEASONS TO STAY COMPETITIVE.

## OCCUPANCY & BOOKING ACTIVITY:

- REVIEW COUNTS INDICATE HIGHER OCCUPANCY IN PEAK MONTHS.
- FEWER DISCOUNTS ARE NEEDED DURING HIGH-DEMAND PERIODS; PROMOTIONS ARE ESSENTIAL DURING SLOWER MONTHS.

## AVAILABILITY:

- LISTINGS ARE LESS AVAILABLE DURING HIGH-DEMAND PERIODS.
- STRICTER BOOKING POLICIES DURING PEAK SEASONS; FLEXIBLE TERMS DURING LOW-DEMAND MONTHS.

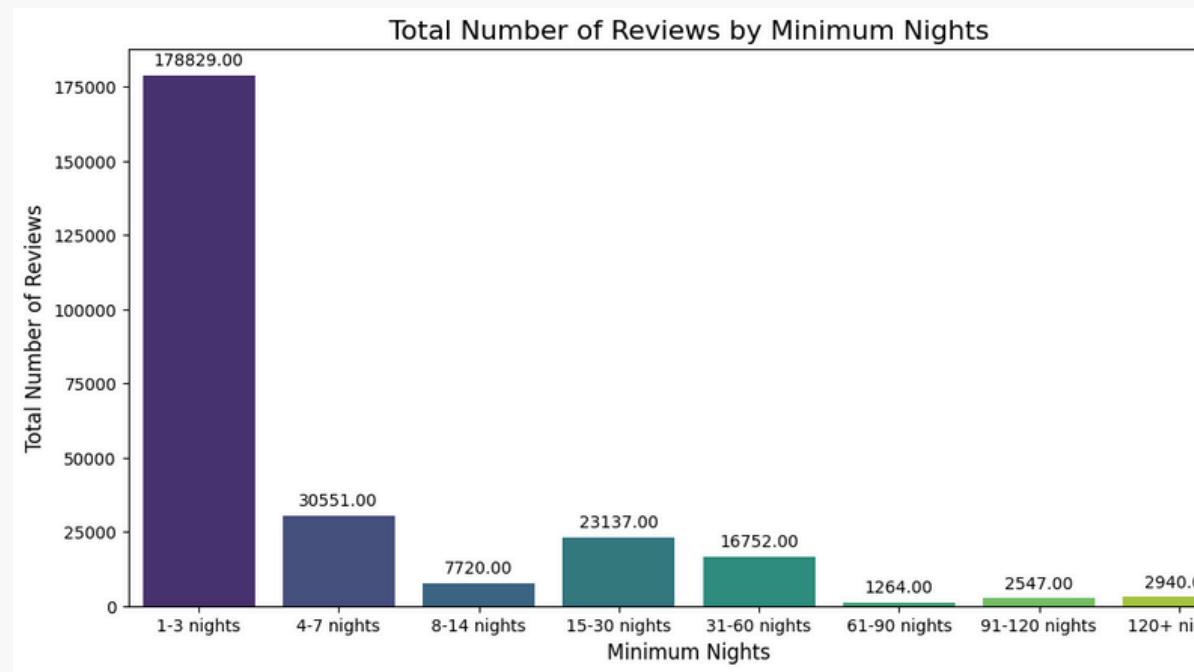
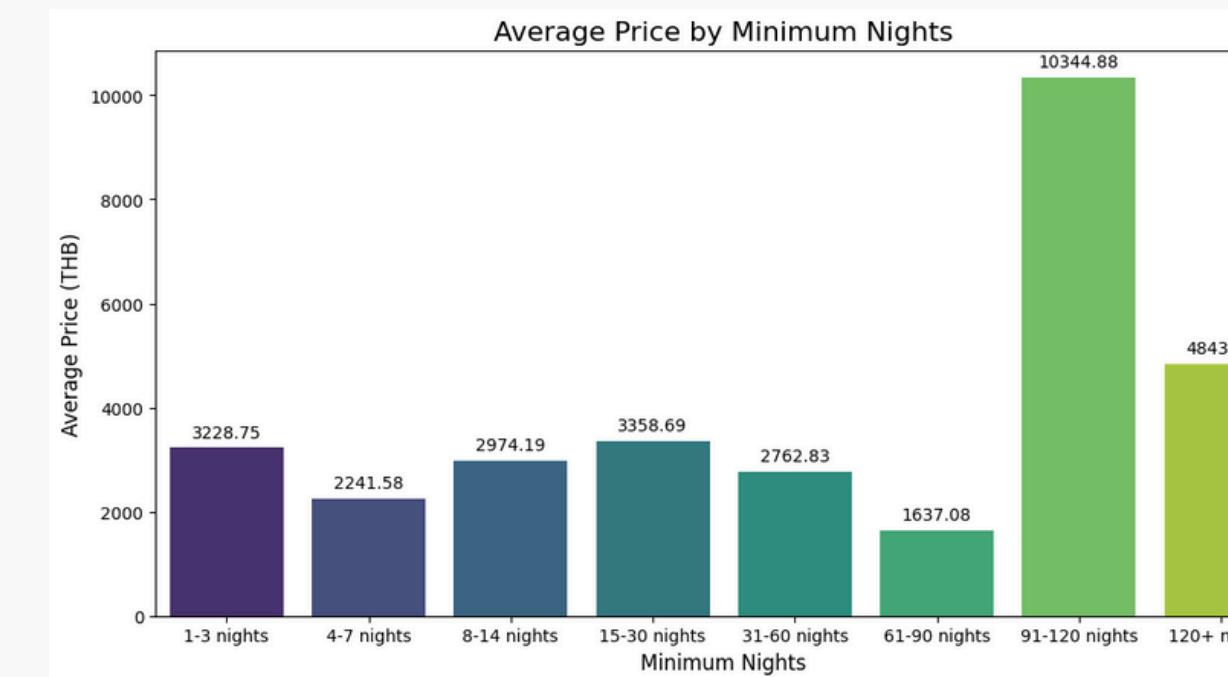
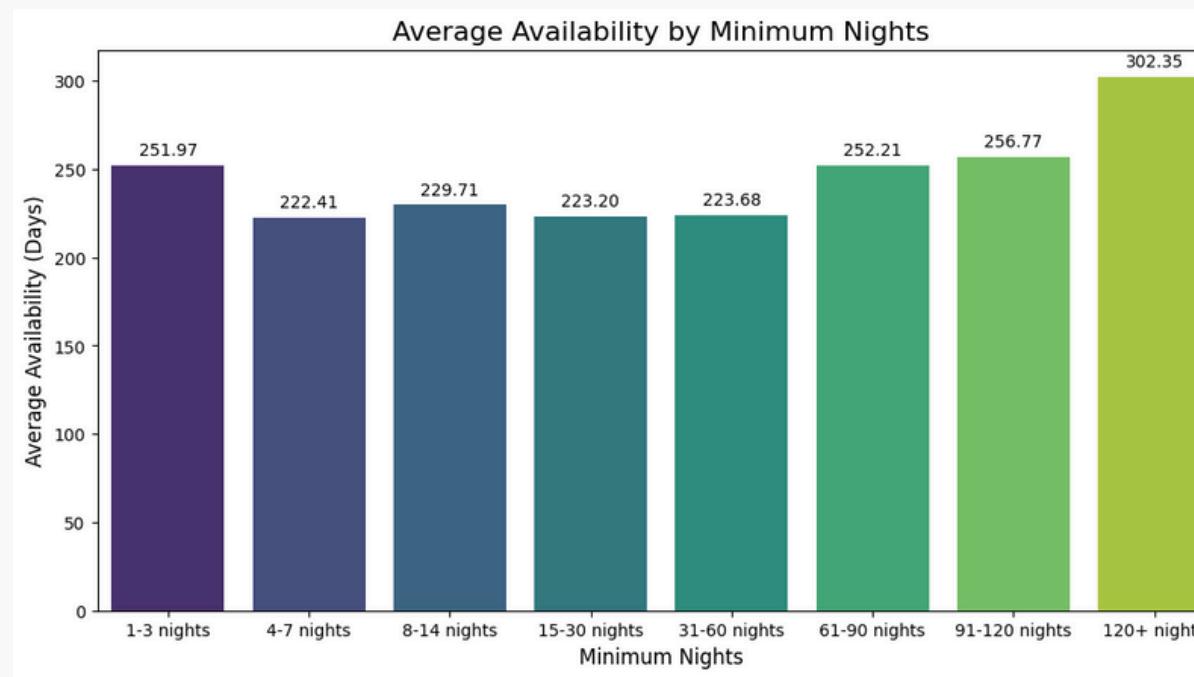
# RECOMENNDATIONS

