

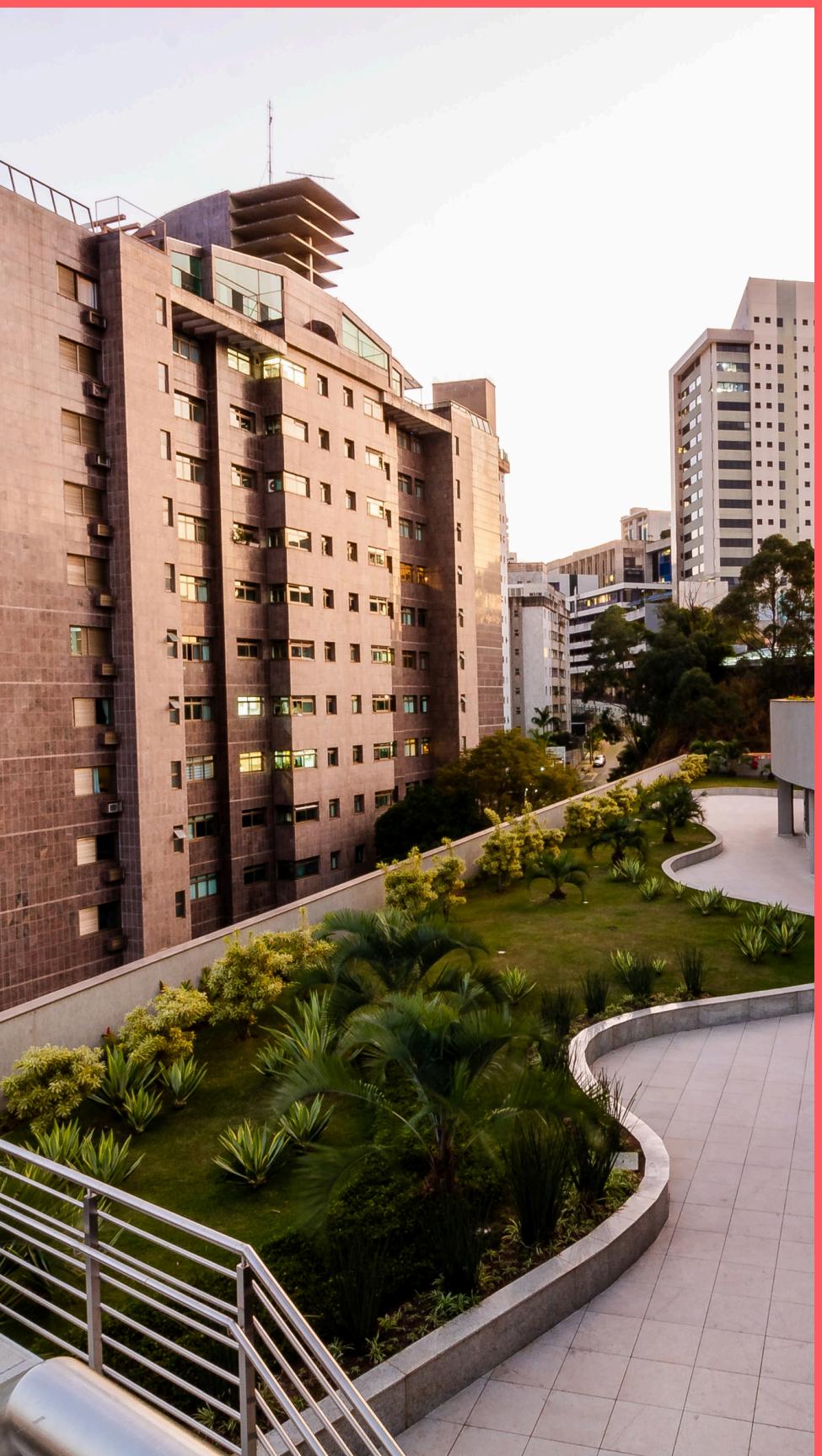


listings and metrics in NYC, NY, USA
(2019)

Presented by **Group 6**

Flory Garima Mehdi Jean Philippe

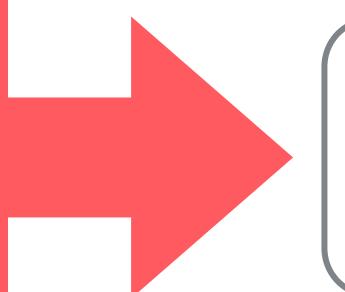




Analysing airbnbs in NYC

Since 2008, guests and hosts have used Airbnb to expand on traveling possibilities and present more unique, personalized way of experiencing the world.

NYC is one of the most popular city destination on the platform and we decided to analyse the listing activity and metrics for 2019.



New York City Airbnb Open Data
Airbnb listings and metrics in NYC, NY, USA (2019)

47k
Listings

11K
Hosts



Business Problem

01

Price:

What factors influence the pricing of Airbnb listings in NYC?

