



Crafting Unforgettable Stays at San Francisco through the Art of Data

*By Ming Thet Paing,
Pone Nyet Aung,
Naing Lin Thu,
Eaint Thet Hmu*



Table Of Content

- *Summary of Data*
- *Data Presentation*



Data Summary

- **Main Theme - Rental Properties**
- **Rows** - 8111
- **Columns** - 9
- **Attributes** - *ID, Latitude, Longitude, Property_Type, Room_Type, Bedrooms, Bathrooms, Minimum_Nights, Price*
- **Data Types**
- **Integer** - *ID, Minimum_Nights*
- **Float** - *Latitude, Longitude, Bedrooms, Bathrooms*
- **Object (String)** - *Price, Property_Type, Room_Type*

```
[8111 rows x 9 columns]
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8111 entries, 0 to 8110
Data columns (total 9 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   id               8111 non-null    int64  
 1   latitude         8111 non-null    float64 
 2   longitude        8111 non-null    float64 
 3   property_type   8111 non-null    object  
 4   room_type        8111 non-null    object  
 5   bathrooms        8099 non-null    float64 
 6   bedrooms         8107 non-null    float64 
 7   minimum_nights  8111 non-null    int64  
 8   price             8111 non-null    object  
dtypes: float64(4), int64(2), object(3)
memory usage: 570.4+ KB
```



Data Cleaning

#	Column	Non-Null Count	Dtype
0	id	8111 non-null	int64
1	latitude	8111 non-null	float64
2	longitude	8111 non-null	float64
3	property_type	8111 non-null	object
4	room_type	8111 non-null	object
5	bathrooms	8099 non-null	float64
6	bedrooms	8107 non-null	float64
7	minimum_nights	8111 non-null	int64
8	price	8111 non-null	object

df_sorted = df.sort_values(by=['minimum_nights'], ascending=True)	print(df_sorted)
	id latitude longitude property_type room_type bathrooms \
0	958 37.76931 -122.43386 Apartment Entire home/apt 1.0
3769	28408154 37.75747 -122.48785 House Private room 1.5
3773	28424542 37.74406 -122.45918 Guest suite Entire home/apt 1.0
3779	28439767 37.72271 -122.40442 House Private room 1.0
3781	28444584 37.73404 -122.47484 House Entire home/apt 4.5
...	...
57	51374 37.76519 -122.45613 Apartment Entire home/apt 1.0
312	505763 37.75081 -122.44524 Apartment Entire home/apt 1.0
3664	19813606 37.79305 -122.42482 Condominium Entire home/apt 1.0
1893	8066789 37.76264 -122.42082 Apartment Private room 2.0
2912	15344978 37.75071 -122.42521 Apartment Private room 1.0
	bedrooms minimum_nights price
0	1.0 1 170.0
3769	1.0 1 120.0
3773	0.0 1 100.0
3779	1.0 1 80.0
3781	4.0 1 260.0
...	...
57	2.0 365 9999.0
312	1.0 265 200.0
3664	1.0 1000 150.0
1893	1.0 1125 3200.0
2912	1.0 100000000 68.0

- **Step 1 - Check**

Check the data source - Airbnb, San Francisco, California, USA

Check Wrong Data Type - Price = Object

Check Missing Data - Bedrooms (4) & Bathrooms(12)

Check Duplicates - No duplicate

- **Step 2 - Replace**

Replace Wrong Data Type - Price - Object to Float

Replace Missing Data - Bedrooms (4) & Bathrooms(12) with Median Value

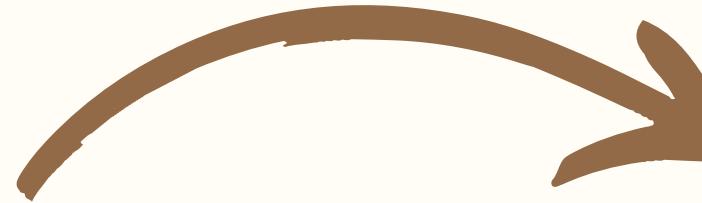
- **Step 3 - Remove**

Remove index 3664, 1893, and 2912 since there are extreme values in minimum nights



Before and After Cleaning

Column	Non-Null Count	Dtype
id	8111 non-null	int64
latitude	8111 non-null	float64
longitude	8111 non-null	float64
property_type	8111 non-null	object
room_type	8111 non-null	object
bathrooms	8099 non-null	float64
bedrooms	8107 non-null	float64
minimum_nights	8111 non-null	int64
price	8111 non-null	object



```
Data columns (total 9 columns):
 #   Column            Non-Null Count   Dtype  
 --- 
 0   id                8108 non-null    int64  
 1   latitude          8108 non-null    float64 
 2   longitude         8108 non-null    float64 
 3   property_type    8108 non-null    object  
 4   room_type         8108 non-null    object  
 5   bathrooms         8108 non-null    float64 
 6   bedrooms          8108 non-null    float64 
 7   minimum_nights   8108 non-null    int64  
 8   price              8108 non-null    float64 
 dtypes: float64(5), int64(2), object(2)
 memory usage: 633.4+ KB
```

```
df.describe()
```