- DYNAMIC PRICING
- TARGETED MARKETING
- ADJUSTING AVAILABILITY



# OPTIMAL MINIMUM NIGHTS POLICY ANALYSIS

# AVERAGE AVAILABILITY, AVERAGE PRICE, AND TOTAL REVIEWS BY MIN. NIGHTS



## AVAILABILITY:

- LISTINGS WITH 120+ NIGHTS MINIMUM STAY HAVE THE HIGHEST AVAILABILITY, WHILE SHORTER STAYS (1-3 NIGHTS) ARE ALSO FREQUENTLY AVAILABLE BUT LESS BOOKED.
- CONSIDER REDUCING THE MINIMUM NIGHT REQUIREMENT TO INCREASE OCCUPANCY.

## PRICING:

- A SHARP PRICE INCREASE FOR 91-120 NIGHTS STAYS INDICATES THESE ARE PREMIUM OFFERINGS COMPARED TO MODERATELY PRICED SHORT-TERM STAYS.
- REEVALUATE PRICING STRATEGY TO BALANCE PROFITABILITY WITH DEMAND.

## REVIEWS:

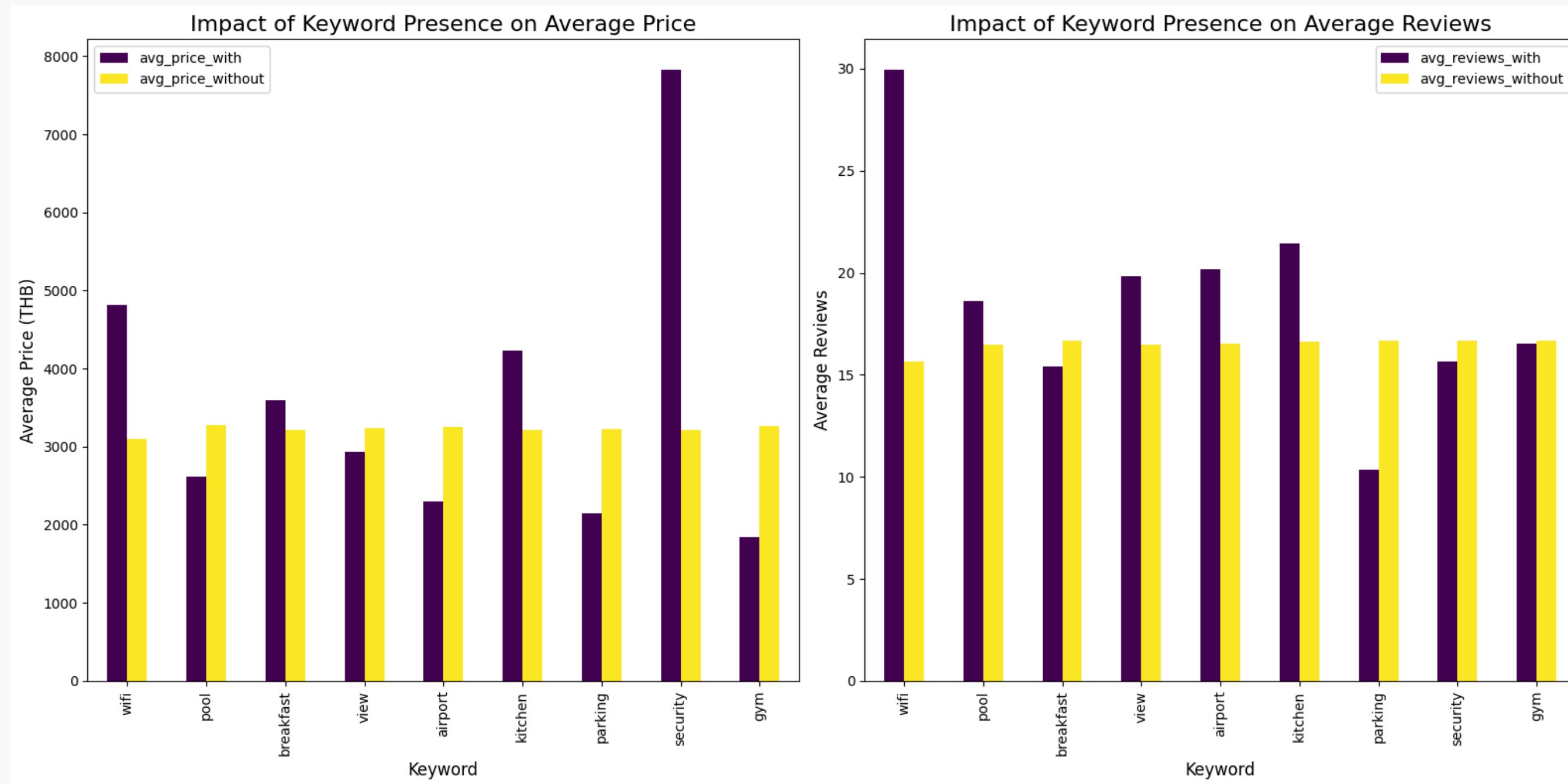
- SHORT-TERM RENTALS (1-3 NIGHTS) RECEIVE THE MOST REVIEWS, REFLECTING HIGHER GUEST TURNOVER AND POPULARITY THAN LONGER STAYS.
- FOCUS MARKETING EFFORTS ON SHORT-TERM STAYS TO CAPITALIZE ON THEIR POPULARITY AND FREQUENT BOOKINGS.