02

Occupancy Analysis:

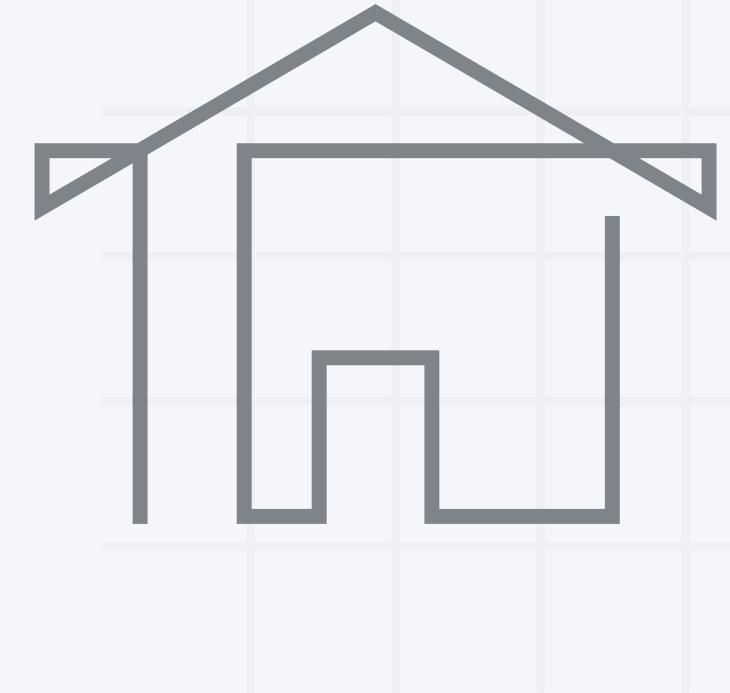
Which neighbourhoods have the highest and lowest occupancy rates?

What is the correlation between Prices and length of stay?

03

Reviews:

Does the number of reviews affects the price of an apartment/house?



Initial Hypothesis

01 hypothesis

Price

Listings in Manhattan are more expensive than those in other neighbourhood.

02 hypothesis

Price

Entire home/apt listings have higher prices than private or shared rooms.

03 hypothesis

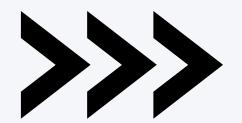
Review

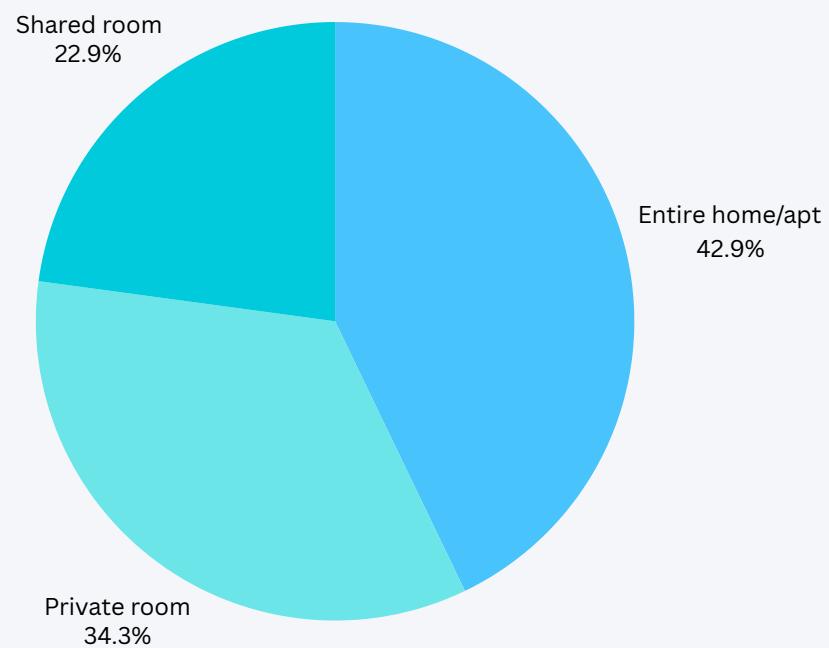
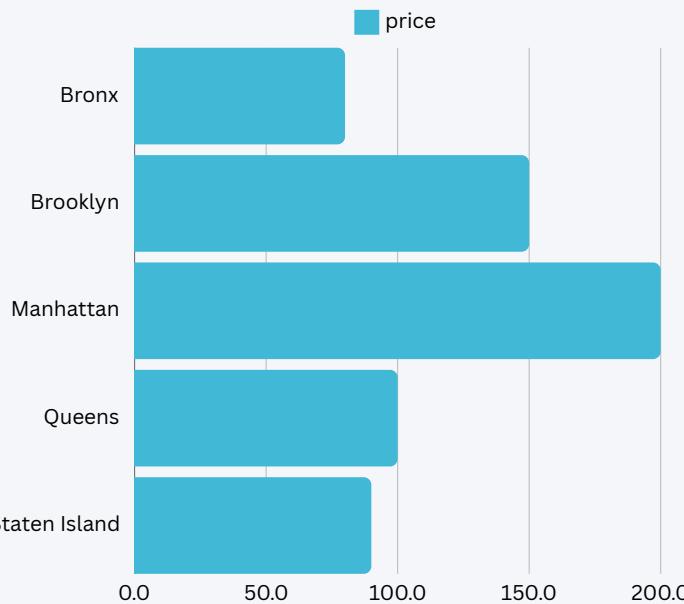
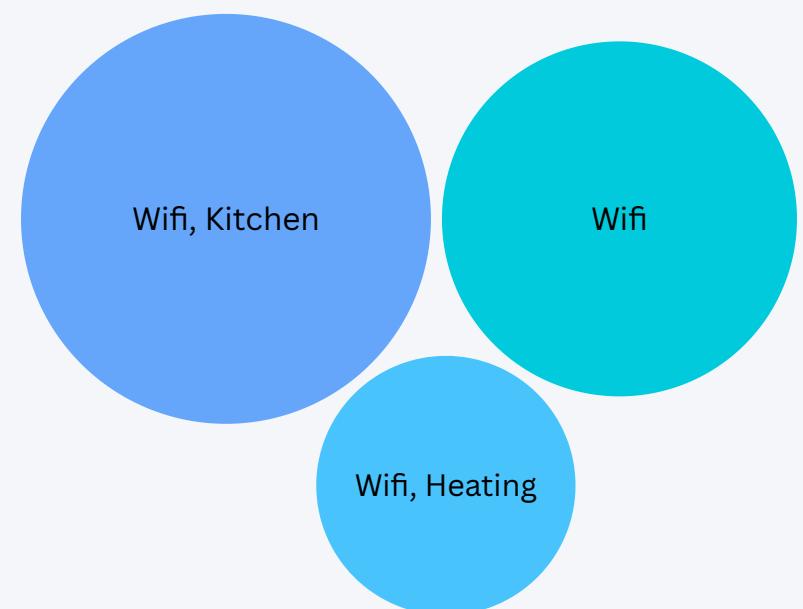
Homes with the most reviews are more expensive.