	id	latitude	longitude	bathrooms	bedrooms	minimum_nights	price
count	8.111000e+03	8111.000000	8111.000000	8111.000000	8111.000000	8.111000e+03	8111.000000
mean	2.024561e+07	37.766054	-122.430107	1.395389	1.345703	1.234526e+04	225.407101
std	1.228565e+07	0.022937	0.026967	0.922655	0.925102	1.110357e+06	412.253039
min	9.580000e+02	37.704630	-122.513060	0.000000	0.000000	1.000000e+00	0.000000
25%	8.905668e+06	37.751450	-122.442830	1.000000	1.000000	2.000000e+00	100.000000
50%	2.161057e+07	37.769150	-122.424650	1.000000	1.000000	4.000000e+00	150.000000
75%	3.120025e+07	37.785670	-122.410615	1.500000	2.000000	3.000000e+01	240.000000
max	3.935418e+07	37.828790	-122.368570	14.000000	14.000000	1.000000e+08	10000.000000

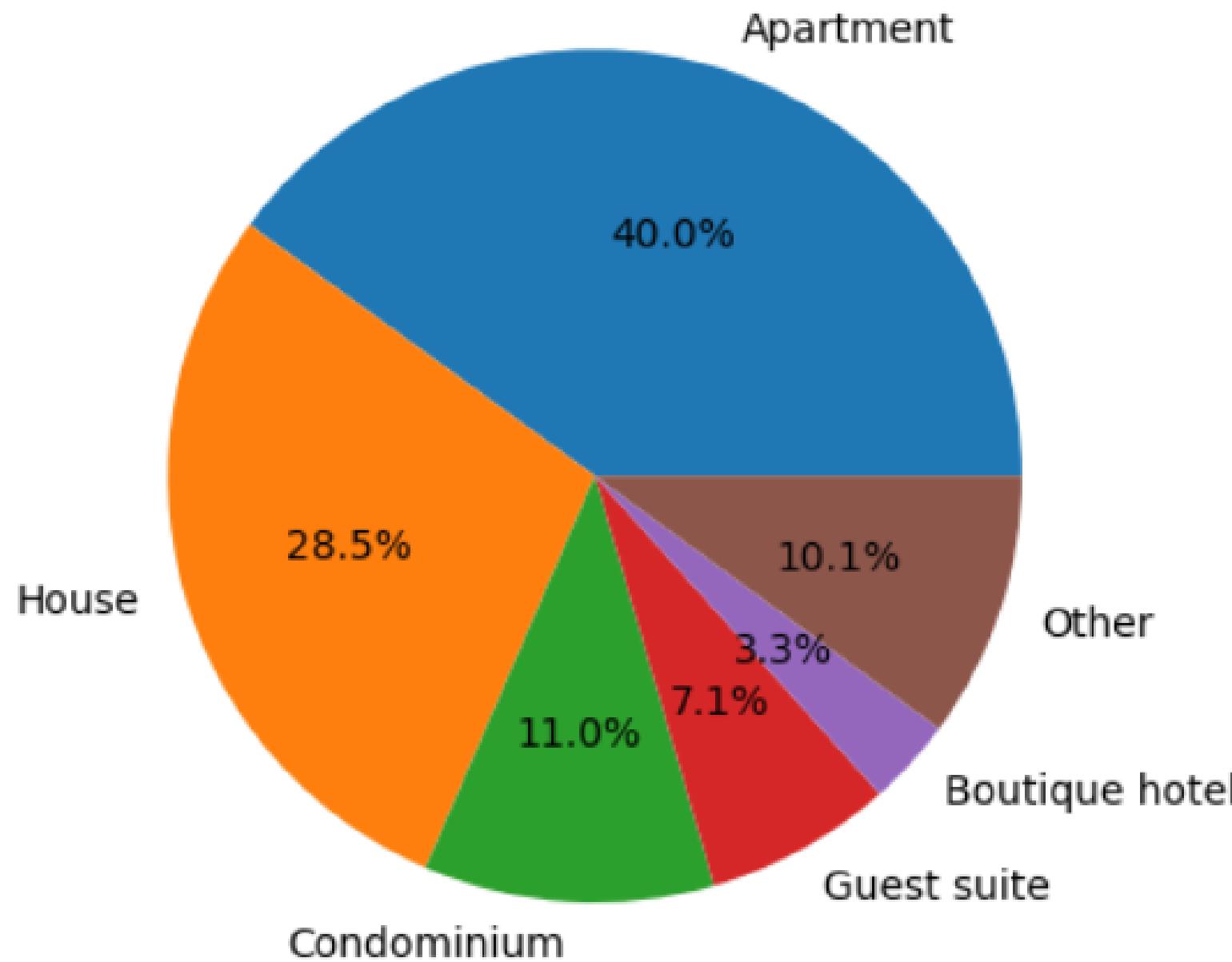


	id	latitude	longitude	bathrooms	bedrooms	minimum_nights	price
count	8.108000e+03	8108.000000	8108.000000	8108.000000	8108.000000	8108.000000	8108.000000
mean	2.024777e+07	37.766053	-122.430110	1.395412	1.345831	16.070054	225.068944
std	1.228706e+07	0.022938	0.026972	0.922781	0.925249	22.219623	410.998977
min	9.580000e+02	37.704630	-122.513060	0.000000	0.000000	1.000000	0.000000
25%	8.906478e+06	37.751450	-122.442833	1.000000	1.000000	2.000000	100.000000
50%	2.161180e+07	37.769155	-122.424645	1.000000	1.000000	4.000000	150.000000
75%	3.120736e+07	37.785670	-122.410610	1.500000	2.000000	30.000000	240.000000
max	3.935418e+07	37.828790	-122.368570	14.000000	14.000000	365.000000	10000.000000