# RECOMENNDATIONS

- BALANCED MINIMUM NIGHTS POLICY
- TARGETING LONG-TERM STAYS
- DYNAMIC POLICY ADJUSTMENTS

gym wifi parking  
bangkok view room oo  
listing description and keyword analysis

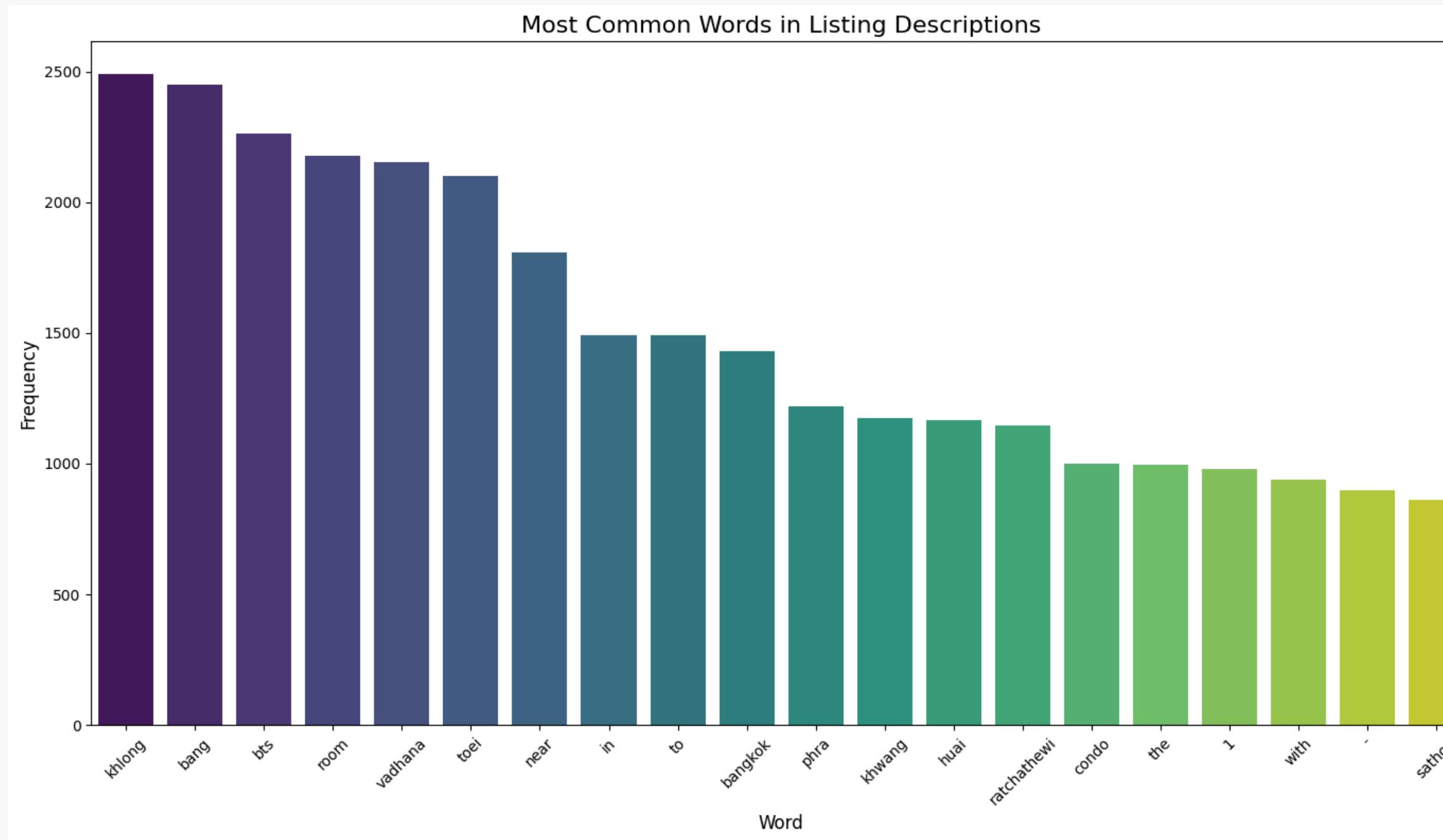
# IMPACT OF KEYWORDS PRESENCE ON AVERAGE PRICE AND AVERAGE REVIEWS



