

Geographical Visualizations with Python

01 INTRO

02 Research Questions

03 Data Profiling

04 Data
Cleaning

06
Preliminary
Insights

05 Data
Profile &
Ethical
Reflection

07
Vision &
Conclusion

INTRO

Exploring Airbnb Amsterdam Listings & Booking Behavior

CONTEXT

Snapshot of Airbnb listings in Amsterdam on December 6th, 2018.

~20,000 listings with detailed information, calendar data for 365 days, reviews, and neighborhood shapefile.

Variables include price, room type, availability, reviews, and location.

OBJECTIVE

Understand factors affecting booking rates and pricing strategies.

Explore trends in guest behavior and listing performance.

MOTIVATION

Helps hosts optimize listings for occupancy and revenue.

Provides insights into market trends for Amsterdam Airbnb listings.

RESEARCH QUESTIONS

What We Want to Know

01

Are there differences in booking rates across room types and neighborhoods?

02

How do price, availability, and reviews affect the likelihood of bookings?

03

What recommendations can hosts or Airbnb implement to improve occupancy?

04

Are there seasonal trends in availability or pricing?

Data Profiling

& Where the Data Comes From

Source

[Kaggle datasets/erikbruin/airbnb-amsterdam](#)

7 Datasets Files

listings.csv → all listings in Amsterdam (~20k).

listings_details.csv → extra variables like amenities, property type, and host info.

calendar.csv → daily availability and prices for 365 days.

reviews.csv → guest reviews (text mining possible).

Neighborhood shapefile → spatial analysis.

Basic Stats Example:

Number of listings: ~20,000

Key attributes: price, room type, availability, number of reviews, review scores

Missing data overview: some listings may not have review scores or full amenities

Data Cleaning

HOW I CLEANED THE DATA

- 01 Understand Each File
- 02 Decide What to Keep
- 03 Cleaning Each File

Check for duplicates (especially in listings, calendar, reviews).

Check missing values and decide how to handle them (drop, fill, or mark as unknown).

Standardize columns (lowercase, no spaces, correct data types).

Convert prices in calendar to numeric (remove \$/€, commas).

Parse dates for calendar and reviews.

04 Merge / Aggregate
 Merge listings + listings_details → main listing table.
 Merge neighbourhoods → add neighborhood info.
 Aggregate calendar → calculate average price, availability rate, maybe seasonal trends.
 Aggregate reviews → calculate average review score, total reviews per listing.

Loaded and merged 1.2M rows from multiple files

Cleaned missing values and standardized formats (prices, dates, availability)

Final shape: 1,200,071 rows × 21 columns

Tools: pandas, numpy, geopandas

```
# Step 1: Import libraries
import pandas as pd
import numpy as np
# GeoPandas is optional if you want to work with neighborhoods
# import geopandas as gpd

# Step 2: Load CSV and GeoJSON files

# Corrected file paths (no extra spaces)
calendar = pd.read_csv('calendar.csv')

# Step 3: Quick check of the datasets
print("Calendar shape:", calendar.shape)

#drop columns
columns_to_drop = ['scrape_id', 'listing_url', 'last_scraped', 'experiences_offered',
listings_clean = listings.drop(columns=columns_to_drop)

#Handle missing values

listings_clean['name'] = listings_clean['name'].fillna('')
listings_clean['summary'] = listings_clean['summary'].fillna('')
listings_clean['space'] = listings_clean['space'].fillna('')
listings_clean['description'] = listings_clean['description'].fillna('')
listings_clean['reviews_per_month'] = listings_clean['reviews_per_month'].fillna(0)

#Convert boolean-like columns
bool_columns = ['instant_bookable', 'is_business_travel_ready', 'require_guest_profile_picture',
'require_guest_phone_verification', 'requires_license']

for col in bool_columns:
    listings_clean[col] = listings_clean[col].map({'t': True, 'f': False})
```

```
import pandas as pd
import numpy as np

# Load calendar (assuming it's already in the notebook)
calendar = pd.read_csv('calendar.csv')

# Convert 'date' to datetime
calendar['date'] = pd.to_datetime(calendar['date'], errors='coerce')

# Convert 'available' to boolean
calendar['available'] = calendar['available'].map({'t': True, 'f': False})

# Clean 'price' column
# Remove $ signs, commas, and convert to float
calendar['price'] = calendar['price'].replace('[\$,]', '', regex=True)
calendar['price'] = pd.to_numeric(calendar['price'], errors='coerce')

# Check for missing values
print(calendar.isna().sum())

# Quick check of cleaned data
print(calendar.head())
```