What are the most common property type?



Distribution of Top 5 Property Types

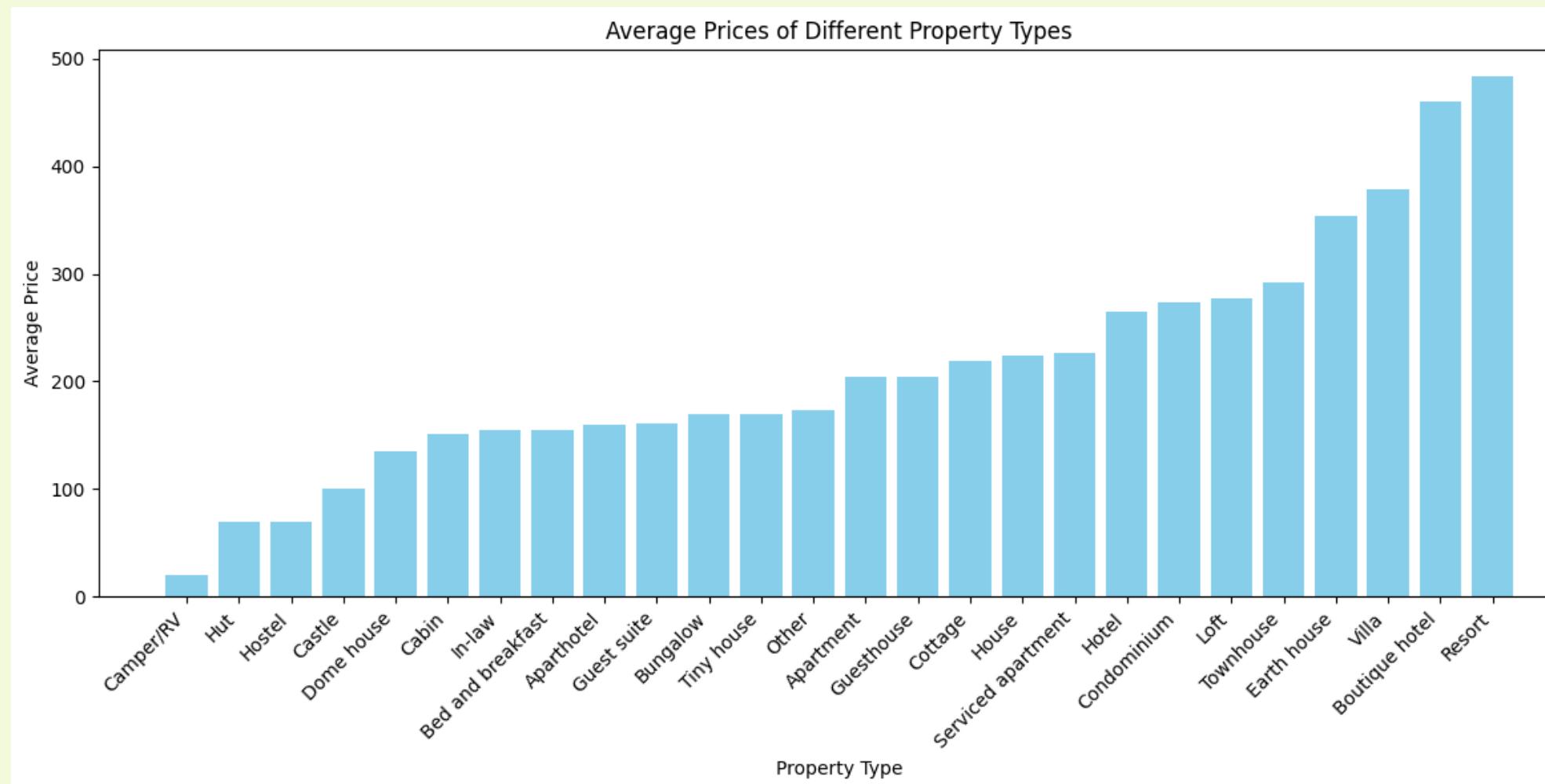


There are **26** different property types in this dataset.

Over one-third of the property types of the records in this dataset are apartments and houses.