KEYWORDS LIKE "BREAKFAST," "WIFI," "KITCHEN," AND "SECURITY" ARE ASSOCIATED WITH HIGHER AVERAGE PRICES.

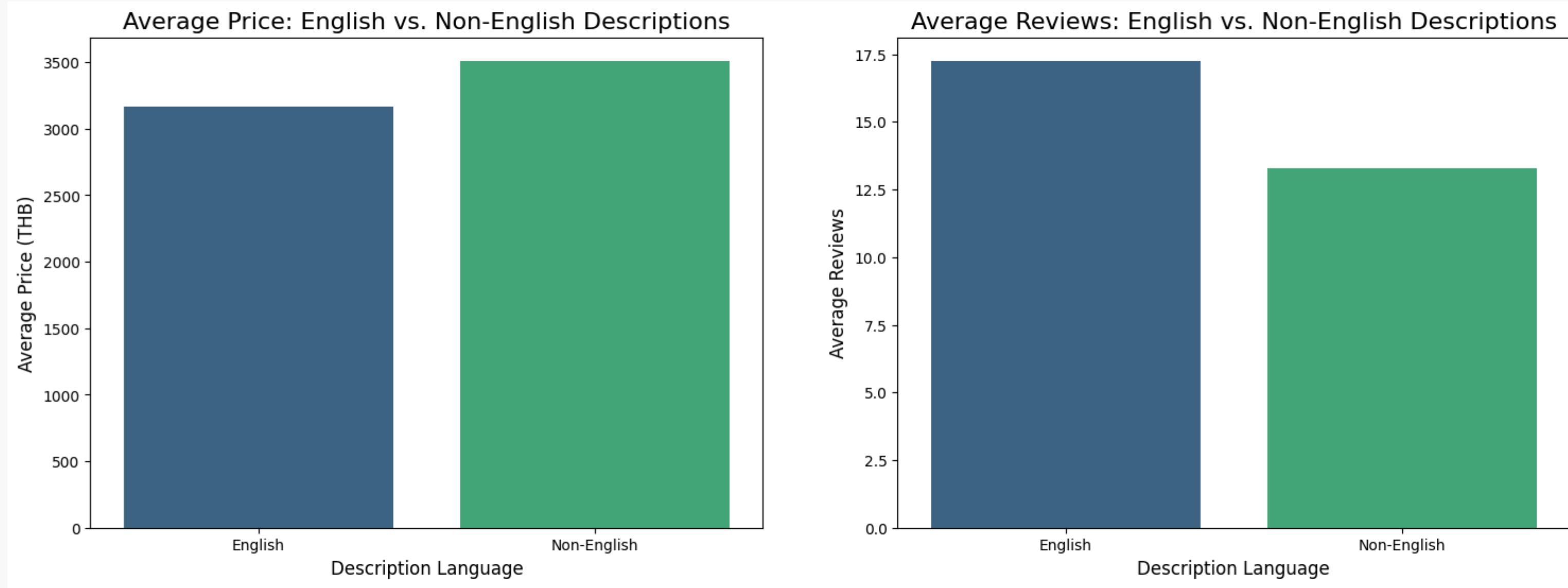
KEYWORDS SUCH AS "WIFI," "KITCHEN," "POOL," "AIRPORT," AND "VIEW" CORRELATE WITH A HIGHER NUMBER OF REVIEWS, INDICATING THEIR IMPORTANCE TO GUESTS.