Data Profile & Ethical Reflection

DATASET OVERVIEW

Overview

Room types: Entire home/apartment, Private room, Shared room.

Average price per night, distribution of bookings, neighborhood differences.

Number of reviews per listing; availability trends.

Limitations

Only Amsterdam listings as of one date → not generalizable to other cities or years.

Seasonal and long-term booking trends are limited.

Missing demographic info about guests.

Ethical Considerations

Protect host and guest privacy.
Avoid overgeneralization based on single-date snapshot.

Understanding what the data tells — and what it doesn't.

Preliminary Insights

Preliminary Insights

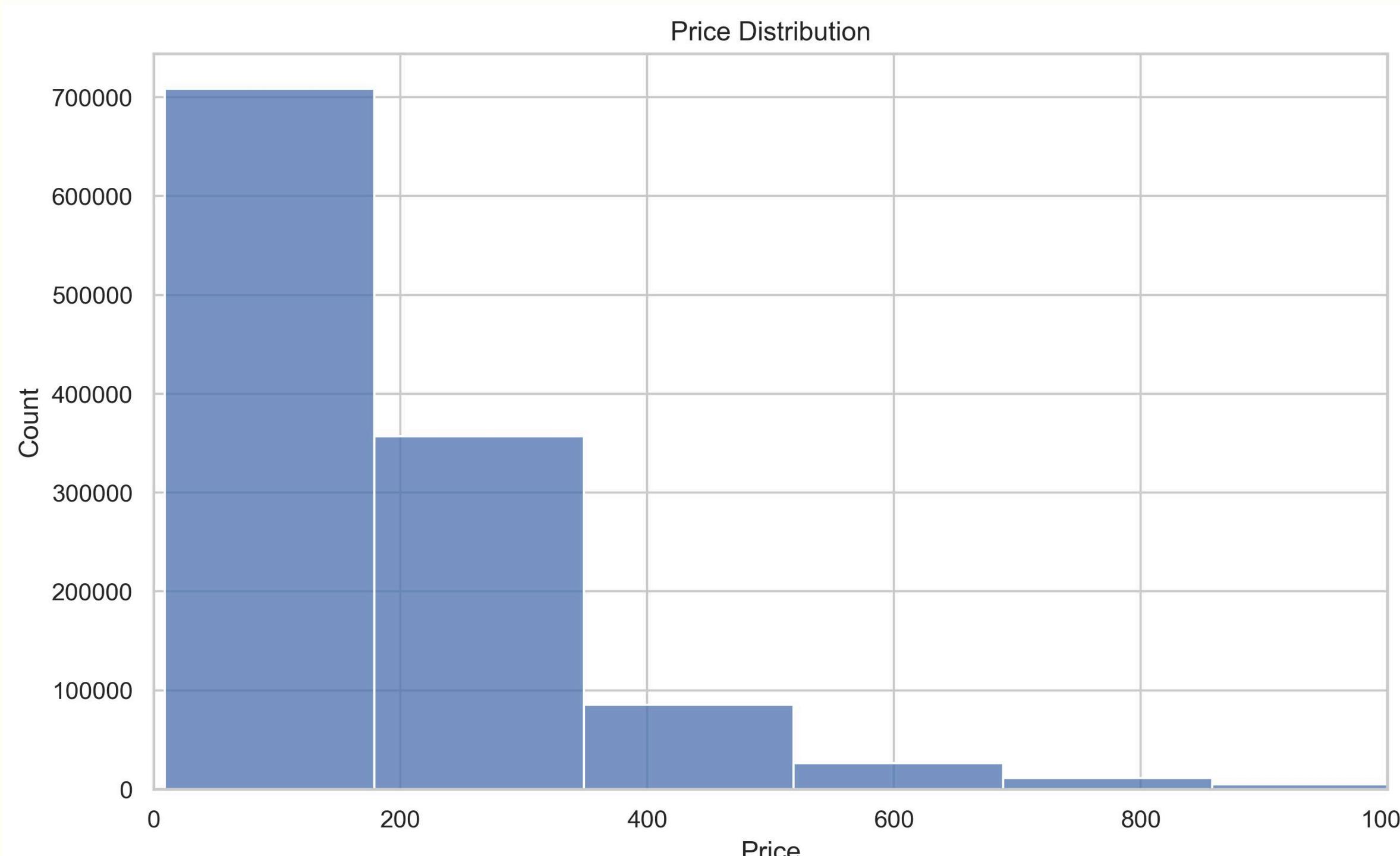
Price distributions across room types
Entire homes are pricier, shared rooms cheaper.