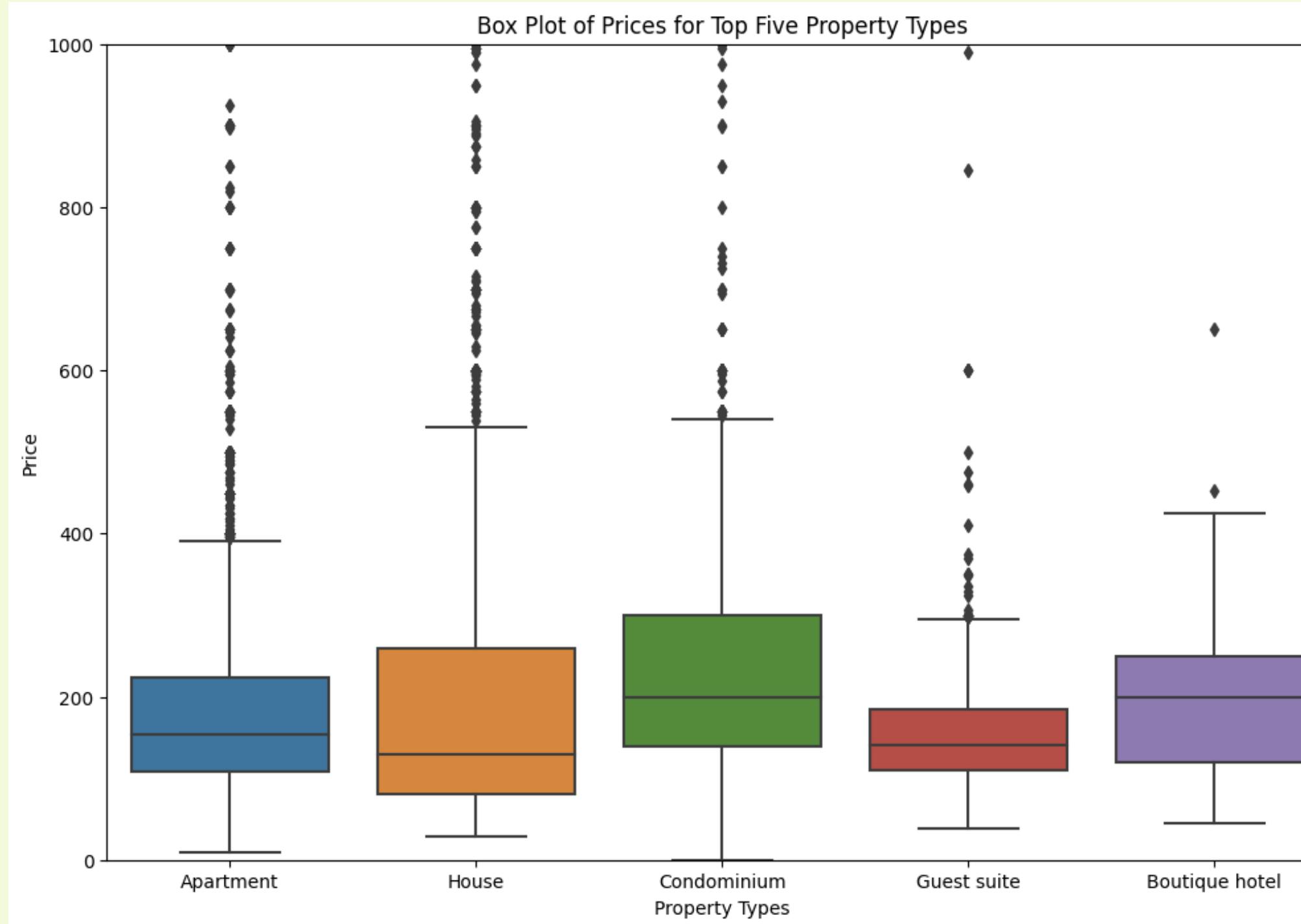


What are the different average prices based on property types?