04 hypothesis

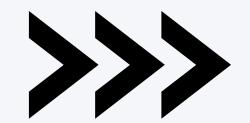
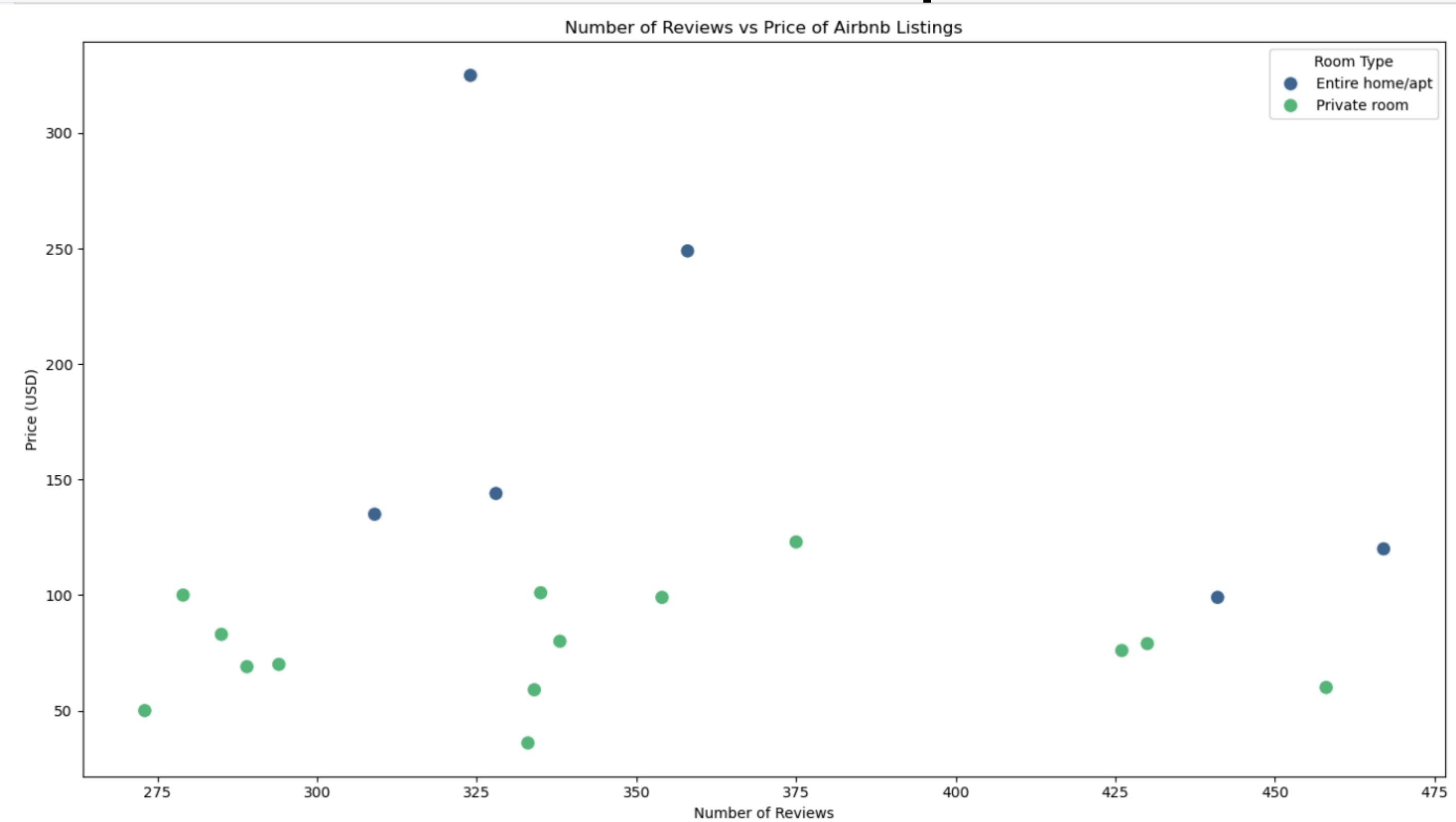
Duration