Booking likelihood higher for well-reviewed and competitively priced listings.

Popular neighborhoods (e.g., city center) have higher occupancy rates.

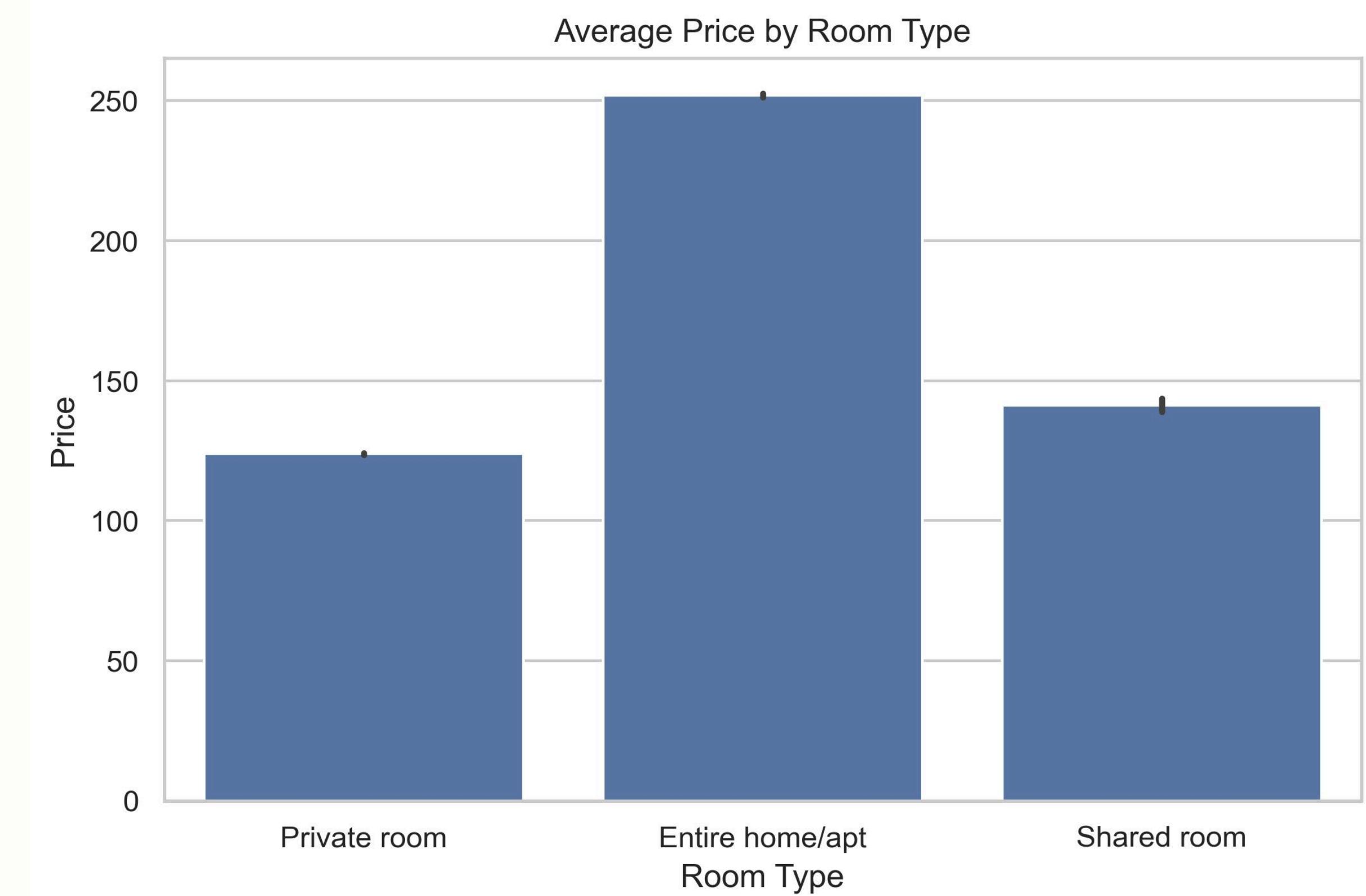
Some listings are available year-round, others are seasonal.