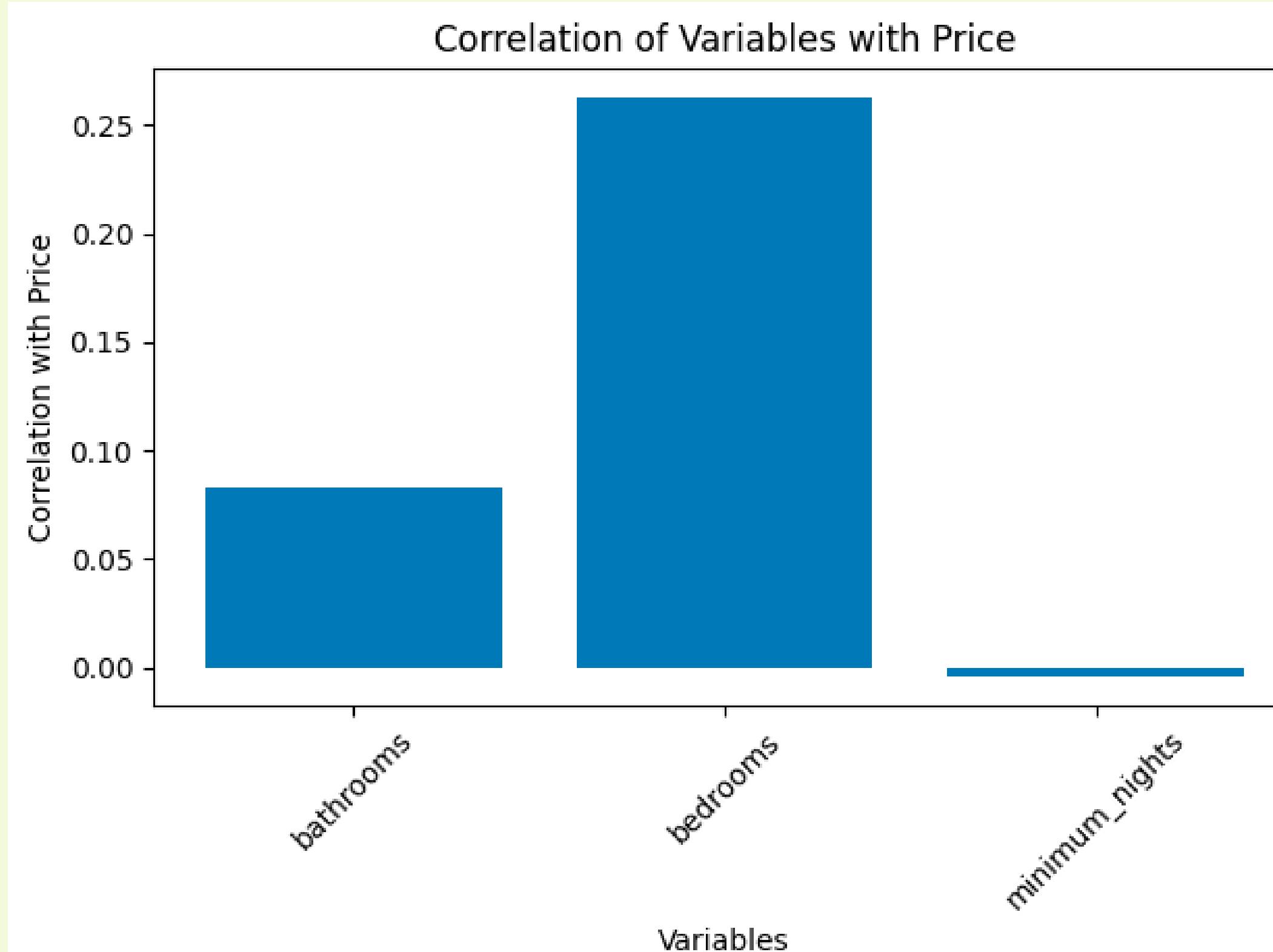
- These are the different average prices for different property types in San Francisco.
- The rent price starts at low as \$20 and as high as \$484.
- Among them, Camper/RV is at the lowest and Resort is at the highest in terms of price.

What are the price differences for the most five common property types?



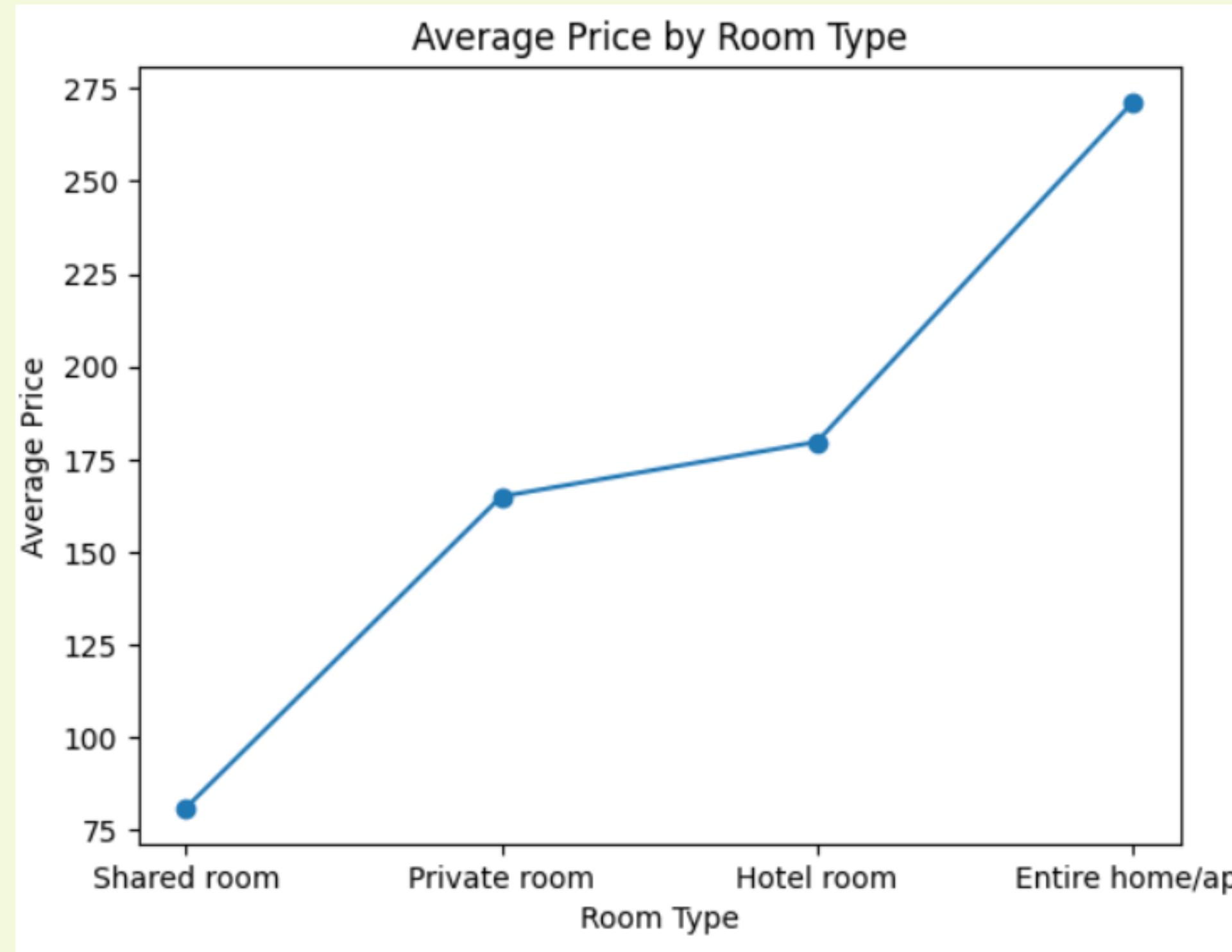
- 25% of the most common property are mostly around \$100 and below (Q1)
- other 25% are nearly \$200 and above (Q3)
- 50% are mostly between \$100 and \$250 (interquartile)
- A few outliers for each property