Prices are affected by length of stay.

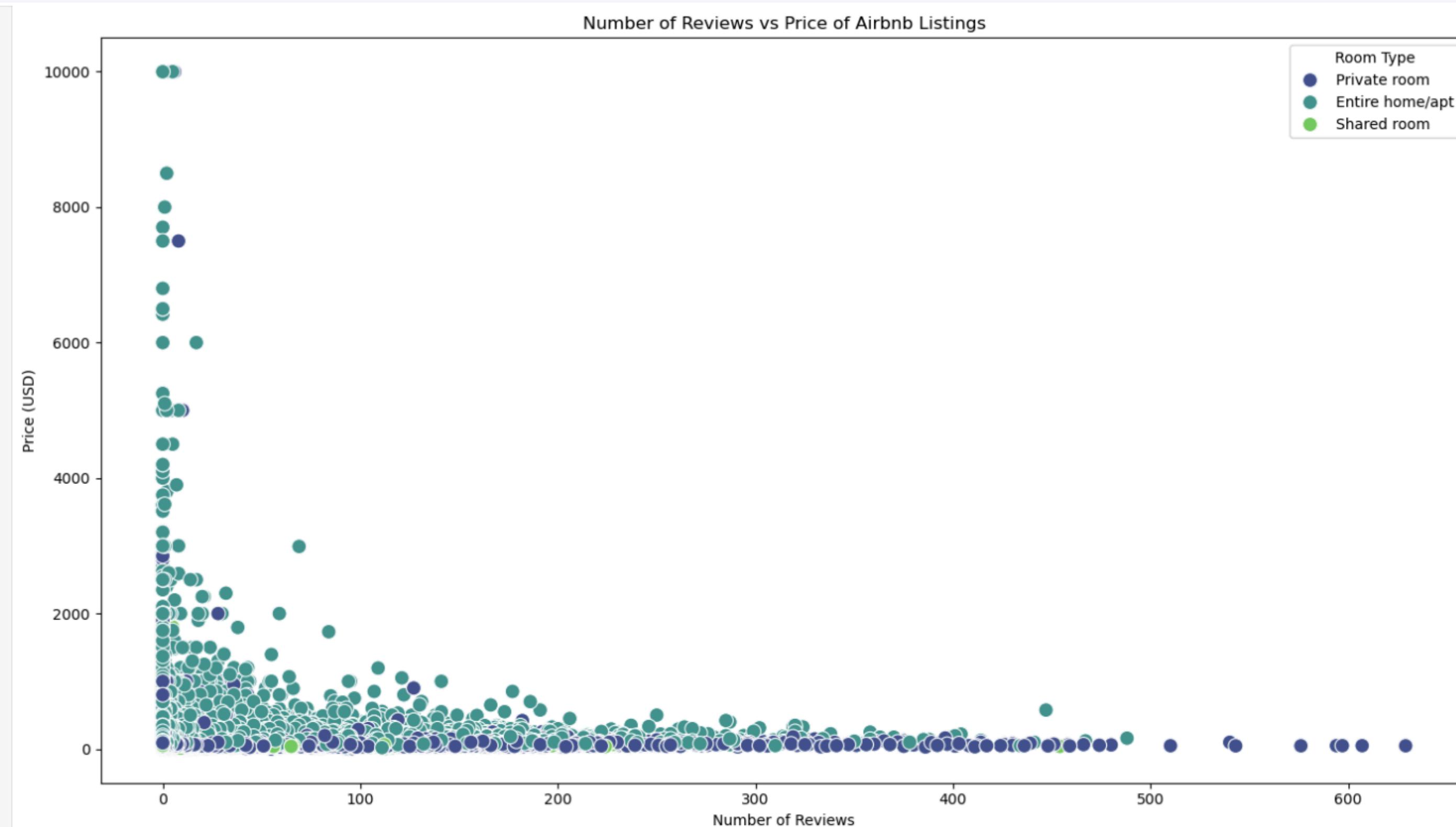


Average Price by Room Type**Average price by neighbourhood****Average price with wifi****Average price by Superhost**

No strong correlation between # of reviews and price (most reviews - top 20)



No strong correlation between # of reviews and price (entire dataset)