Insights



PRICE DISTRIBUTION (HISTOGRAM)

Most listings priced between €50–€250
Outliers above €500 are rare, mostly in premium neighborhoods



AVERAGE PRICE BY ROOM TYPE (BARPLOT)

Entire homes have the highest average price
Private rooms are the most common listing type

Exploratory Analysis



HEATMAP

Cleanliness and overall review scores rise together, while most other variables show little correlation.

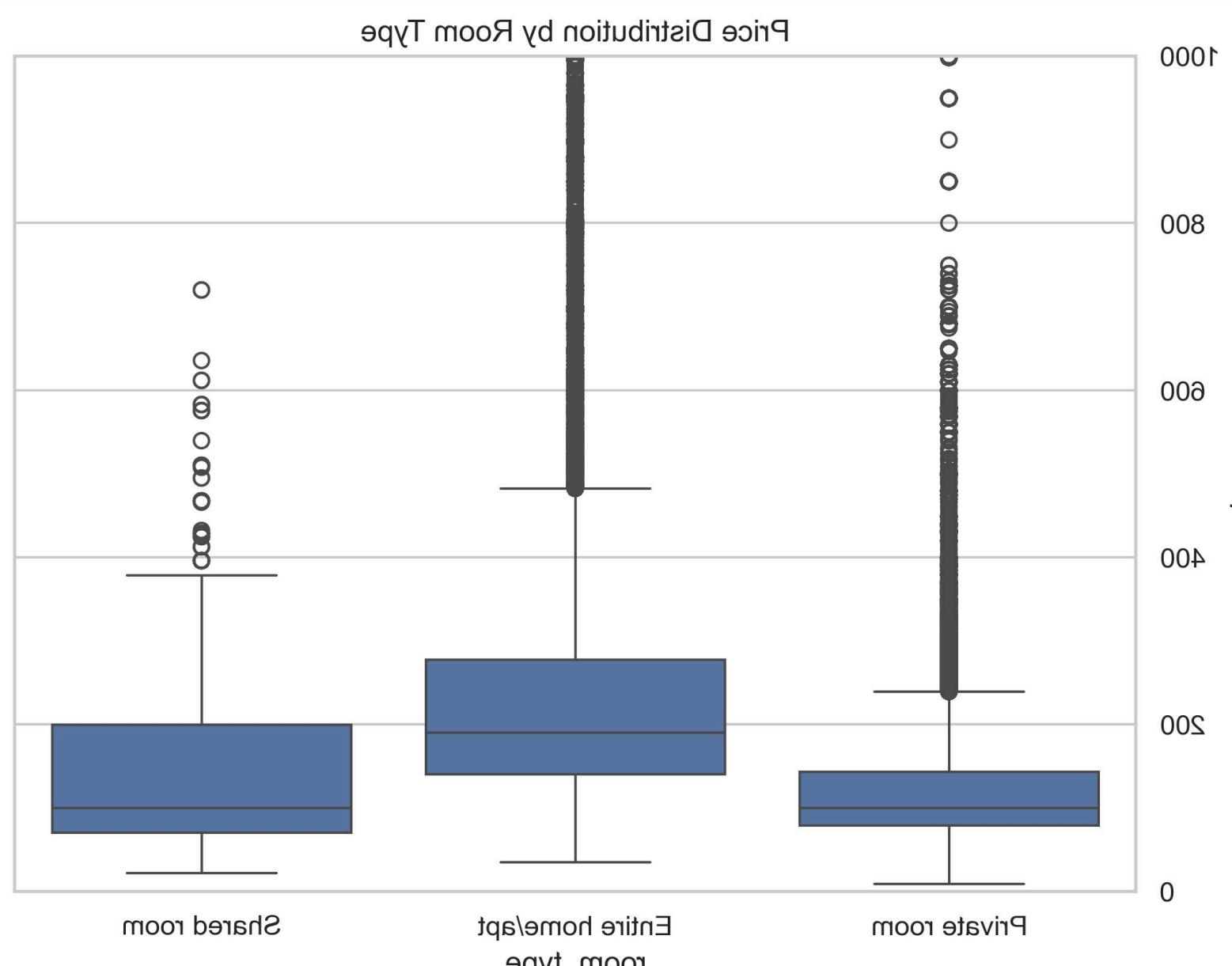
Positive = both increase together; negative = one rises as the other falls.