# IMPACT OF KEYWORDS PRESENCE ON AVERAGE PRICE AND AVERAGE REVIEWS



**COMMON TERMS INCLUDE "ROOM," "NEAR," "BANGKOK," AND "CONDO," INDICATING THAT LOCATION AND PROPERTY TYPE ARE FREQUENTLY EMPHASIZED.**

# IMPACT OF KEYWORDS PRESENCE ON AVERAGE PRICE AND AVERAGE REVIEWS



**NON-ENGLISH DESCRIPTIONS LEAD TO HIGHER AVERAGE PRICES, WHILE ENGLISH DESCRIPTIONS ATTRACT MORE REVIEWS, SUGGESTING BETTER BOOKING RATES DUE TO GREATER VISIBILITY AND GLOBAL UNDERSTANDING.**

# RECOMENNDATIONS

- **EMPHASIZE HIGH-VALUE FEATURES**
- **FOCUS ON ESSENTIAL AMENITIES**
- **USE CLEAR AND DESCRIPTIVE LANGUAGE**
- **CONSIDER ENGLISH DESCRIPTION**

# CONCLUSION

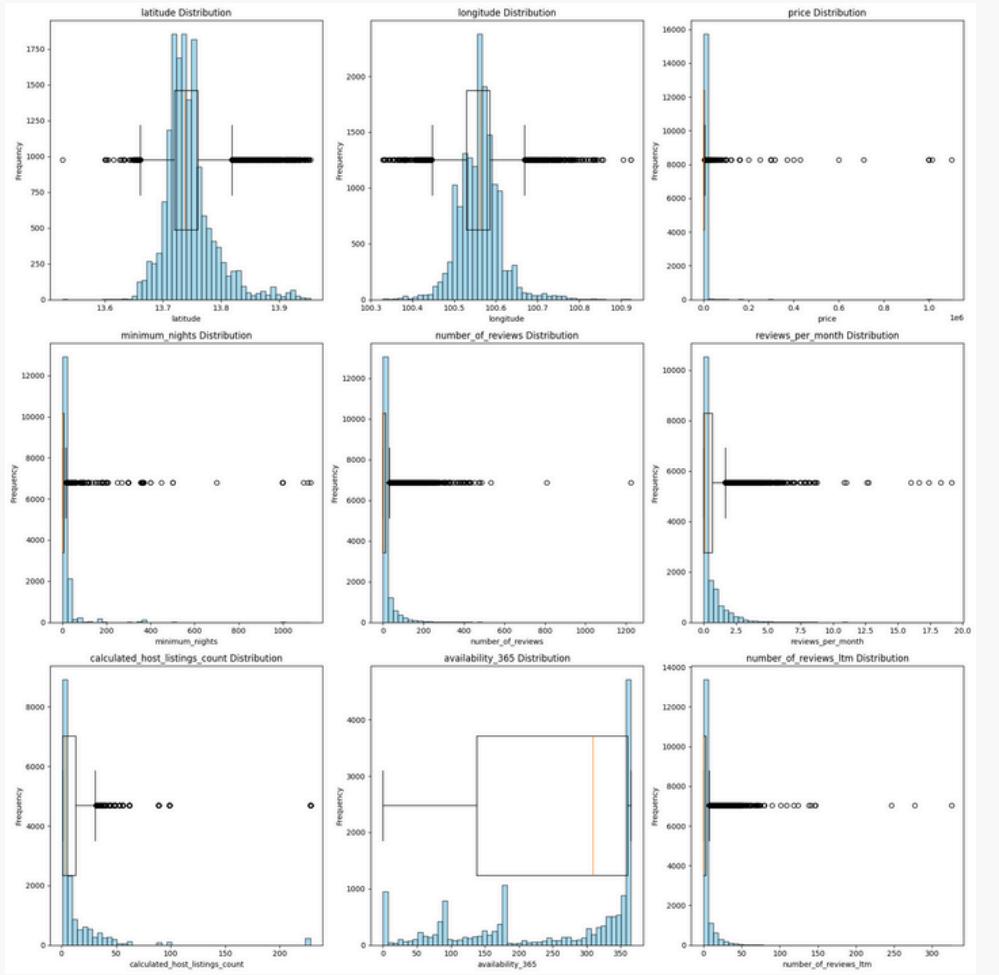
- **DIFFERENTIATE IN COMPETITIVE AREAS:** FOCUS ON UNIQUE FEATURES AND COMPETITIVE PRICING TO STAND OUT IN HIGH-DENSITY NEIGHBOURHOODS. NEW LISTINGS SHOULD CONSIDER LESS SATURATED AREAS FOR EASIER MARKET ENTRY.
- **LEVERAGE SEASONAL TRENDS:** IMPLEMENT DYNAMIC PRICING ALIGNED WITH TOURIST SEASONS TO MAXIMIZE REVENUE AND MAINTAIN OCCUPANCY YEAR-ROUND.
- **OPTIMIZE MINIMUM NIGHT POLICIES:** FAVOR SHORTER MINIMUM STAYS (1-3 NIGHTS) TO BOOST OCCUPANCY AND REVIEWS, ADJUSTING DYNAMICALLY BASED ON SEASONAL DEMAND.
- **ENHANCE DESCRIPTIONS:** USE STRATEGIC KEYWORDS AND THE ENGLISH LANGUAGE TO IMPROVE VISIBILITY, ATTRACT BOOKINGS, AND COMMAND HIGHER PRICES.