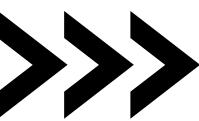
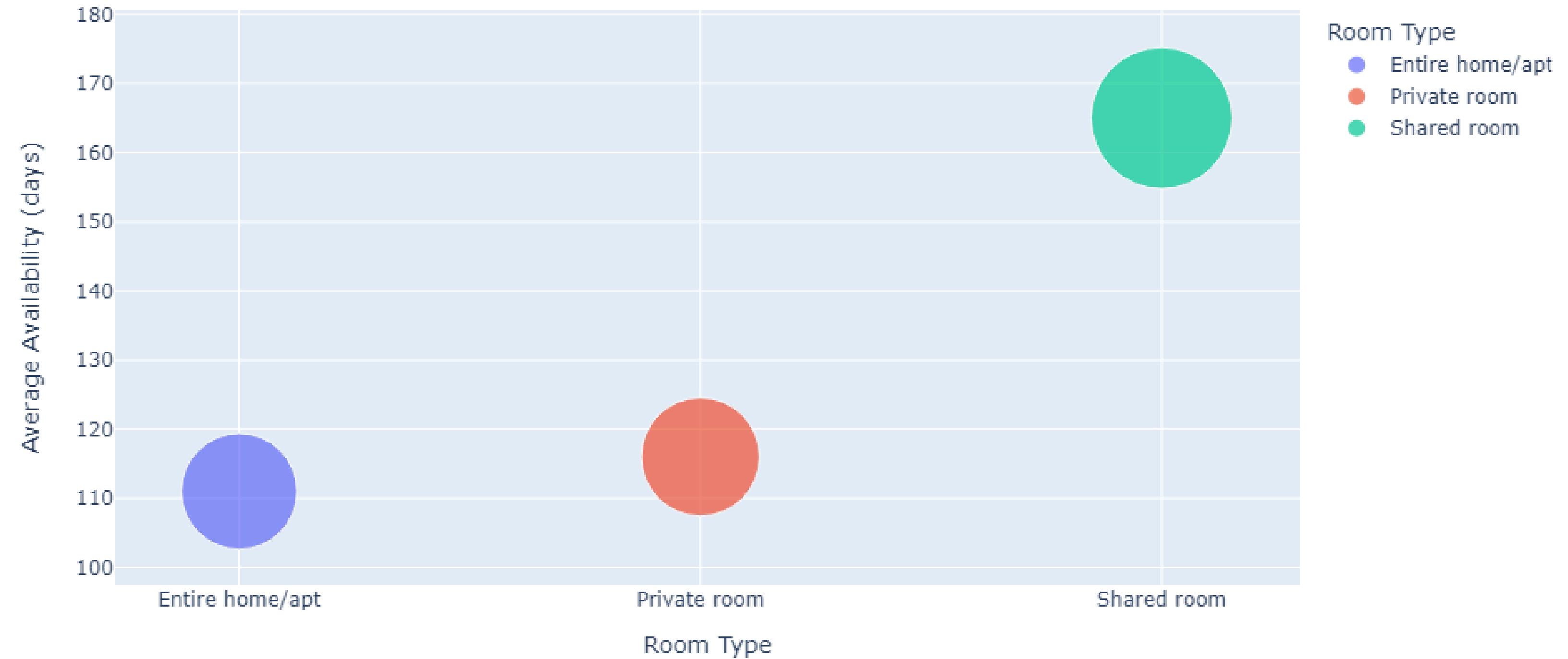
NEIGHBORHOOD WITH HIGHEST AND LOWEST OCCUPANCY

	Neighborhood	Neighborhood Group	Occupancy Rate (%)
Highest Occupancy	Bay Terrace, Staten Island	Staten Island	100.00
Lowest Occupancy	Co-op City	Bronx	0.27