AVERAGE PRICE BY ROOM TYPE (BARPLOT)

Scatterplot Insights:

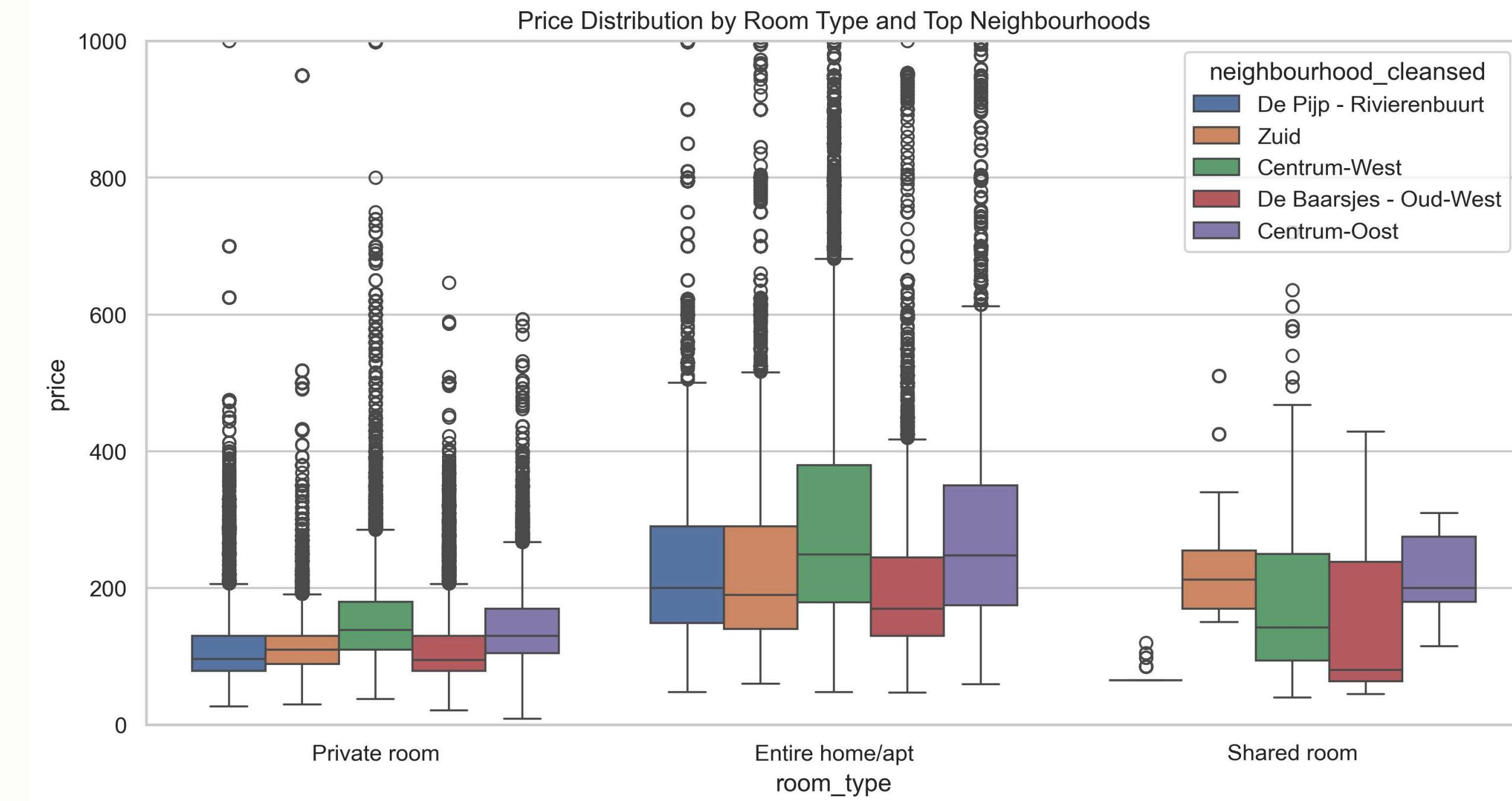
Larger apartments—more bedrooms or higher accommodates—generally cost more. Outliers exist (e.g., 1-bedroom units above €8,000), showing that other factors like location or amenities also affect price.