What factors (property types, minimum nights, etc.) have the most significant impact on pricing?



- Bedrooms have a significant effect on the prices since it has a 0.3 positive correlation with price which is more than the other factors.
- It is less likely to have a strong relationship between the minimum nights required and the price since there is a negligible correlation between them.

How do the prices differ based on room types?



Average price by room type

Shared room = 80.933610

Private room = 165.002075

Hotel room = 179.710145

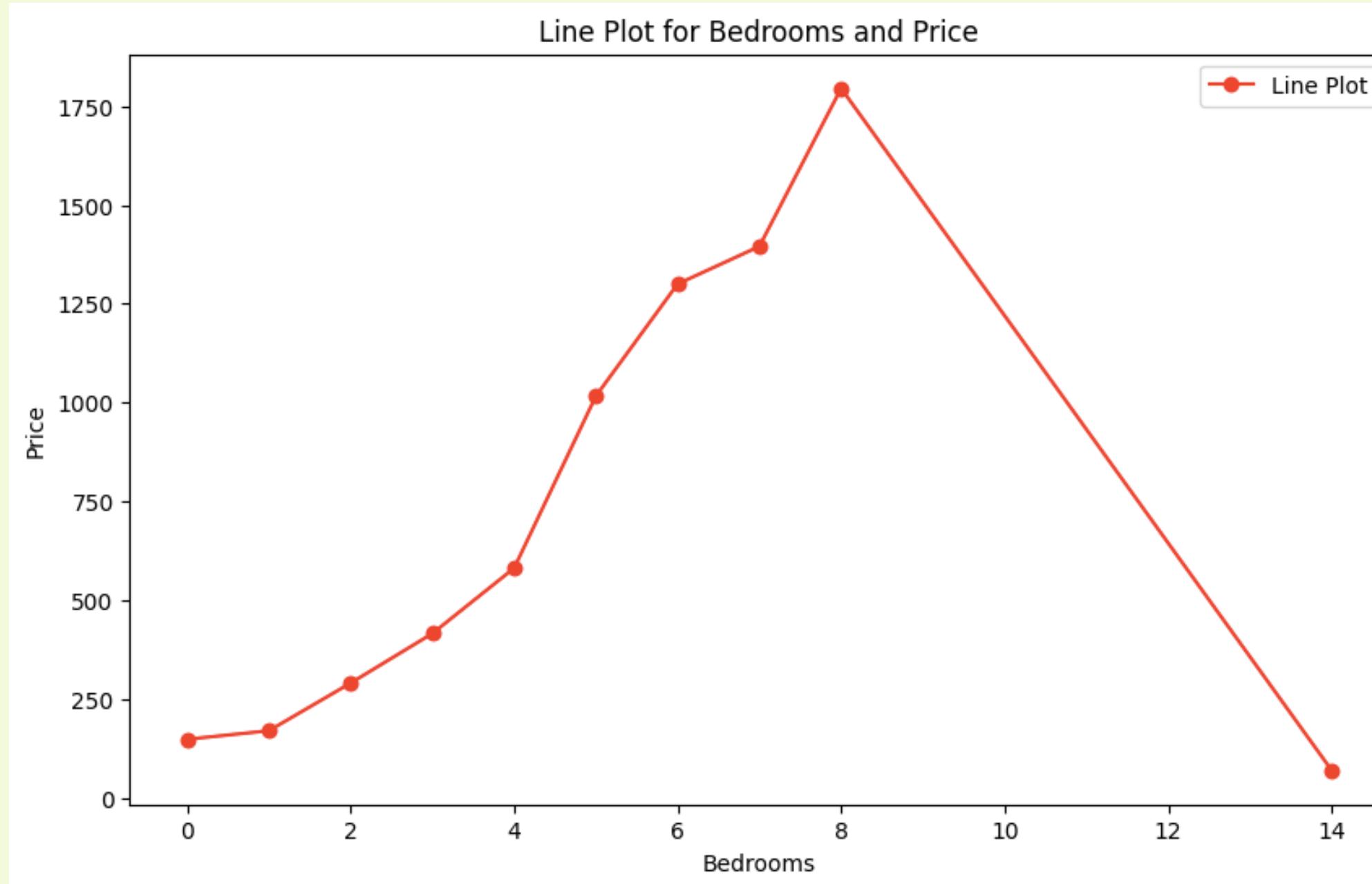
Entire home/apt = 271.280595

Entire home/apt is the most expensive room type and shared rooms is the cheapest in average.

Entire homes/apt are over 3 times more expensive than shared rooms

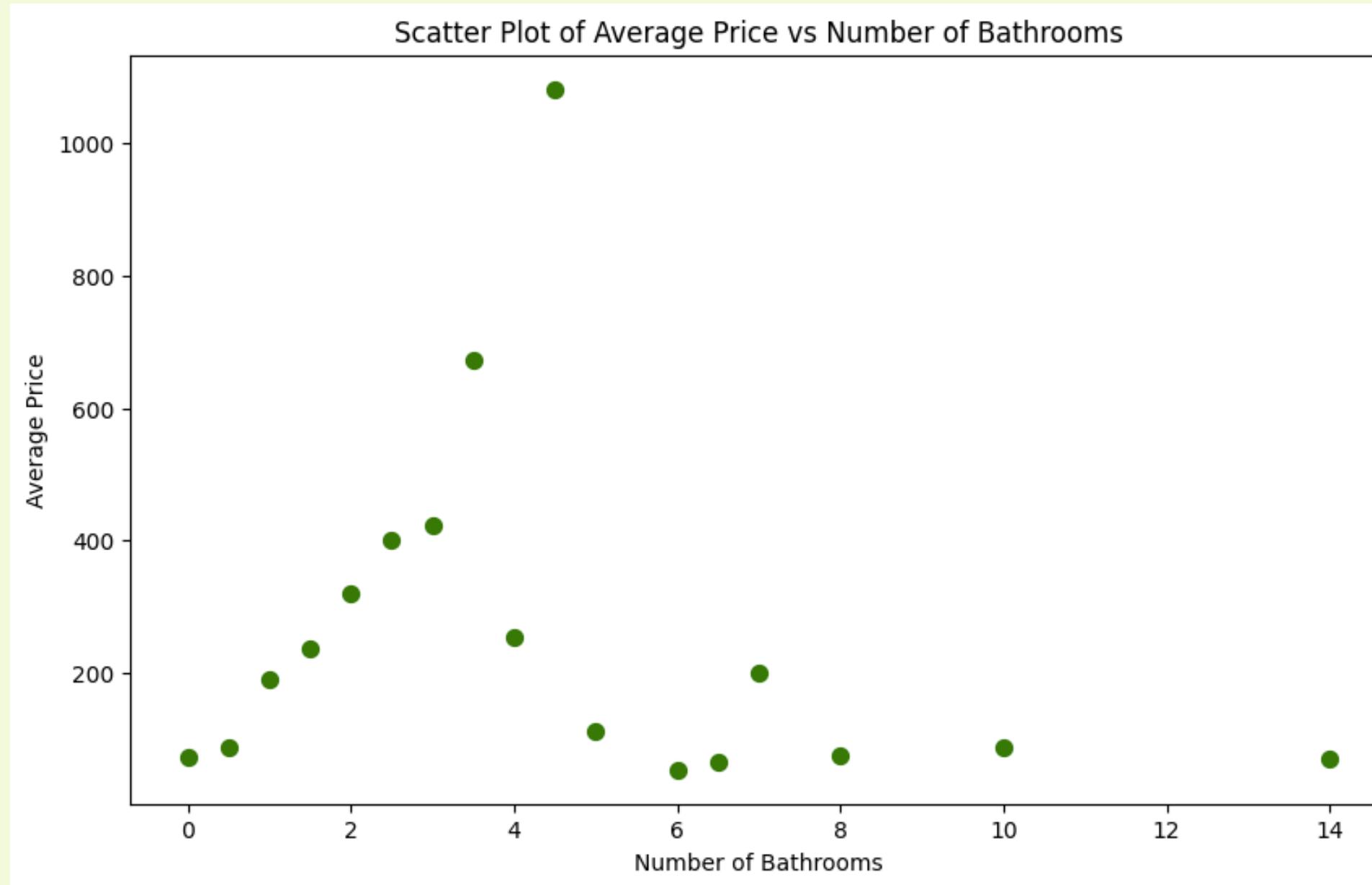
Hotel rooms are only a bit more expensive than private rooms.