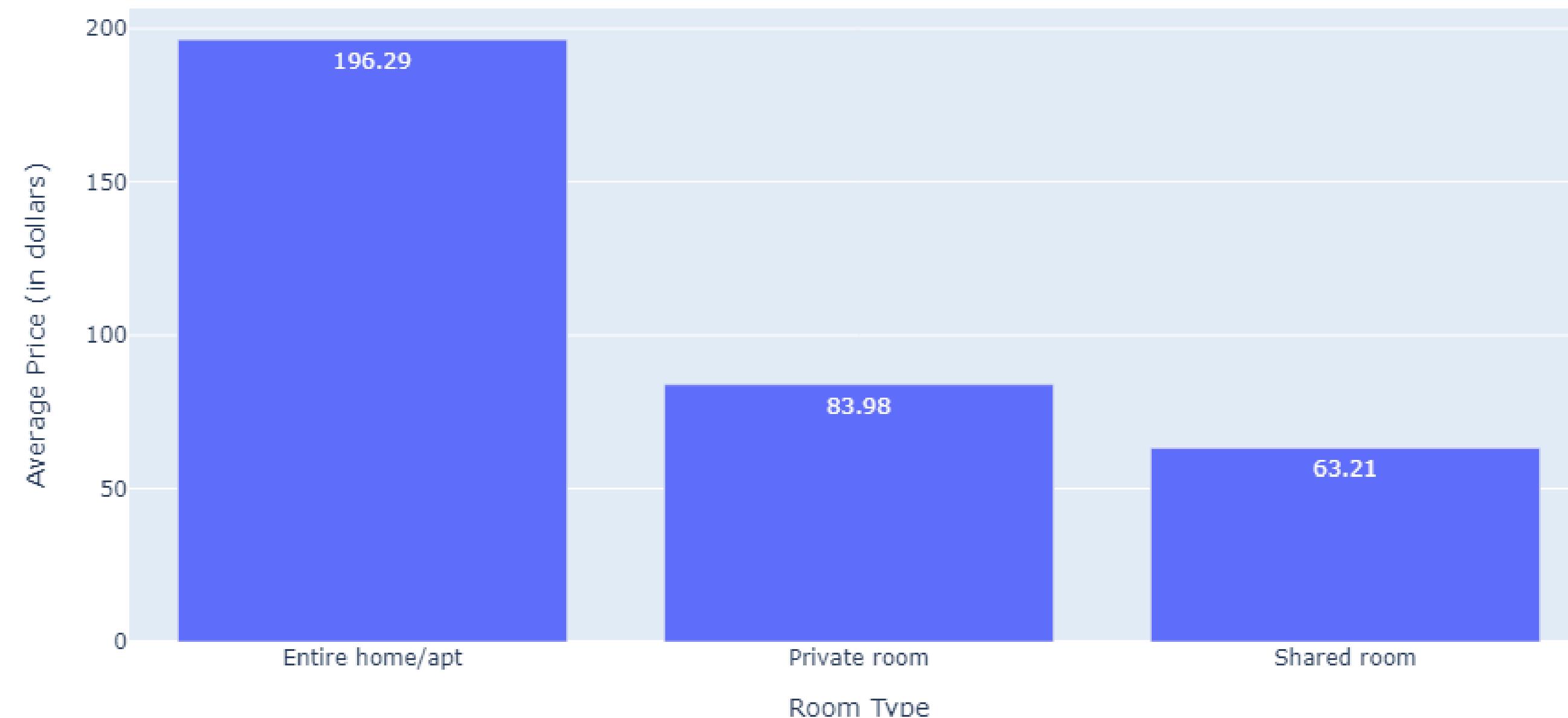


Room Type And Availability

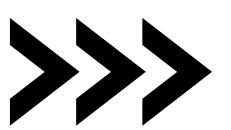


Homes/Apartments Have Higher Prices Over Other

Average Prices by Room Type

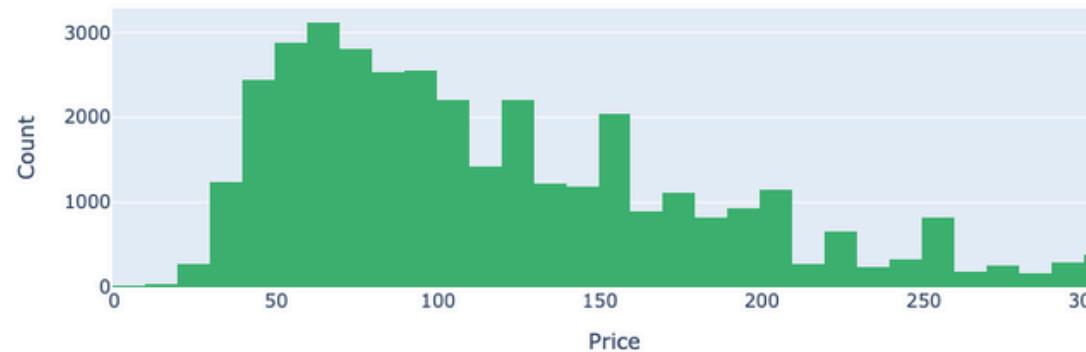


Hypothesis:
Full listings have
higher Prices.



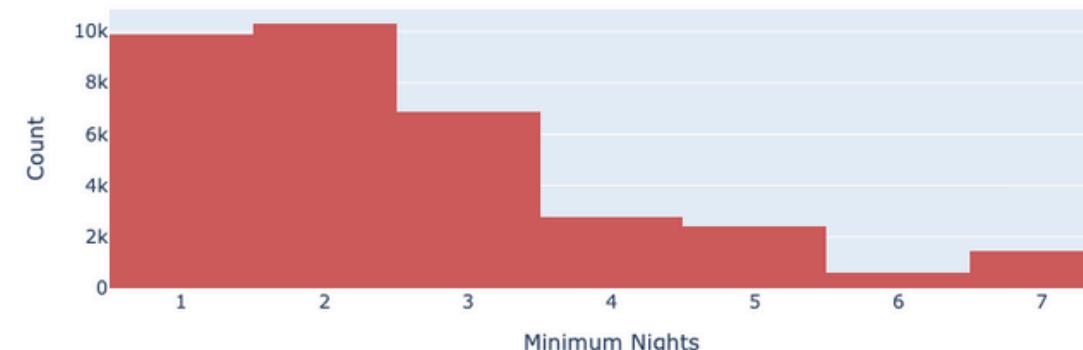
What is the correlation between Prices and length of stay?