Exploratory Analysis



BOXPLOT

Entire homes have the widest price range.
Shared rooms rarely exceed €200.
Room type strongly influences price, but outliers and variability exist within each category.



BOXPLOT

Entire homes have the widest price range.
Shared rooms rarely exceed €200.
Room type strongly influences price, but outliers and variability exist within each category.

Correlations

STRONG

accommodates ↔ bedrooms
price ↔ accommodates

WEAK

price ↔ reviews_per_month
(indicates popularity doesn't always mean higher price)

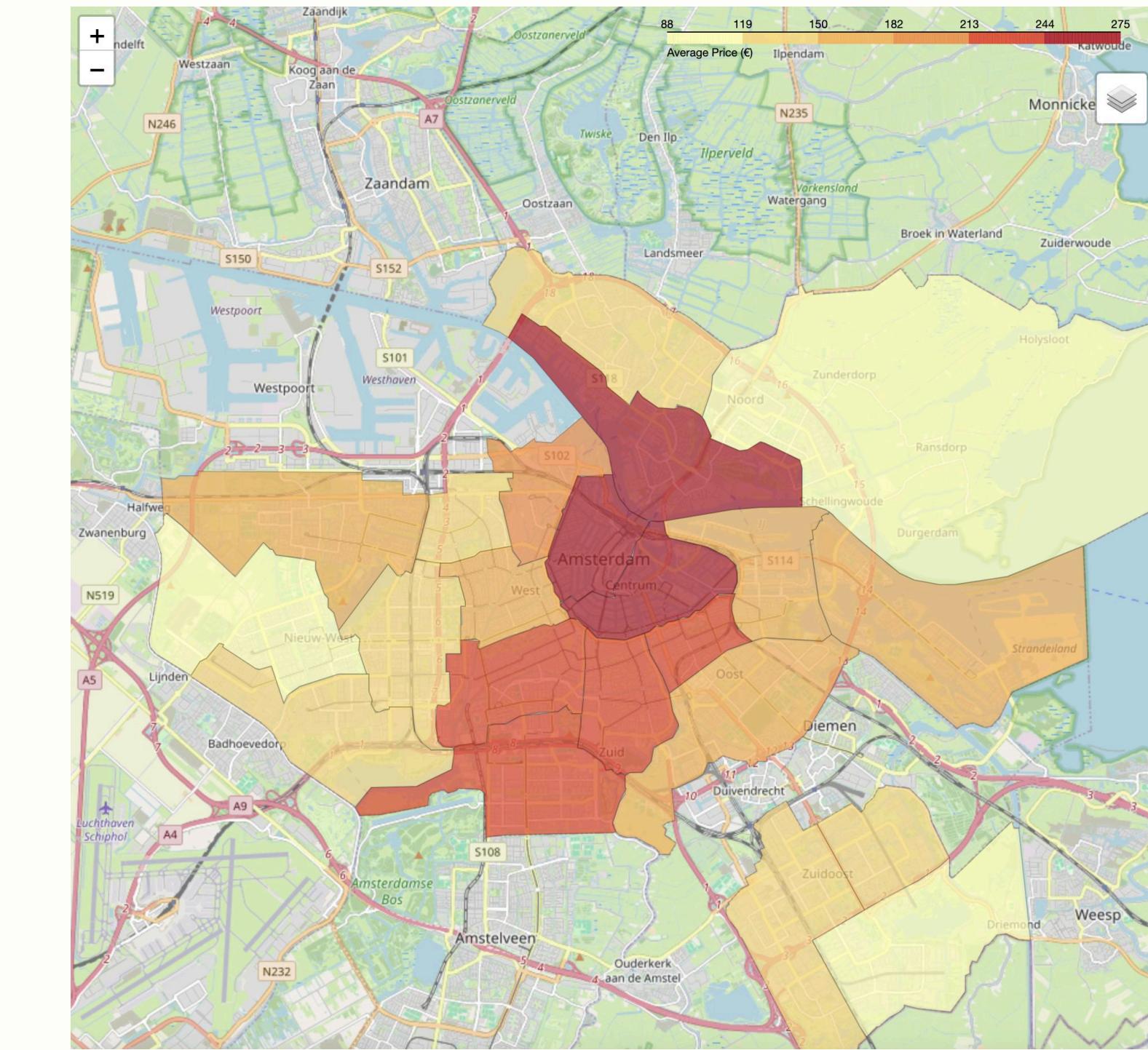
Geo-Spatial Analyse

AVERAGE AIRBNB PRICES BY NEIGHBORHOOD

The **Choropleth map** shows the average nightly price for Airbnb listings in each neighborhood.

Darker colors indicate higher average prices, while lighter colors indicate lower prices.

This visualization highlights which areas are more expensive and helps identify patterns across the city.



Questions & Hypothesis

Hypotheses

01

HOW DO LOCATION AND
NEIGHBORHOOD INFLUENCE
PRICE?

02

ARE HIGH-REVIEW LISTINGS
CONSISTENTLY MORE
EXPENSIVE?

03

ARE EXTREME
MINIMUM_NIGHTS VALUES
AFFECTING THE
DISTRIBUTION?

Listings with more bedrooms and
accommodates will have higher
prices.

Entire homes are priced higher than
private or shared rooms, controlling
for location.

Higher review scores correlate with
slightly higher prices.

New Questions

04

WHY DO CERTAIN
NEIGHBORHOODS
COMMAND HIGHER PRICES
— LOCATION, AMENITIES,
OR HOST QUALITY?

05

ARE THERE CORRELATIONS
BETWEEN PRICE AND
REVIEW SCORES OR
NUMBER OF LISTINGS PER
HOST IN THESE AREAS?

The choropleth map shows the average nightly price for Airbnb listings across neighborhoods. Central and highly touristic areas generally have higher prices, while peripheral neighborhoods tend to be more affordable.

This analysis answers the research question about spatial price distribution and highlights which neighborhoods are most expensive.

Vision & Conclusion

Future Vision

RECOMMENDATIONS

01

Optimize pricing based on neighborhood and room type.

02

Highlight amenities and maintain high review scores.

03

Track seasonal trends to adjust availability and pricing.

04

Use text analysis on reviews to identify guest preferences.

VISION

Implement continuous monitoring of listing performance.
Incorporate insights into Airbnb recommendations and host guidelines.
Expand analysis to multiple years for trend detection.