**THANK YOU**

# APPENDICES

# DATA DISTRIBUTION

	Shapiro-Wilk Statistic	Shapiro-Wilk p-value	D'Agostino Statistic	D'Agostino p-value	Skewness	Skew Direction	Normally Distributed
latitude	0.90341	0.0	4356.959559	0.0	1.401384	Positive	False
longitude	0.95221	0.0	2077.433462	0.0	0.559228	Positive	False
price	0.044418	0.0	43046.717756	0.0	33.882198	Positive	False
minimum_nights	0.279787	0.0	22190.129178	0.0	8.229398	Positive	False
number_of_reviews	0.441057	0.0	19163.17247	0.0	6.261361	Positive	False
reviews_per_month	0.579128	0.0	15621.035696	0.0	4.628295	Positive	False
calculated_host_listings_count	0.417437	0.0	16474.501173	0.0	5.386583	Positive	False
availability_365	0.832341	0.0	12885.350519	0.0	-0.576219	Negative	False
number_of_reviews_ltm	0.411577	0.0	24895.698981	0.0	9.725723	Positive	False



# OUTLIERS

	Column	Total Outliers	Percentage (%)	Lower Bound	Upper Bound
0	latitude	1094	6.900902	13.660975	13.818615
1	longitude	602	3.797389	100.446500	100.668340
2	price	1403	8.850060	-1393.500000	4722.500000
3	minimum_nights	3168	19.983599	-8.000000	16.000000
4	number_of_reviews	2240	14.129818	-19.500000	32.500000
5	reviews_per_month	1471	9.279001	-1.005000	1.675000
6	calculated_host_listings_count	1832	11.556172	-17.000000	31.000000
7	availability_365	0	0.000000	-195.000000	693.000000
8	number_of_reviews_ltm	2219	13.997351	-4.500000	7.500000