Distribution of Prices (Up to \$300)



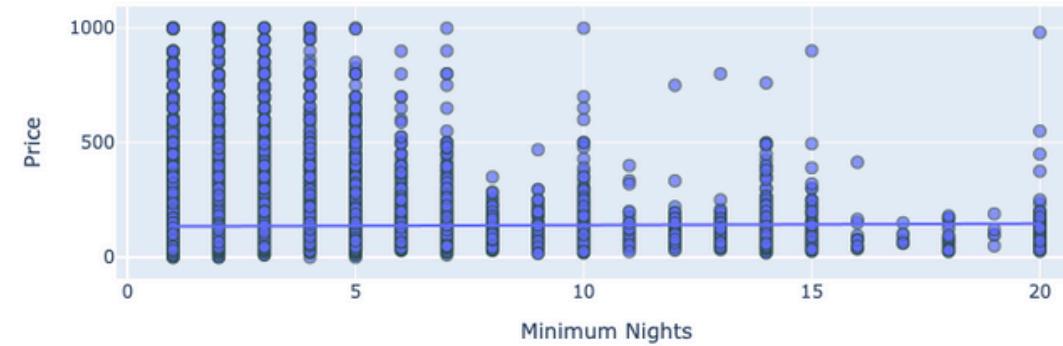
→ Most listings analysed are priced between \$20 to \$260 per night.

Distribution of Length of Stay (Minimum Nights, Max 7 Nights)



→ The most popular length of stay imposed per listings ranges between 1-3 nights with **2 Nights Min Stay** the most popular.

Correlation Between Price and Minimum Nights (Up to 20 Nights, Price <= \$1,000)

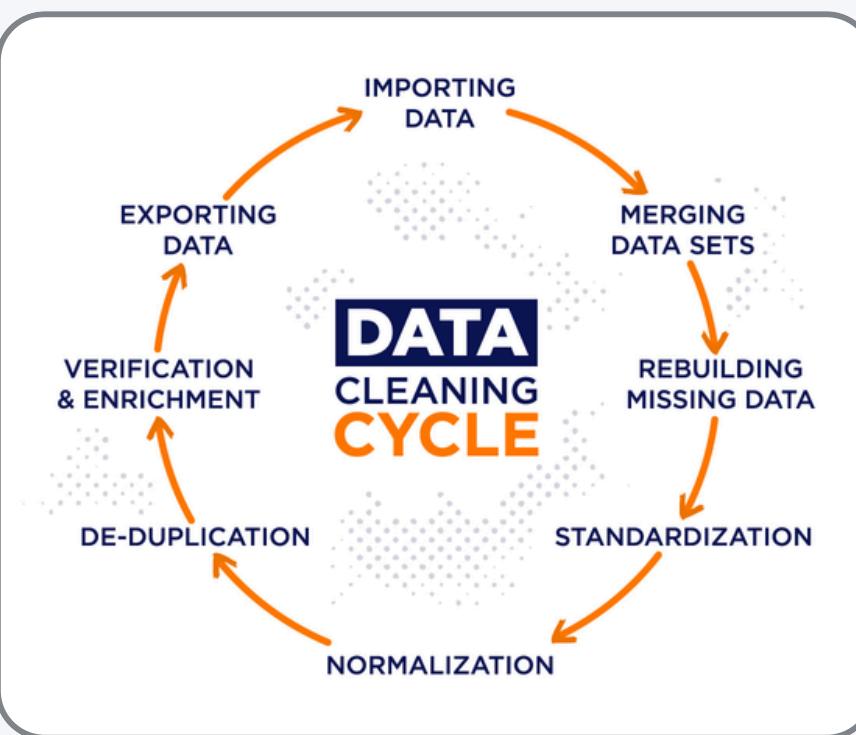
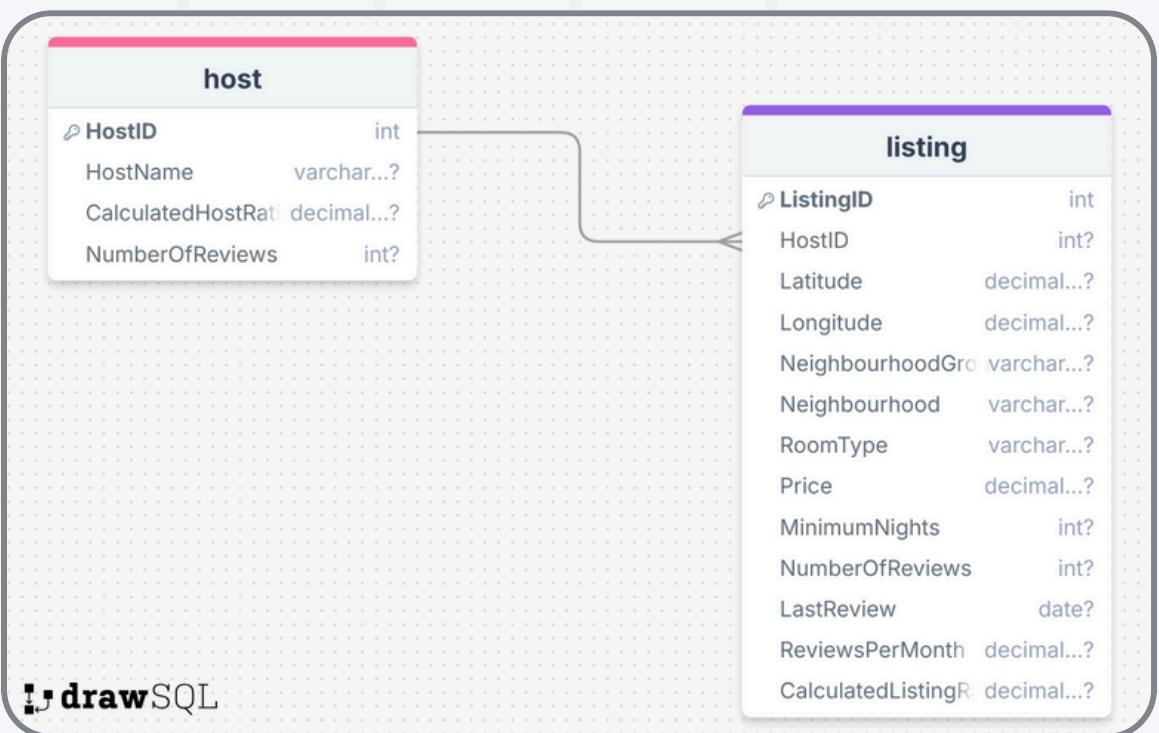