How do the prices differ based on number of bedrooms?



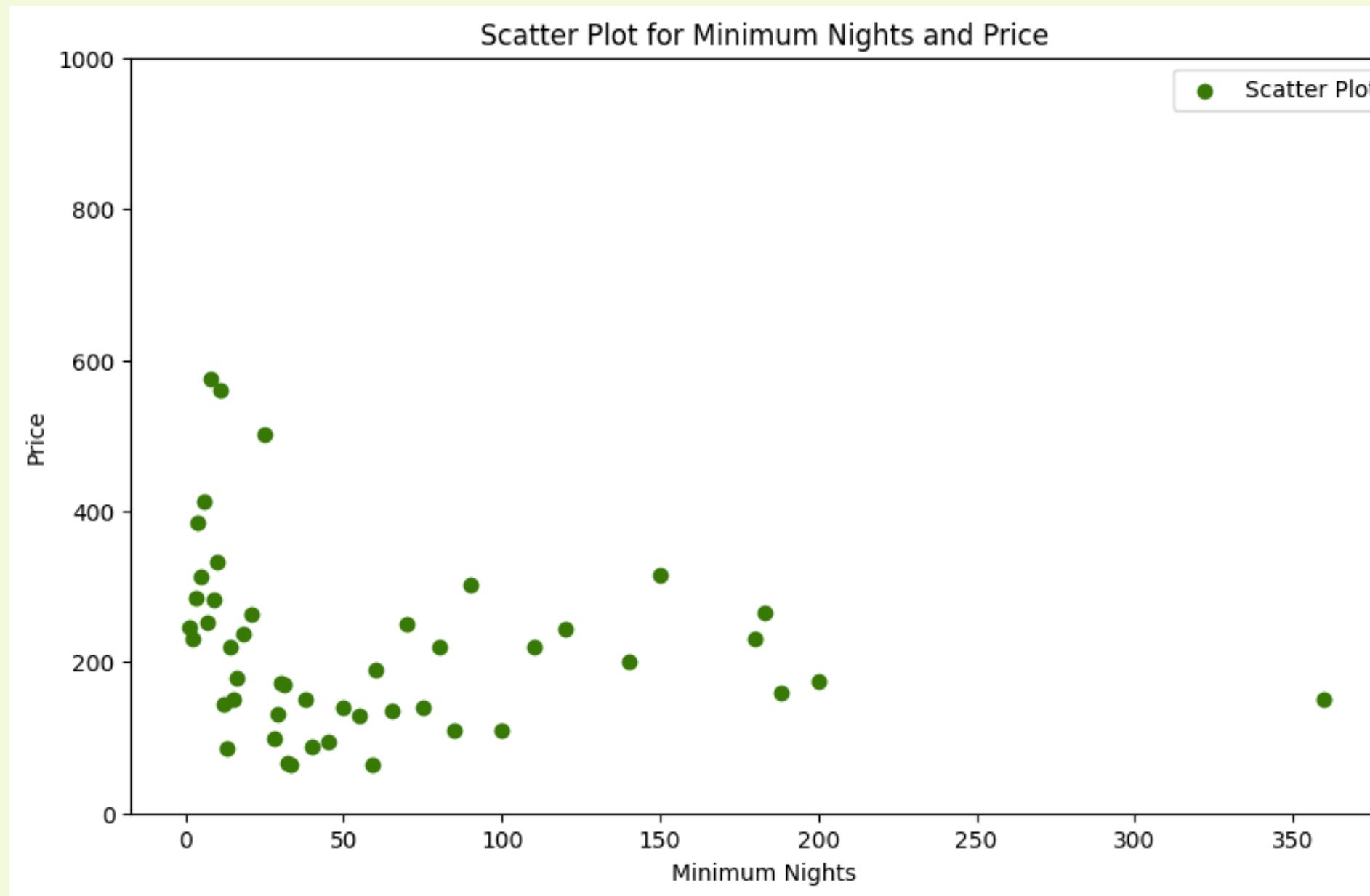
- Positive correlation between price and bedrooms
- As the number of bedroom increases, the price will also increase
- But, there is an outlier for 14 bedroom hotel

How do the prices differ based on number of bathrooms?



- A moderate positive correlation between the number of bathrooms and the average price of the property
- Properties with more bathrooms tend to be more expensive
- Correlation is not perfect, however, as there is some scatter in the data. This means that there are some properties with more bathrooms that are less expensive than properties with fewer bathrooms, and vice versa.

How do the prices differ based on minimum nights?



- A weak positive correlation between the minimum number of nights you have to stay and the price.
- The more nights you stay, the more expensive the price is likely to be.
- A few outliers in the data, such as the point at (200 nights, \$400)



Thank You

Contact Us!

+1 855-424-7262

www.airbnb.com or

hello@airbnb.com

888 Brannan St, San Francisco, CA 94103, USA

Ming Thet Paing (*Senior Sales Executive*)

mingthetpaing.airbnb@gmail.com

Naing Lin Thu (*Junior Sales Executive*)

nainglinthu.airbnb@gmail.com