- Short Stays tends to **cost higher** compared to longer stays.
- **\$ 142** is the mean price and **5 nights** is the mean length of stay of studied listings in NY.
- 75% of listings costs below **\$170**.

Hypothesis:

Prices are affected by length of stay.





Our Method

01 Created our Database Schema drawing to have a clear understanding of the relationship between our tables.

02 Cleaned our raw data using python.

03 Analysed the cleaned data using MySQL query and Python and used libraries like plotly to represent our findings graphically.





Our Difficulties

01

Creation and Importation of the database into MySQL workbench with many errors.

02

Switch between MySql and Jupyter and not being able to fetch MySql data into Jupyter. needed duplicated work/files.





The project proved how data analysis support decision making, as one of our initial hypothesis was false.

Data analysis must support business decision making

01

Manhattan is indeed the most expensive neighbourhood.



02

The number of reviews do not have a direct impact of the price



03

The longer the length of stay is, the lower will be the price.



04

The availability of shared rooms is greater than full listings and whole listings have higher prices than private listings.



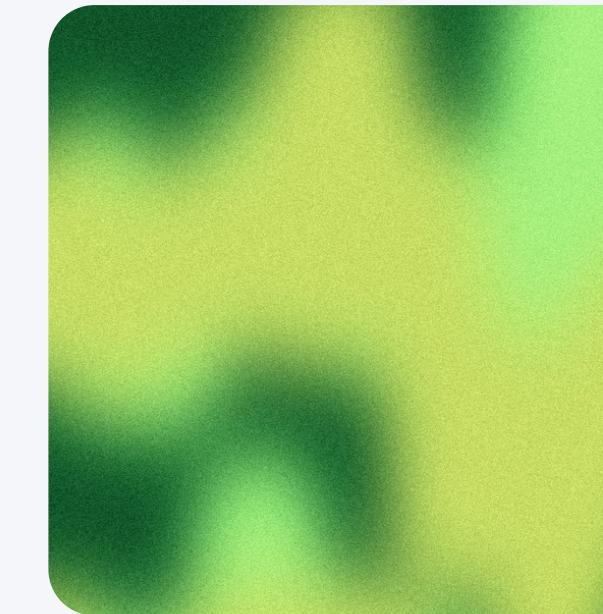
Thank You



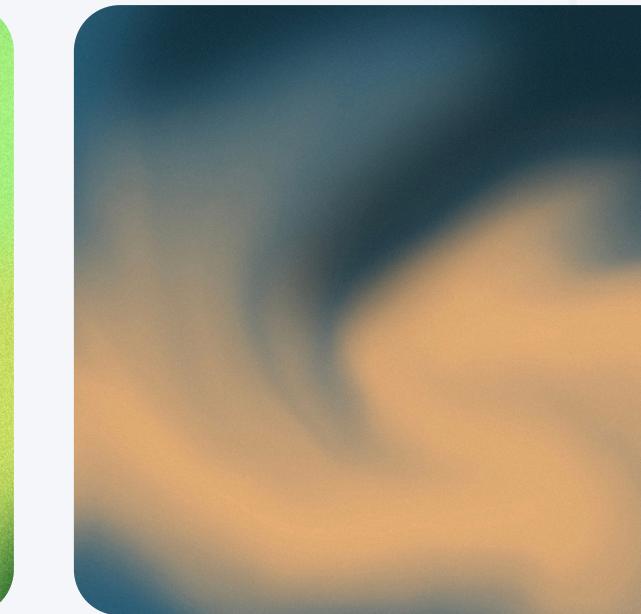
Flory
Data Analyst



Garima
Data Analyst



Mehdi
Data Analyst



Jean Philippe
Data Analyst

